

**THE INFLUENCE OF FEATURES OF COLLOCATIONS ON THE
COLLOCATIONAL KNOWLEDGE AND DEVELOPMENT OF KURDISH HIGH
SCHOOL STUDENTS: A LONGITUDINAL STUDY**

Submitted by Jamal Muhammad Ramadhan, to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Education

in June 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other university.

Signature:

Abstract

This study explored the influence of four features of collocations- frequency of occurrence, syntactic structure, semantic transparency, and congruency with L1- on the collocational knowledge and development of 252 Kurdish high school learners of English as a foreign language. The importance of collocations in learning English as a second or foreign language and the difficulties that challenge learners at different levels of language proficiency have been well established. However, few studies have adopted a longitudinal research design or a hybrid definition of collocations, incorporating both frequency-based and phraseological views. The present study took this approach to explore learners' collocational knowledge and development and the influence of features of collocations on their collocational knowledge and development at the high school level of learning English as a foreign language. The study employed two tests: an appropriateness judgement test to measure learners' receptive knowledge and a gap-filling test to measure their productive knowledge of collocations.

The data were collected in two waves, one at the beginning of their school year and the other at the end. Data analyses were conducted to determine the relationship between features of collocations and learners' collocational knowledge and development. The results revealed frequency of occurrence as the most influential factor affecting learners' knowledge and development. Influence of the syntactic structure of collocations on the learners' knowledge and development came second whereas congruency with L1 occupied the third position. Semantic transparency seemed to have the least influence on their collocational knowledge and development. Gender appeared as an influential factor in the individual tests. However, its influence was not significant in terms of overall knowledge development. In general, the results indicated that learners' productive collocational knowledge lagged behind their receptive. However, receptive and productive collocational knowledge did not increase at the same rate over the study period. While learners' receptive collocational knowledge did not show an increase in knowledge, their productive knowledge increased significantly over the school year. The results also revealed that grammatical collocations were less challenging than lexical collocations at this level of language learning. Finally, according to the study results, some

pedagogical implications and suggestions for further studies are presented.

Acknowledgements

I would like to thank all those contributed to accomplishing this study. Special thanks to my supervisors Dr. Phillip Durrant and Dr. Esmael Abdollahzadeh for their help, understanding, patience, and comments. I would like to also express my sincere gratitude to the University of Exeter in general and Alison Stamps in particular who sympathised with me and supported me financially while I was going through a severe financial crisis in my third and fourth years of the study and without their help, I would not be able to complete my study. I would also like to thank the library staff of the University of Exeter, the staffs and students of the high schools who participated in the study, and General Directorate of Education in Duhok-Kurdistan for their patience and support to the study. I would like to also thank those who participated in the pilot studies and collocations extraction process, natives and non-natives of English and Kurdish language teachers. I would like to also thank the examiners of this dissertation for the time and effort exerted in examining and evaluating the content of the study. Finally, I would like to thank my family for all the suffering and difficulties they faced through the years of the current study.

Table of contents

<u>Contents</u>	<u>Page</u>
Abstract.....	2
Acknowledgements.....	4
Table of contents.....	5
List of tables.....	11
List of symbols.....	12
List of abbreviations.....	12
Chapter one: Introduction.....	14
1.1 The research problem.....	14
1.2 The research study.....	16
Chapter two: Collocations.....	18
2.1 Introduction.....	18
2.2 Definition of collocations.....	18
2.2.1 The frequency-based definitions.....	19
2.2.2 The phraseological definitions.....	25
2.2.3 The rise of reconciling definitions: The phrasal frequency-based definitions.....	29
2.3 The importance of collocations.....	31
2.3.1 The importance of collocations for native-like production.....	31
2.3.2 The role of collocations in supporting cognitive fluency.....	33
2.3.3 The importance of collocations for pedagogy.....	34
2.3.4 The importance of collocations for linguistic applications.....	36
2.3.5 The theoretical linguistic importance of collocations.....	37
2.3.6 The role of collocations in meaning.....	39
2.4 Classifications of collocations.....	40
2.4.1 Introduction.....	40
2.4.2 Types of collocations.....	40
2.4.3 Patterns of collocations.....	41
2.4.4 The importance of classifying collocations.....	42
2.5 Collocational knowledge.....	43
2.5.1 Introduction.....	43
2.5.2 Dimensions of collocational knowledge.....	44
2.5.2.1 Receptive knowledge of collocations.....	47

2.5.2.2 Productive knowledge of collocations.....	47
2.5.3 How much collocational knowledge is needed?.....	49
2.6 The linear and nonlinear view of language knowledge and development.....	51
2.6.1 Linear perspective of language knowledge and development.....	52
2.6.2 The nonlinear view of language knowledge and development	54
2.6.2.1 The Dynamic Systems Theory.....	54
2.6.2.2. The dynamic view of language.....	57
2.7 Conclusion.....	65
Chapter three: Factors affecting collocational knowledge and development.....	67
3.1 Introduction.....	67
3.2 Learners' characteristics.....	67
3.3 Contextual factors	70
3.4 Features of collocations.....	76
3.4.1 The influence of frequency.....	77
3.4.2 The influence of the syntactic structure.....	84
3.4.3 The influence of the semantic features.....	89
3.4.4 The influence of congruency with L1	96
3.4.5 The influence of the phonological features	101
3.5 The current model of collocational knowledge and development.....	103
3.6 Concluding and identifying the gap.....	111
3.7 The current study framework.....	112
Chapter four: Methodology.....	115
4.1 Introduction.....	115
4.2 The longitudinal design of the study.....	117
4.3 Procedure.....	120
4.4 Design.....	124
4.5 Participants.....	125
4.6 Data collection.....	126
4.6.1 Introduction.....	126
4.6.2 Curricula textbooks.....	126
4.6.3 Collocation extraction and selection procedure.....	128
4.6.3.1 The operationalisation of the frequency criterion.....	128
4.6.3.2 The operationalisation of the syntactic criterion.....	136

4.6.3.3 The operationalisation of the combinability criterion.....	137
4.6.3.4 The operationalisation of the native intuition criterion.....	138
4.6.4 Classifying collocations.....	139
4.6.5 The receptive knowledge tests.....	143
4.6.6 The productive knowledge tests.....	145
4.6.7 Collocation types of the tests.....	145
4.7 Pilot tests.....	148
4.7.1 Native pilot tests.....	149
4.7.2 Non-native pilot tests.....	149
4.8 Coding and scoring the tests.....	150
4.8.1 Coding and scoring the receptive tests.....	150
4.8.2 Coding and scoring the productive tests.....	150
4.9 Item analysis.....	151
4.9.1 Item analysis-receptive test 1.....	151
4.9.2 Item analysis-receptive test 2.....	152
4.9.3 Item analysis-productive test 1.....	152
4.9.4 Item analysis-productive test 2.....	153
4.10 Reliability of the constructs.....	153
4.10.1 Reliability of the receptive test 1.....	153
4.10.2 Reliability of the receptive test 2.....	154
4.10.3 Reliability of the productive test 1.....	154
4.10.4 Reliability of the productive test 2.....	154
4.11 Waves of data collection.....	157
4.12 Ethics (UK and Kurdistan).....	157
Chapter five: Data analyses and results.....	160
5.1 First wave data analyses.....	160
5.1.1 Participants' characteristics.....	160
5.1.2 Exploring relationships.....	160
5.1.2.1 Receptive test 1 analyses.....	161
5.1.2.1.1 Gender.....	161
5.1.2.1.2 Frequency.....	161
5.1.2.1.3 Syntactic structure.....	162
5.1.2.1.4 Semantic transparency.....	163
5.1.2.1.5 L1 congruency.....	164

5.1.2.2 Productive test 1 analyses.....	164
5.1.2.2.1 Gender.....	164
5.1.2.2.2 Frequency.....	164
5.1.2.2.3 Syntactic structure.....	165
5.1.2.2.4 Semantic transparency.....	165
5.1.2.2.5 L1 congruency.....	166
5.2 Second wave data analyses.....	166
5.2.1 Receptive test 2 analyses.....	166
5.2.1.1 Gender.....	166
5.2.1.2 Frequency.....	167
5.2.1.3 Syntactic structure.....	167
5.2.1.4 Semantic transparency.....	168
5.2.1.5 L1 congruency.....	168
5.2.2 Productive test 2 analyses.....	169
5.2.2.1 Gender.....	169
5.2.2.2 Frequency.....	170
5.2.2.3 Syntactic structure.....	170
5.2.2.4 Semantic transparency.....	170
5.2.2.5 L1 congruency.....	171
5.3 Summary.....	171
5.4 Collocational knowledge development.....	172
5.5 Findings summary.....	184
5.5.1 The influence of gender on the participants' collocational knowledge and development.....	184
5.5.2 The influence of frequency on the participants' collocational knowledge and development.....	185
5.5.3 The influence of the syntactic structure on the participants' collocational knowledge and development.....	185
5.5.4 The influence of the semantic transparency on the participants' collocational knowledge and development.....	186
5.5.5 The influence of the L1 congruency on the participants' collocational knowledge and development.....	187
5.5.6 The ordinal influence of features of collocations on the participants' collocational knowledge and development.....	187

5.5.7 How does participants' collocational knowledge develop over a school year?.....	187
5.5.8 Influence of the test designs on the data quality.....	188
Chapter six: Discussion of the findings.....	189
6.1 The influence of gender on the participants' collocational knowledge and development.....	190
6.2 The influence of frequency on the participants' collocational knowledge and development.....	192
6.3 The influence of the syntactic structure on the participants' collocational knowledge and development.....	195
6.4 The influence of the semantic transparency on the participants' collocational knowledge and development.....	199
6.5 The influence of the L1 congruency on the participants' collocational knowledge and development.....	202
6.6 The ordinal influence of the collocational features on the participants' collocational knowledge and development.....	205
6.7 Participants' collocational knowledge and development.....	206
6.8 Influence of the test designs on the data quality.....	213
Chapter seven: Conclusions.....	214
7.1 Summary.....	214
7.2 Contributions of the study.....	216
7.3 Pedagogical implications.....	219
7.4 Limitations of the study.....	221
7.5 Directions for further research.....	223
Appendices.....	226
Appendix (A): Ethical Approval Certificate and Consent Form.....	226
Appendix (B): Test items list.....	234
Appendix (C): Classifying collocations.....	239
Appendix (D): The receptive tests.....	242
Appendix (E): The productive tests.....	244
Appendix (F): Characteristics of the test items.....	248
Appendix (G): Item analysis-receptive 1.....	250
Appendix (H): Item analysis-receptive 2.....	252
Appendix (I): Item analysis- productive 1.....	254

Appendix (J): Item analysis-productive 2.....	256
Appendix (K): Item-total statistics- receptive test 1.....	258
Appendix (L): Receptive test 1 item-total statistics after deleting four items.	260
Appendix (M): Receptive test 2 item-total statistics.....	262
Appendix (N): Item total statistics-productive test1.....	264
Appendix (O): Productive test 2 item total statistics.....	266
References.....	268

List of tables

Table (1): Statistics of the textbooks.....	126
Table (2): BBI's and Gitsaki's collocation types and extracted items.....	135
Table (3): Balance among the items.....	145
Table (4): Item analysis-receptive test 1 statistics.....	150
Table (5): Receptive test 2 descriptive Statistics.....	151
Table (6): Item analysis-productive test 1 statistics.....	151
Table (7): Item analysis-productive test 2 statistics.....	152
Table (8): Receptive test 1- gender group statistics.....	159
Table (9): Receptive test 1- syntactic structure statistics.....	160
Table (10): Receptive test 1- syntactic structure post-hoc Games-Howell....	160
Table (11): Productive test 1- gender group statistics.....	162
Table (12): Productive test 1- syntactic structure of collocations.....	163
Table (13): Productive test 1- syntactic structure of collocations descriptives.	163
Table (14): Receptive test 2- gender group statistics.....	164
Table (15): Receptive test 2- syntactic structures statistics.....	165
Table (16): Receptive test 2- syntactic structures ANOVA.....	166
Table (17): Productive test 2- gender group statistics.....	167
Table (18): Productive test 2- syntactic structure of collocations descriptives.	167
Table (19): Productive test 2- syntactic structure of collocations ANOVA.....	168
Table (20): Collocational knowledge analyses and results summary.....	169
Table (21): Collocational knowledge development results summary.....	171
Table (22): Collocational Knowledge of the contrasted variables.....	177

List of symbols

?: inappropriate use

italic: highlighted

+: has the feature, or increase

-: lack the feature, or decrease

List of abbreviations

adj: adjective

adv: adverb

ANOVA: analysis of variance

AWL: Academic Word List

BBI: Benson, Benson & Ilson

BERA: British Education Research Association

BNC: The British National Corpus

CITC: Corrected Item-Total Correlation

CLAWS: The Constituent Likelihood Automatic Word-tagging System

COCA: Corpus of Contemporary American English

COLLMATCH: collocate matching

COLLEX: collocating lexis

det: determiner

df: degree of freedom

DST: The Dynamic Systems Theory

e.g.: exempli gratia, for example

EAP: English for Academic Purposes

EFL: English as a Foreign Language

ESL: English as a Second Language

etc.: et cetera, and so forth

F: F-test (ANOVA)

fv: facility value

i.e.: id est, that is

inf: infinitive

KSA: Kingdom of Saudi Arabia

L1: First Language

L2: Second Language

LOCNESS: Louvain Corpus of Native English Essays

M= mean
MCQ: multiple choice questions
MD: medical
MI: Mutual Information
n: noun
N: number
NLP: Natural Language Processing
o: object
p: p-value (calculated probability)
prep: preposition
pron: pronoun
r: correlation coefficient
s: subject
SD: standard deviation
SE: standard error
Sig.: significance
SMS: Short Message Service
SPSS: Statistical Package for the Social Sciences
Std: standard
t: t-test (pared-samples, independent-samples)
T: time
TOEFL: Test of English as a Foreign Language
UG: Universal Grammar
UK: United Kingdom
UWL: University Word List
v: verb
vs: versus

Chapter one

Introduction

1.1 The research problem

Learning a foreign language in the contemporary societies has emerged to become one of the main necessities. In addition, the need for effective communication has increased due to the technological advances in every facet of the modern human life; in politics, economics, sciences, and education. Internationally, English language has risen as one of the main active languages which is widely used as a means of communicating and conveying information among nations. Accordingly, research aiming at developing and accelerating learning English language has increased rapidly. Research in this field has revealed that one of the major challenges ahead of English language learners is producing proper collocations. In spite of learners' memorising of thousands of words, they seem to be inadequately able to produce correct collocations and this is a challenge even for those learners who are at advanced levels of language proficiency (Bahns, 1993; Nesselhauf, 2003).

Significance of collocations in developing learners' language proficiency has been increasingly underscored in the last three decades (e.g., Fan, 2009; González Fernández & Schmitt, 2015; Nation, 1990; Schmitt, 1997). However, demarcating learners' collocational knowledge and showing their knowledge development patterns need research into the role of features of collocations, such as frequency of occurrence, syntactic structure, semantic transparency, and congruency with learners' first language, in the learners' collocational knowledge and development. Features of collocations have been found as one of the key factors affecting learners' collocational knowledge development. The influence of features of collocations on learners' collocational knowledge is well-established in the literature of the field (e.g., Gitsaki, 1999; Kellerman, 1978; Koya, 2005; Kurosaki, 2012; Nesselhauf, 2005). However, studies have revealed that these features vary in their degrees of influence on the learners' collocational knowledge and development. Moreover, studies in this field have not yielded a consensus on how these features affect learners' collocational knowledge and development. Furthermore, some studies have shown even contradictory results. Accordingly, this situation requires more investigations into how this influence outlines or directs the general collocational knowledge

and development of the second language (L2 henceforth) learners and which features are more influential than the others.

Exploring the influence of features of collocations on knowledge and development has been conducted from various perspectives and in different study designs. Studies have generally focused on frequency of occurrence, syntactic structure, semantic transparency, and congruency with learners' first language as the main influential features on learners' knowledge and development (e.g., Dörnyei, Durow, & Zahran, 2004; Gitsaki, 1999; Fitzpatrick, 2012; Kurosaki, 2012; Nesselhauf, 2003, 2005; Zheng, 2010). However, what can be noticed on these studies is the snap-shot nature of most of the previous studies. The investigations have been short-term or cross-sectional in which data were gathered at certain points of time and were unable to follow up learners' knowledge and development of collocations over a long period of time. In this concern, Ortega & Byrnes (2008) argue that since L2 learning is a long-term process, the best way to gain better understandings and insights into collocational knowledge progress and patterns of ability development of this linguistic phenomenon is by conducting longitudinal studies. Accordingly, the present study investigated the impacts of four key features of collocations- frequency of occurrence, syntactic structure, semantic transparency, and congruency with learners' first language- on the learners' collocational knowledge and development in a longitudinal design.

The significance of investigating collocation features and their relations with L2 learners' collocational knowledge and development can be realised in various aspects. First, its importance lies in the potential roles that these features can play in explaining part of learners' errors and weaknesses in producing proper collocations, for example, knowledge deficiency in the grammar and the syntactic structures of the learned language can result in producing incorrect combinations, consequently, improper collocations (Fan, 2009). Secondly, some features, even at advanced levels of second language use, may affect not only correctness at the syntactic or semantic level, but also at the style and register level of a text, for example, there are certain collocations which are used exclusively in specific contexts such as those within media contexts, e. g., *press release* (Leśniewska, 2006). Third, being aware of the collocational features and their impacts on the learners' knowledge and development can

raise second language teachers' and curriculum designers' awareness of the different types of challenges and learners' collocation errors and help them decide on how to incorporate them into curricula textbooks at the various levels of language studying in schools and language learning institutes to promote learners' collocational knowledge and use (Ha, 2013). Accordingly, understanding collocation features from its various facets may aid accelerating L2 learners' collocational knowledge and development to produce native-like language.

Finally, locally, i.e., in Kurdistan Region, despite the evidenced significance of collocations to learners' language proficiency development, the topic has rarely and very limitedly been investigated and it has never, to my knowledge, been conducted on high school students in this depth, level, and time span. Moreover, the explicit teaching of this linguistic phenomenon seems pedagogically neglected in most of the schools in the region.

1.2 The research study

The current research is an investigation into the influence of four features of collocations- frequency, syntactic structure, semantic transparency, and congruency with learners' mother language- on the learners' collocational knowledge and development. The underlying premise of this study is that more investigations into this aspect of collocations in different contexts can explain and reveal more information about learners' collocational knowledge and the influence of features of collocations on their knowledge development. As previous investigations into learners' collocational knowledge and influence of features of collocations on their knowledge development have resulted in rather contradictory findings, the present study aims to provide more information to disambiguate the situation by exploring Kurdish high school students' collocational knowledge.

Although the study was carried out in Kurdistan Region, it has its significance to the neighbouring countries. Its significance to the neighbouring countries can be related to the that these countries, e.g., Iraq, Turkey, Iran, and Syria, have nearly similar educational contexts. Accordingly, what applies to Kurdish learners may generally apply to L2 learners in the neighbouring countries as well. Finally, in a wider sense, the study findings can be of significance to the international educational community since the study findings can be related to

L2 learners in other countries as L2 learners generally share some common characteristics as it is evidenced in various studies.

Chapter two

Collocations

2.1 Introduction

Collocations have been investigated thoroughly in the previous studies. The conducted studies have mainly dealt with collocations in terms of their definitions, types, acquisition, teaching, and significance to the L2 learners' competence. However, these studies have been carried out adopting various views of collocations. Consequently, different definitions of collocations came to existence and various approaches were adopted and suggested for a better understanding of this linguistic phenomenon.

This chapter reviews the key studies that explored and examined collocations in terms of definitions, importance, features, types, patterns, knowledge dimensions, and linear and non-linear views of language knowledge and development.

2.2 Definition of collocations

Terminologically, it has been argued that it is merely a terminological issue to classify collocations as a separate combination type or as a sub-division of idiomatic expressions (Bolinger, 1979). This is because using terms such as "combination" or "co-occurrence" can encompass the same range of linguistic area, accordingly, it is not obvious what else more can be obtained from using the term "collocation" rather than using "combination" or "co-occurrence" as substitute terms (Granger & Paquot, 2008; Schmid, 2003). Accordingly, they claim that using a separate term for this type of combination is unwarranted since it does not offer extra information that distinguish collocations as a separate group from the other types of combinations. However, I would argue that the use of the term "collocation" is warranted and worth using since it saves effort and time for research purposes. Additionally, it is evidenced in various studies, as it will be discussed later in this chapter, that characteristics of the combinations encompassed under the umbrella of "collocations" are different from those of the other types of combinations. Accordingly, the term "collocation" is used throughout the current study as a distinctive linguistic phenomenon.

Historically, the concept of “collocation” as a technical term in linguistics is credited to the contributions of Palmer (1966), who is regarded as a pioneer in studying and classifying collocations, and Firth (1957), who is regarded as the father of collocation, (Bartsch, 2004; Xiao & McEnery, 2006). Palmer (1966) was mainly interested in restricted collocations in which an element is utilised in a specialised sense that occurs exclusively in combination with the other elements (Pawley, 2007). Palmer (1966, p.5) defines collocations as “a succession of two or more words that must be learned as an integral whole and not pieced together from its component parts”. Various definitions of the term “collocation” have been presented since its appearance as a linguistic concept. However, there seems to be no general agreement on a specific definition of collocations (Granger & Paquot, 2008). The diversity in defining collocations is attributed to various reasons such as the complex nature of collocations, various perspectives from which this linguistic phenomenon has been viewed, and the different criteria used for classifying collocations (Pearce, 2002). However, collocation definitions can be grouped into two distinctive, and at times, overlapping orientations. The first view of collocations is called the frequency-based approach (Nesselhauf, 2004), statistical approach, distributional approach (Evert, 2005), or quantitative approach. The second view of collocation is called the phraseological approach (Nesselhauf, 2004). However, a third view of collocation seems to be under establishment which can be called the hybrid or the phrasal frequency-based approach (e.g., Altemberg & Granger, 2001; Kurosaki, 2012; Nizonkiza, van Dyk, & Louw, 2013). The third view is a reconciling approach that combines both perspectives in one pot. These views of collocations will be explained in detail in the following sections.

2.2.1 The frequency-based definitions

A number of features are attributed to collocations which have been proved to be useful in defining and identifying collocations. However, considering these features resulted in suggesting various definitions and investigating designs of collocational knowledge and development. Though, due to the complex nature of collocations, it might not be sufficient to define and explore the various aspects of collocations relying solely on one feature or from one perspective (Koya, 2005). Accordingly, describing collocations from different angles seems

to be an inevitable consequence. Among these features, frequency is one of the main properties employed in defining and identifying collocations.

Collocations are generally characterised by their recurrent appearance in spoken and written texts. Frequency of use is considered as an important indicator of the behaviour characteristic of a word association, for example, the word *love* is strongly associated with the words *affair* and *in* to the extent the appearance of the word *affair* or *in* in a text gives a high probability of being preceded in the former word and followed in the latter by the word *love* (Brezina, McEnery, & Wattam, 2015). Accordingly, these chunks *love affair* and *in love* are important chunks in the English language. The habitual recurrent co-occurrence of certain expressions provides evidence for the existence of such collocations (Gledhill, 2000).

The definition which is established on the frequency of co-occurrence of the constituent parts as the underlying assumption is called the frequency-based definition of collocations. Among the key representatives who have participated in establishing and developing the frequency-based view of collocations are Firth (1957), Sinclair (1966, 1987, 1991), Halliday (1966), Greenbaum (1974), Pawley & Syder (1983), Hoey (1991, 2003, 2005), Clear (1993, 2005), and Stubbs (1995, 2001). According to this view, the term “collocation” generally refers to the habitual co-occurrence of patterns of words (Xiao & McEnery, 2006). The syntactic relationships between the lexical elements do not have a part in deciding whether a certain combination is a collocation or not (Nesselhauf, 2005).

A distinction is usually made between a statistically significant co-occurrence from a co-occurrence that happens by chance using statistical association measures such as frequency, log-likelihood (G^2) MI (mutual information), and t-scores (t) (Bartsch & Evert, 2014). However, within this view, researchers varied in their reliance on the raw frequency of co-occurrence criterion. According to Moon (1998), the term “collocation” encompasses the co-occurrences of all frequencies as collocations. Differently, Stubbs (2001) limits the term “collocation” to only those of frequent co-occurrences while Kjellmer (1982) counts combinations of recurrent co-occurrences, i.e., that occur more than once, in a given corpus as collocations. The focus in this view is on collocation

as a linguistic phenomenon and the role it plays in the linguistic meaning constitution (Howarth, 1996).

Firth (1957, p. 179) contends that a word can be known by “the company it keeps” and that collocation adds to the meanings of the individual words that form the collocation. Firth (1957, p. 181) defines collocations as “statements of the habitual places of that word”. In other words, the habitual and recurrent association of semantically related words (Bartsch & Evert, 2014). Accordingly, the concept of collocation is that "meaning by collocation is an abstraction at the syntagmatic level" (Firth, 1957, p.193), i.e., the sequential relationship between two or more items in a sentence boundary. In the linear relationship between the component items of the collocation *dark night*, for example, one of the meanings of *night* is its collocability with *dark* and one of the meanings of *dark* is its collocability with *night*. According to Firth, collocation is a separate level of meaning.

After Firth’s definition of collocations, further definitions have been introduced. However, these definitions seem to be merely rewording of the same idea (Jabir, 2011) with regarding or disregarding certain features of collocations such as the syntactic relations of the constituent items. While Firth’s (1957) quantitative definition of collocations relies on the raw frequency of semantically related items, Halliday (1966) takes frequency of co-occurrence and the syntactic relations of the collocational components as bases for his definition. Halliday (1966) appears to be mainly interested in collocations and the way lexical sets membership can be defined in the framework of collocation. According to Halliday, collocability refers to the syntagmatic relationships or the sequential co-occurrence of lexical items at certain distance from each other. In addition, Halliday claims that the probability of the co-occurrence of a collocation component is greater than that chance could predict. Accordingly, frequency of occurrence is regarded as an important element in defining collocations. However, he takes two elements in defining collocations: frequency of co-occurrence and the syntactic relations of the collocational components.

Léon (2007) argues that Halliday adds a paradigmatic dimension to the syntagmatic dimension proposed by Firth in his definition of collocations since he suggests investigating lexical patterning in the light of lexico-grammar. The

additional dimension, to Léon, is evidenced in introducing a probabilistic turn, which is seen as an essential element for computational corpora studies. Accordingly, Halliday has reinterpreted the characteristic of mutual expectancy suggested by Firth's definition of collocation in terms of lexical patterning, where the inclination of words to co-occur is not predicted by chance.

However, Halliday (1966, p. 150) argues that grammar is unable to explain the reasons behind all collocations such as why *strong tea* and *powerful car* are acceptable collocations whereas *?strong car* and *?powerful tea* are not proper collocations. This indicates that both collocability and the grammatical structure rely on the syntagmatic relations between lexical sets components but collocability relations are independent from the grammatical structures of these sets items.

Another definition is introduced by Greenbaum (1974, p. 82) who defines collocations as "a frequent co-occurrence of two lexical items in the language". Although Greenbaum accepts the statistical definition of collocations, he considers a co-occurrence as a collocation only if the collocation is statistically significant (Xiao & McEnery, 2006).

With a little modification in the Firth's use of the term "collocation", Hoey (1991) uses the term "collocation" only if a lexical item occurs with other items in a textual context more than a chance can predict. Hoey (1991, pp. 6-7) defines collocations as lexical items that co-occur with other items "with greater than random probability in its context". This is a textual definition of collocations and it implies that the constituent parts of collocations predict each other in that the existence of a constituent in a textual context makes the appearance of the other more probable than chance would predict in other contexts (Durrant & Doherty, 2010).

Adding a space criteria to the definition, Sinclair (1991, p. 170) defines collocations as "the occurrence of two words or more within a short space of each other in a text". Additionally, he argues that words generally do not carry their meanings independently from each other. Rather, according to him, the individual words meanings comprise merely parts of the whole meanings expressed through the word combinations that structure the text. In other words, words meanings are determined according to the textual context or environment in which they are utilised (Sinclair, 1996). For example, the collocation *top*

drawer can mean “*the uppermost drawer in a cabinet*” or “*something that is the best of its class*” as in the sentence: *He bought a top drawer car* (Macis & Schmitt, 2017, p. 51). This definition emphasises the space between the component parts and the lexical text in which the collocation is used.

Stubbs (2001, p. 29) describes collocations as "frequent co-occurrence" of lexical items and that collocations are types of syntagmatic relations between these items. Additionally, he argues that the co-occurrence relations between collocational items are constructed in a linear cluster. In other words, in a collocation, the appearance of a node implies the existence of the collocate before or after it. In accordance, this necessitates investigating words in collocations rather than studying them individually.

Briefly put, collocation within Firth's and his proponents' view consists of several elements. The first element is co-occurrence of the items, i.e., words go together. The second element is the recurrence of the combination, i.e., a collocation is a combination which frequently occurs in a language more than chance can predict. Space limit between the node and its collocate is the third element. The distance between the node and its collocate should not exceed few words on either side of the node. The component words of a collocation may occur next to each other without intervening words between nodes and collocates such as *heavy rain*, or they may occur at a distance from each other intervened by few words, as with more flexible collocations such as *the strength of his argument* (Manning & Schütze, 1999). The fourth element is that meaning of words is mostly perceived by the accompanied words. This element highlights the lexical context role in determining meaning of words. Finally, although the notion of collocation according to this view may imply a grammatical relation between collocation constituents (Bartsch, 2004) and that some proponents, such as Halliday (1966), attempted to explain collocability based on the grammatical relations of the lexical component items of collocations, the syntactic relations of the collocational constituents do not seem to have a key role in defining collocations according to this approach.

The frequency-based definition of collocations is found to be mostly objective since it is based on a real existence in a context, frequent occurrences, and collocational span. However, Henriksen (2013) argues that disregarding the semantic aspects from the frequency-based approach analysis can result in

identifying recurring lexical combinations which have little psycholinguistic validity such as *and the* as collocations. This occurs due to depending solely on frequency of occurrence of certain lexical chunks which co-occur frequently but they are not perceived as collocations by native speakers. However, the frequency-based approach has well-known tools for overcoming this problem such as the use of association measures; manual filtering rather than depending on raw frequency of occurrence (Simpson-Vlach & Ellis, 2010). Frequency-based studies identify collocations relying mainly on the frequent co-occurrence of their component items in large corpora.

Furthermore, despite the unquestionable importance of frequency of co-occurrence in recognising collocations, some researchers (e.g., Gavioli & Aston, 2001; Kjellmer, 1984) see that research should not depend on this feature only as some other features exist which can also be employed in defining collocations such as semantic transparency and syntactic features. Moreover, they argue that this feature, frequency of occurrence, is deduced from corpora that are thought to be insufficient to mirror the whole real experience of language users. Accordingly, for a reliable frequency, corpora should be sufficiently large to represent the target language and to reflect the real experience of L2 learners (Trofimovich, 2011). However, it is believed that no corpus in whatever size can reflect the whole real experience of the target language and that each corpus is usually examined for identifying specific features depending on the study aims (Gavioli & Aston, 2001). Moreover, compiling a corpus that is representative of all language users will not be workable due to the impossibility of undertaking this in practice in addition to that each language has various dialects and varieties which makes deciding on which variety to be based is far from being easy or achieved (Gabrielatos, 2005).

Another limitation of this feature is that it includes all possibilities into account which results in a huge number of combinations of which some may be syntactically unrelated to each other (Seretan, Nerima, & Wehrli, 2003). In consequence, this may result in blurring the boundaries among the different types of combinations and makes it difficult to distinguish among them to the extent considering a structure like (v+ the) as a collocation which is not.

In conclusion, despite the rather fuzziness and issues intervening the application of frequency, it remains one of the essential key features in defining collocation. However, it seems that it is not sufficient to rely solely on frequency as a decisive feature in collocational linguistic investigations and analyses, except where the study purposes require employing this feature alone.

2.2.2 The phraseological definitions

Some researchers (e.g., Cowie, 1981, 1994, 1998; Gitsaki, 1996; Howarth, 1996, 1998; Nesselhauf, 2003, 2005), who are highly influenced by the Russian phraseology, view collocations as being constrained by specific rules governing their structures. Accordingly, they define collocations based on typological grounds and take the grammatical structure, semantic transparency, and substitutability of the component items as guiding principles (Gyllstad & Wolter, 2015). The underlying assumption of this view assumes a direct syntactic relationship between the component items of collocations. According to this tradition, collocations are regarded as a type of word combinations that occupies the space between idioms and free word combinations. Idiomatic expressions such as *to kick the bucket* differ from collocations in that the whole meaning of the expression cannot be inferred from the semantics of its individual components. Free word combinations can be distinguished from collocations in that constituents can be freely substituted by others without seriously affecting the overall meaning of the combination (McKeown & Radev, 2000).

Howarth (1996), driven by degree of restriction, semantic transparency, and substitutability of the items, suggests four combination types. The first type is called free combinations in which the constituent elements are used in their literal meanings and elements can be substituted with items from the same semantic field such as *drink tea*. The second type of combinations is called restricted collocations where one element is employed in its literal sense while the other has a specialised or figurative sense function. In this type of combinations, substituting elements is possible but it is limited such as *perform a task*. The third combination type is classified as figurative idioms that have figurative reference with a very little degree of commutability which is characterised by its flexibility of substituting one of the elements by another, and the literal meanings of the items are maintained such as *do a U-turn*. The fourth

type of combinations is called pure idioms in which the overall semantics cannot be retrieved from the literal meanings of their elements and this type of combinations has no space for substituting any items in the juxtaposition such as *blow the gaff*.

Cowie (1998, p. 191) argues that defining collocations as “groups of words which frequently occur in combination with each other” is completely an ambiguous and an unsatisfactory definition. He explains that this is because this definition tells nothing about number of elements in a collocation, degree of frequency of occurrence, or classes of words which can combine as collocations. Cowie (1994) makes a distinction among four sorts of combinations. Though the distinction does not outline clear-cut boundaries among these combinations as they overlap into each other’s domains. According to him, degrees of opacity and fixedness of the four groups of combinations go on a continuum line in which the opaquest and most fixed combinations are placed at one end and the most transparent and variable ones at the other. In accordance, this continuum consists, from the least to the most opacity and fixedness, of free combinations, restricted collocations, figurative idioms, and pure idioms. Although idioms can also be defined as “groups of words which frequently occur in combination with each other”, he agrees that collocations are different from idioms (Cowie, 1998). Elements of free combinations are characterised by being transparent in meaning and freely combined, i.e., substitutable by other items. However, items of figurative idioms and restricted collocations are limitedly substitutable. Finally, pure idioms elements are restricted and are not substitutable without affecting the whole meaning of the idiom (Cowie, 1981). His classification of word combinations is mainly based on two criteria: ‘commutability’, i.e., the substitutability of the combination components without affecting the semantics of the whole combination, and the opacity or ‘transparency’ of the combination semantics, i.e., whether the lexical items are used in their literal or figurative meanings.

Nesselhauf (2005) sets three criteria on a combination constituents to test their collocability: syntactic relations; semantic transparency, and substitutability. Syntactic relations are exemplified in the relationships which exist between the combination items such as v+n, adj+n, adv+adj, and v+adv relations. Semantic transparency means that meanings of collocations can commonly be retrieved

from the individual meanings of their constituent words. Finally, substitutability indicates a degree of substitutability of their constituents with similar items in the same semantic field. Accordingly, for a combination to be accounted as a collocation, first, the component items of the combination should fall in one of the syntactic relations such as the v+n relation, i.e., there should be a grammatical relation between the items. If there is no syntactic relation between the constituents as in *study the*, this criterion will not be met and the combination will not be counted as a collocation. Second, semantic transparency implies that the meaning by a collocation should be generally perceived from the constituent parts of the combination, e.g., in the collocation *make a decision*, the general meaning of the collocation can be retrieved from the meanings of *make* and *decision* though the word *make* is not used completely in its literal sense. Third, substitutability allows a limited degree of freedom to substitute a word by another in the same word class such as *take* in place of *make* in *make a decision*, i.e., *take a decision*. The application of the three criteria on the collocability of the component items of a v+n combination can be explained in the combination *draw a conclusion* as an example. Syntactically, *draw a conclusion* is an acceptable v+n combination since the relationship between the items is a v+direct object relationship, i.e., v+n relationship. Semantically, the general meaning of this combination can be deduced from the individual combination items. As for the substitutability criterion, the verb *draw* is substitutable by *reach* to produce *reach a conclusion*, but it cannot be replaced by the verb *produce* to yield *?produce a conclusion* which is unacceptable to native speakers. This shows that one of the combination constituents is limitedly substitutable, which is one of the characteristics of collocations. Accordingly, this combination can be regarded as a collocation based on the three criteria.

The importance of the syntactic feature is not seen only in defining collocations but also in patterning collocations. Hausmann (1984), for instance, maintains that the syntactic category of the component items of collocations is one of the important criteria which can be relied on in classifying collocations. According to him, collocations have a hierarchical structure in which one word, i.e., the node, dominates over the other and determines its choice, i.e., the collocates. Additionally, the grammaticality of the collocational component items can facilitate the direct collocation extraction from the source corpus (Evert, 2004),

and it may help in studying, learning, and teaching collocations. It can also help in identifying and distinguishing the different types of the multi-word combinations (Seretan, Nerima, & Wehrli, 2003) and collocations in a text. The importance of this feature lies also in its use in predicting meaning of the lexical items. For example, Renouf & Sinclair (1991) state that the existence of a grammatical item and its order in an expression can help in guessing the meaning of a lexical item, for example, in an expression such as *an X of as in some of*, *some* here is usually a quantity (Renouf & Sinclair, 1991).

In sum, the phrasal definition of collocations consists of several elements. The first element is the syntactic structure of collocations. Only specific grammatical categories which hold particular syntactic relations are considered collocations. Semantic transparency, which entails that the meaning by a collocation should be perceived from the constituent parts of the combination, is the second element. However, one of the collocational constituents may be used in its figurative meaning. The semantic transparency of the combination should occupy the space between completely transparent and entirely opaque ones. Commutability or substitutability is the third element. Substituting collocational constituents are restrictedly possible without affecting the whole meaning of the combination. Degree of restrictedness occupies the distance between free combinations and pure idioms.

However, Hausmann (1989) believes that the syntactic criterion alone is not adequately an effective tool to differentiate between the different types of combinations and that adding other criteria such as a semantic criterion is essential for this purpose. In line with, Gledhill (2000) contends that the inadequacy of the syntactic criterion can be attributed to that the syntactic criterion mainly deals with the abstract aspects of collocations. Hence, for a better depiction of this linguistic phenomenon, involving other criteria such as frequency of occurrence, which deals with the concrete aspects of collocations, sounds necessary for providing tangible evidence from the real-life context, i.e., spoken and written texts.

Furthermore, the phraseological approach is thought to be subjective because it takes the syntactic relations between the constituents into account (Henriksen, 2013). Accordingly, the decision about the collocability of a combination or syntactic structure is made based on the intuition of the researcher and it

disregards the actual occurrence of collocations and their frequency in real contexts use.

Moreover, the phraseological approach has been criticised for that it cannot tell the actual frequency of use of collocations, consequently, it may result in incorporating pedagogically inappropriate collocations into learning materials such as *judicial organ* or *ruggedly handsome* which are collocations of very low frequency that learners rarely face in their academic life (Henriksen, 2013).

In conclusion, regardless the evidenced significance of the syntactic criterion in the phraseological approach in defining, patterning, and discriminating collocations from the other types of combinations. The general view points to the notion that this feature is not solely sufficient for encompassing the key characteristics of this linguistic phenomenon, i.e. collocations unless used for specific research purposes. Accordingly, incorporating the other key feature in the definition, i.e., frequency, may offer a more comprehensive view of collocations.

2.2.3 The rise of reconciling definitions: The phrasal frequency-based definitions

It can be inferred from the definitions of the two previous approaches that the frequency-based view of collocations relies mainly on the textual co-occurrence of certain combinations, and observing their frequency of occurrence as the main guiding principle. This perspective puts emphasis on the concrete or physical aspects of collocations, i.e., their actual co-occurrences and recurrences in texts. In contrast, the second tradition defines collocations within the syntactic/semantic frameworks which are concerned more with the abstract relationships between words and the potentiality of lexical combinability excluding frequency and their probability of occurrence from its consideration (Gledhill, 2000; Henriksen, 2013; Nation, 2013). In this approach, collocations are recognised as a type of partly restricted word combinations, and classifying collocations into certain grammatical structures takes the priority in its agenda. Phraseologists have understood collocations as syntactic combinations where they mix lexical and syntactic analyses together, accordingly, they have suggested various syntactic patterns of collocations (Poulsen, 2005). In addition to the syntactic sub-classifications, their definition of collocations is based on motivation, commutability, and selectional restrictions grounds (Schilk, 2011).

However, each of these two approaches has been criticised for neglecting a key feature that is adopted by the other approach, though neither approach denies the existence of the collocational feature adopted by the other approach.

Accordingly, in a way to fill the gap and avoid some of the shortcomings in these two approaches, a third linguistic perspective of collocations is under development. The third, rather new, growing approach, tries to provide a more comprehensive view of collocations by combining both frequency-based and phraseological approaches in one unified approach. This hybrid view, i.e., the phrasal frequency-based approach, attempts to utilise the advantages of both perspectives, i.e., combining the best of the two approaches to provide better insights into the study of phraseology in general and collocations in particular (Granger & Paquot, 2008), through involving the frequency, semantic, and syntactic features altogether in defining collocations (e.g., Alternberg & Granger, 2001; Gyllstad & Wolter, 2015; Kurosaki, 2012; Nizonkiza, van Dyk, & Louw, 2013; Wolter & Gyllstad, 2011).

In brief, the operationalisation of the hybrid approach to identify collocations as suggested by researchers adopting this view can be done in two phases. Firstly, the most frequently occurring combinations in the target corpus are identified in accordance with the frequency-based procedures. Secondly, certain word combinations to be maintained or discarded according to the identified semantic and syntactic criteria (e.g., Henriksen, 2013; Li & Lu, 2011; Todiraşcu et al., 2008).

Proponents of this perspective believe that neither all frequent recurrent combinations nor all syntactic patterns of combinations are collocations. Only combinations of specific grammatical patterns that consist of restrictedly commutable items and frequently co-occur can be considered collocations. This view of collocation seems to have emerged to fill the gap and deficiency in each approach separately on the one hand, and the growing awareness of the role of collocations in the L2 learners' competence development, on the other hand (Barfield & Gyllstad, 2009; Granger & Paquot, 2008; Gyllstad, 2007; Nesselhauf, 2005). However, I would argue that two points need to be borne in mind with adopting the hybrid approach. First, it should not imply abandoning the other two approaches since each perspective seems to have its significance for specific linguistic analyses. For example, the frequency-based approach fits

more in automatic corpus-based methods of extraction whereas the phraseological approach suits better the linguistic analyses studies (Granger & Paquot, 2008). Second, the third perspective should be built on clear rigorous assumptions that prevent adding more fuzziness to the existing situation and so as not be merely combining two different perspectives without obvious boundaries of the approach.

The hybrid perspective tries to reconcile and combine the two perspectives together for a more comprehensive definition of collocations (e.g., Barfield & Gyllstad, 2009; Granger & Paquot, 2008; Gyllstad, 2007; Gyllstad & Wolter, 2015; Nesselhauf, 2005; Nizonkiza, van Dyk, & Louw, 2013). Establishing on this, the definition which can be inferred from this perspective, based on the conducted studies adopting this view of collocation, is that a collocation is “ a recurrent co-occurrence of two or more partially transparent and restrictedly commutable items in a specific syntactic pattern that occurs more than chance can predict”. As the reconciling approach is a blend of the two approaches, frequency-based and phrasal approaches, the underlying criteria of defining collocations are assumed to be frequent co-occurrence of constituents, falling in certain syntactic categories, semantic transparency, and limited degrees of substitutability of the constituents.

In conclusion, it is hoped and believed that the hybrid approach will provide, taking study aims into account, deeper insights into this linguistic phenomenon, i.e., collocations. This expectation is established on the assumption that the hybrid view will gain the advantages of both approaches collectively and refrain their disadvantages separately. The present study adopted this view of collocations to explore learners' collocational knowledge and development as will be explained in more detail in the next chapter (Section 3.7).

2.3 The importance of collocations

The role of collocations can be perceived in various aspects as explained in the following sections.

2.3.1 The importance of collocations for native-like production

Collocations have increasingly been found of an essential significance in developing L2 learners' language competence and as a sign of their language proficiency levels. However, collocations seem to be the most daunting component of vocabulary knowledge even at advanced language proficiency

levels of which L2 learners have been found to be in short of the expectations (Ahmadian & Darabi, 2012; Farrokh, 2012). For example, Palmer (1966) notes that L2 learners may face challenges in producing collocations such as *to ask a question*, *to do a favour* or *to give trouble* and present unacceptable combinations such as *?to make a question*, *?to perform a favour* and *?to do trouble* respectively.

The increasing interest in collocations as a linguistic phenomenon can be attributed to a number of reasons. Among the main factors in this concern is the well-established significance of collocations in the L2 learners' proficiency in both written and spoken modes (Fan, 2009; González Fernández & Schmitt, 2015; Keshvarz & Salimi, 2007; Nation, 1990; Nizonkiza, van Dyk, & Louw, 2013; Schmitt, 1997). In other words, learners' collocational knowledge can play a significant role in improving their proficiency to produce native-like language in both written and spoken forms (Pawley & Syder, 1983).

Another reason for why collocations matter is related to the nature of collocations themselves. Collocations are commonly characterised by being arbitrary, frequent in their co-occurrence and recurrence (Sinclair, 1991). Moreover, collocations are thought to be fixed to an extent (Wu, Witten, & Franken, 2010), limitedly compositional, and sometimes difficult to deduce the whole semantic meanings from their constituent parts (Manning & Schütze, 1999). Consequently, the intricate nature of collocations makes it difficult for non-natives to learn and understand. Accordingly, comprehensive, thorough, and profound investigations into this phenomenon are required by linguists and lexicographers of their various aspects.

Furthermore, due to learners' inadequate knowledge of collocations, Mohajeri & Ketabi (2013) and Mounya (2010) argue that collocational knowledge appears to be an unavoidable aspect of the learned language that needs to be acquired. Additionally, they note that it is one of the significant criteria which are utilised for distinguishing native from non-native speakers of a language. Accordingly, it is an essential step to be taken in order to produce a natural or native-like language.

In addition to the evidenced role of collocations in developing learners' language proficiency, this linguistic phenomenon is seen as an element of the text or discourse coherence and as a distinctive competence marker that

differentiates between native and non-native speakers of the language (Bahns & Eldaw, 1993; Gitsaki, 1999; Hill, 2000; Jabir, 2012; Nesselhauf, 2005; Pawley & Syder, 1983; Stubbs, 2002; Wray, 2002). Additionally, Yule (2003) states that knowing a language implies not only knowing the meaning of the words, but also knowing their conventional patterns of collocations. This makes collocations a distinctive and prerequisite part of language knowledge that needs to be acquired for a proper use of a learned language. Moreover, non-native speakers have been found to be in short of sufficient collocational knowledge (Nizonkiza, Van Dyk, & Louw, 2013; Takač & Lukač, 2013) even at advanced levels of language proficiency (Arnaud & Savignon, 1997; Channell, 1981; Howarth, 1996; Nesselhauf, 2005).

The abundance and wide-spread use of collocations in the various genres of writing whether they are technical or non-technical can be an additional reason for this interest in collocations. For example, collocations have been found to be much more frequent in native speakers' academic writings, approximately 34% of the total texts, than idioms and frozen expressions, about 5% of the total academic texts (Erman & Warren, 2000; Howarth, 1996).

In conclusion, due to their pervasiveness in use, deficiency in using and producing collocations affect inevitably the naturalness of language productions in the context in which it is used (Smadja, 1993) whether in written or spoken forms. Accordingly, collocational knowledge occupies an exceptional priority in language learning.

2.3.2 The role of collocations in supporting cognitive fluency

Significance of collocations in enhancing cognitive fluency can be explained in that collocational knowledge facilitates constructing utterances. Language users become less dependent on their creativity in constructing utterances, grammar, and lexis. Learners' less reliance on creativity leads to less linguistic processing, accordingly, greater fluency (Bonk, 2000; Ellis, 2003; Sinclair, 1991; Siyanova-Chanturia & Martinez, 2015). According to Mueller (2011), L2 learners resort to collocational knowledge to fill the gap in their semantic knowledge when they fail to construct correct combinations as in the case of using prepositions. When learners cannot predict the correct combination based on the meaning of the components, they refuge to the memorised collocations of this category.

2.3.3 The importance of collocations for pedagogy

Manning & Schütze (1999) argue that neglecting collocations in the previous linguistic studies despite their crucial role in developing learners' language proficiency necessitates more investigations into this significant linguistic phenomenon. Zaabalawi & Gould (2017) argue that despite the credited significance of collocations, the best teaching means which can efficiently develop L2 learners' mastery of collocations still seems unclear. However, they hold that raising awareness combined with frequent exposure to particular exemplar collocations is likely to considerably develop L2 learners' knowledge and use of the encountered items. Additionally, they suggest that teachers should incorporate a reading activity component, as a source of collocational content, when they teach English writing. In addition, they recommend teaching English language as lexical phrases, such as collocations, because they believe that, in line with the modular perspective of language, English language consists of ready-made lexical chunks which should mostly be committed to memory. Similarly, Farrokh (2012) believes that raising awareness of collocations in terms of knowledge and use may help, in theory and practice, set effective means of improving learners' level of proficiency in ways that save effort, time, and costs specifically in pedagogical contexts. In line with, it is maintained that teaching collocations is unquestionably fundamental and of great importance in developing L2 learners' proficiency (Channell, 1981; Lewis, 1993, 1997, 2000; Marton, 1977; Wu, 1996; Wu, Witten, & Franken, 2010; Yazdandoost, AmalSaleh, & Kafipour, 2014) due to their central role in building up L2 learners' competence who endeavour to reach a high level of accuracy and fluency in the target language (Nesselhauf, 2003).

Additionally, it is believed that (Ellis, 2003; Krashen, 1989; Shooshtari & Karami, 2013) vocabulary, including collocations, is positioned at the core of language learning process and that a comprehensible input is a fundamental ingredient for making a significant development in learning a language. Accordingly, much attention should be paid to learning and teaching vocabulary and collocations. Krashen (1989) attributes importance of vocabulary in the learners' language proficiency to practical and theoretical reasons. Practically, it is significant firstly, because gaining a large vocabulary is fundamental for proficiency in a language. Secondly, L2 learners' awareness of vocabulary importance makes

them carry dictionaries rather than grammar books because learners often attribute their language deficiency to their shortage of vocabulary knowledge. Theoretically, Krashen (1989) argues that studying the process of vocabulary learning can assist in understanding the language learning process in general. Additionally, he believes that learning vocabulary is, in principle, similar to learning the other aspects of language.

As far as the general view of vocabulary is concerned, the notion of vocabulary has shifted from the traditional view which thinks of vocabulary as individual words into the view that thinks of vocabulary as lexical units that may consist of more than one item. Schmitt & Carter (2000), for example, argue that the traditional view of vocabulary is inadequate due to the existence of large numbers of units which are made up of more than one orthographic word such as *give up*. Accordingly, vocabulary has become to include, in addition to individual words, various types of lexical phrases such as collocations. Schmitt & Carter (2000) assert the crucial role of these lexical phrases in the vocabulary learning and developing learners' language proficiency.

Nattinger & DeCarrico (1992) highlight the importance of teaching lexical phrases, which include prefabricated units and collocations, rather than focusing on individual words and recommend teaching them since language fluency comes from ability to use lexical phrases skilfully. Additionally, they argue that there is no much difference between the acquisition of the first language (henceforth L1) and learning a foreign language since L2 learners use prefabricated language in a very similar way to how L1 learners produce, understand, and analyse a new language. Accordingly, they assert that lexical phrases can play a vital role and function as an effective basis for L2 teaching. Nattinger & DeCarrico (1992, p. 12), give language acquisition of a child as an example to show how L1 and L2 learners are similar in learning a new language. They state that a child first uses a chunk such as "*I want- to- go*" with a fixed sense, then by time, the child recognises the syntactic potentials of the chunk and detaches it from its context to create new chunks such as "*I want to get up*". Accordingly, they argue that L2 learners behave the same way when they start learning a foreign language; they first learn a prefabricated phrase, then after being able to analyse the syntactic structure of the chunk, they isolate the chunk from its current context and begin to produce new chunks in another

context based on the first one. Finally, they argue that adopting the lexical approach in teaching avoids the shortcomings result from too heavily relying on linguistic theories of competence and communicative competence to use language appropriately as native speakers. Accordingly, they suggest utilising lexical phrases as practical tools inside the class for teaching conversations and the other skills of language to EFL and ESL learners.

Additionally, Lewis (1993), the first who introduced the term “Lexical Approach”, believes that language is grammaticalised lexis and not lexicalised grammar. To him, the greatest part of learning a language is learners’ ability of producing and comprehending lexical phrases as chunks, i.e., chunks are the building blocks of any learned language. Accordingly, lexis should be the central focus not grammar since it plays a substantial role in creating meaning while grammar mainly plays a managerial function in language. For this, according to him, more time should be spent on learning phrases such as collocations and less time on grammar. In addition, he suggests identification of lexical chunks as a basic classroom activity to raise learners’ awareness of collocations. Additionally, he emphasises on teaching frequent fixed expressions and collocations in spoken language with using dictionaries since they provide textual contexts for the learned expressions and collocations. In other words, he calls for developing learners’ language proficiency through focusing on words and word combinations rather than grammar, and using real language not created language by textbook designers based on grammatical patterns that may not be used in reality.

Shooshtari & Karami (2013) note that raising learners’ awareness of collocations through explicit teaching may contribute to promoting their language proficiency in general. Similarly, Alali & Schmitt (2012) contend that formulaic sequences form an important element of discourse and ought to be emphasised in the pedagogical contexts and teaching pedagogy.

In sum, a great deal of emphasis should be put on the aspects that relate to learning and teaching collocations since reaching a fluent, native-like communication level without a considerable collocational knowledge development seems a hard objective to achieve.

2.3.4 The importance of collocations for linguistic applications

Collocations can be of use in a wide range of applications such as Natural Language Processing software (Agrawal, Sanyal, & Sanyal, 2014), making dictionaries, machine translations, parsing phrases, and corpus linguistic research (Stubbs, 1996). These applications can be of considerable benefit for L2 learners in the pedagogical contexts, language learning institutes, and self-study environments. The use of dictionaries, for example, specifically those take L2 learners' needs in consideration such as the Collins COBUILD English Dictionary, BBC English Dictionary, and Oxford Advanced Learners Dictionary can help minimise L2 learners' collocational errors and building and enriching their collocational knowledge (Sadeghi, 2010). Additionally, McKeown & Radev (2000) argue that employing collocations in these linguistic applications can be utilised in a wide range of linguistic fields such as machine translations, statistical applications, language generation, and statistical approaches to words sense disambiguation applications.

Accordingly, I would argue that the importance of collocations in these applications exceeds the boundary of developing L2 learners' collocational knowledge and language proficiency to include the other aspects of linguistic investigations and applications. Additionally, I expect, with the advance of technology and statistical software, further useful linguistic applications that can be employed for increasing L2 learners' proficiency and linguistic studies.

2.3.5 The theoretical linguistic importance of collocations

Being a universal linguistic phenomenon; words are always used in collocations not in isolation (Duan & Qin, 2012), collocation can have its importance in the theoretical knowledge development of language theories such as generative grammar, usage-based models, idiom principle, and lexical priming.

However, the literature in this aspect of linguistics shows that the study of collocations has presented certain theoretical issues because it is thought to be not fitting well into some accepted models of linguistic description, such as Chomsky's generative grammar theory, due to its being a "rather ill-defined area of linguistic patterning that is neither clearly syntactic nor clearly semantic" (Clear, 1993, p. 271).

Nevertheless, I would argue that collocation can partly support even a linguistic theory such as Chomsky's generative grammar theory. Generative grammar is a linguistic theory which is based on Chomsky's theoretical linguistics of

grammar from the 1950s to the 1970s. Generative grammar describes a language in terms of a set of universal logical rules that are innately acquired to be able to generate the infinite number of possible sentences according to the correct structure of that language (Chomsky, 1976). Although the generativists have mainly focused on the study of syntax to provide models for native speakers' competence of language, and collocation is not explicitly mentioned in the linguistic theory of generative grammar literature, the theory has important implication for the view of collocations (Bartsch, 2004). Establishing on the phraseological perspective of collocations, one of the joint assumptions of the generative grammar theory and collocations is their production according to specific syntactic patterning structures. Generativists confine generating sentences in a language according to certain grammatical structures, similarly, phraseologists realise a combination as a collocation only when it is produced in a certain grammatical structure. According to this assumption, I would argue that the concepts of collocations and the generative grammar theory support each other in following certain syntactic patterns in phrases, i.e., collocations, and sentence production.

However, the focus has shifted from the algorithmic, generative models mentioned earlier to the usage-based models of language as proponents of the latter (e.g., Barlow & Kemmer, 2000; Langacker, 2000) believe that usage-based models demonstrate far greater cognitive and neurological plausibility than algorithmic, generative models. According to the language conception of these models, the structural properties of languages emerge from their usage. Accordingly, I would argue that this conception is similar to that is used for defining collocation which is also established on a usage-based principle, specifically, the frequency-based view of collocations. As a result, they both can assist in the theoretical development of each other.

The conception of collocation can also go in good harmony with the Sinclair's (1991) idiom principle theory which posits that a language user has a large number of semi-prefabricated phrases, which can include collocations, that make single choices in the mental lexicon even when they appear to be analysable into single constituents (Sinclair, 1991). Accordingly, prefabricated language may assist in facilitating production and seemingly interpretation of utterances.

Similarly, the lexical priming is a linguistic theory developed by Hoey (2005), which seeks to relate a concept such as collocation to the experimental findings of psycholinguists interested in the retardation and acceleration of word association, can also be related to the conception of collocations. According to this theory, characteristics such as frequency, recurrent occurrence of multiword expressions, and predictability have central roles in accelerating the process of word associations. For example, on hearing or reading the beginning of a collocation, e.g., *extenuating*..., a mature language user is compelled to complete it with one of the most likely words such as *circumstances* (Siyanova-Chanturia & Martinez, 2014). Accordingly, the theory has an important implication for collocations as it can explain why collocations exist; a listener can recognise a word more quickly when a related word is given (Pace-Sigge, 2013).

Another aspect of the significance of collocations is the outstanding overlap between the concept of collocations and the other multi-word entities such as phrases, idioms, and named entities (Karan, Sanjdar, & Basic, 2012). As clear-cut boundaries have not been established, up to date, between collocations, on the one hand, and the other multi-unit entities, free combinations and idioms, and the overlaps exist among the various types of combinations, it may necessitate including this aspect of linguistics, collocations, in the study whenever the other two aspects are investigated. In conclusion, collocations may play a role not only in developing learners' proficiency but also in the theoretical knowledge development of language.

2.3.6 The role of collocations in meaning

An additional significance of collocations can be its assumed role in being part of the meaning of a word. According to Firth's (1957, p.179) theory of meaning, "you shall know a word by the company it keeps". This means that the meaning of a word is in part determined by the characteristics of its collocates. For example, a part of the meaning of the word *dark* is known by its accompany with the word *night* which means "little or no light". As *dark* can mean different senses in different lexical contexts, for example, *dark green* "dark= approaching black in shade"; *dark secret* "dark= mysterious"; *dark L* "dark= velarised form of the sound of the letter l" (Oxford Dictionary of English, 2015). Hence, deficiency in the collocational knowledge results in an inadequate picture of what words

mean or at least losing part of the meaning and accordingly misunderstanding in communication.

This short overview of the significance of collocations demonstrates that collocations are important not only in theory but also in practice. Theoretically, collocations are important due to their connectedness and overlapping with other types of word combinations. Practically, collocations are important because of their role in generating natural or native-like expressions specifically in academic writings and computerised applications such as dictionaries and machine translations of texts and phrases.

2.4 Classifications of collocations

2.4.1 Introduction

Collocations have been classified into various types (e.g., lexical, grammatical, upward, downward) and patterns (e.g., v+n, adj+n, prep+n) differently according to authors' perspectives and the criteria used for classifying collocations (e.g., Benson, Benson, & Ilson, 1986; Howarth, 1996; Sinclair, 1991). Consequently, the divergent understanding of the concept of collocation and lack of consensus in this aspect resulted in generating various collocation types and patterns (Shokouhi & Mirsalari, 2010), as explained in the following sections.

2.4.2 Types of collocations

Different types of collocations have been suggested by different researchers. Sinclair (1991, pp. 115-116) distinguishes between “upward” and “downward” collocations. Accordingly, the concept “upward collocation” denotes that words habitually collocate with words that are more frequently used than they are themselves in the English language, e.g., the word *back* collocates with *at*, *down*, *from*, *into* and *on*, all of which are more frequent words than *back*. In contrast, the concept “downward collocation” means that words habitually collocate with words that are less frequent than they are, e.g., the words *arrive*, *bring* and *climbed* are examples of less frequently occurring words which collocate with the word *back*. This distinction demonstrates a difference in the grammatical structure of both types of collocations. While constituents of the upward collocations are inclined to be mostly grammatical frames such as grammatical elements and subordinators since nodes in this type of collocations usually collocate with prepositions, adverbs, conjunctions, and pronouns, nodes

of downward collocations habitually collocate with semantic words such as nouns and verbs.

Similarly, Howarth (1996) classifies collocations based on the criterion of commutability and makes a distinction between five levels of restrictedness. Additionally, while classifications of collocations of some linguists (e.g., Hausmann, 1989) are exclusively based on open-class words, most linguists' classifications of collocations (e.g., Benson, Benson, & Ilson, 1986) include function words and syntactic structures in addition to open-class words.

Concluding from the majority of the previous studies (e.g., Benson, Benson, & Ilson, 1986; Baker, 1992; Gitaski, 1996; Wei, 1999), collocations can mainly be divided into lexical and grammatical collocations. Lexical collocations generally comprise an array of associations of verbs, nouns, adjectives, and adverbs, for example, *launch a missile*, *withdraw an offer*, *a crushing defeat*, *storms rage*, *a world capital*, *deeply absorbed*, and *appreciated sincerely* (Benson, Benson, & Ilson, 1986, pp. 253-254). Alternatively, grammatical collocations commonly consist of a verb, a noun, or an adjective combined with a preposition or a grammatical structure such as an infinitive or a clause, e.g., *account for*, *adjacent to*, *advantage over*, and *to be afraid that* (Bahns & Eldaw, 1993). The dichotomy (lexical vs grammatical collocations) seems widely acceptable among linguists and researchers. Accordingly, it is employed as the basic distinctive criterion in differentiating between types of collocations in the current study.

2.4.3 Patterns of collocations

Various patterns of collocations have been suggested by researchers. Generally, the patterning seems to be mainly established on the syntactic structures of the combinations. Hausmann (1989), for example, suggests the following patterns of collocations: n+adj, n+v, v+n, v+adv, adj+adv, n+(prep)+n. Benson, Benson, & Ilson (1986), on their parts, classified collocations into the following patterns: v+n, n+adj, n+v, n+prep+n, adj+adv, v+adv whereas Smadja (1993) introduces the following patterns: n+adj, n+v, v+n, v+prep, v+adv, v+v, n+prep, n+det.

A further classification is suggested by Gitsaki (1996, pp. 164-166) through analysing a large number of assumed collocations. She distinguishes among 37 types of collocations of which 8 as lexical collocations and 29 as grammatical

collocations. The suggested lexical collocation types are v+n/pron., v+n, adj+n, n+v, n+ of +n, adv+adj, v+adv, and n+n. The suggested grammatical collocation patterns are n+ prep, n+to+inf, n+to+inf, n+that-clause, prep+n, adj+prep, predicate adj+ to+inf, adj+that-clause, svo+to+ o/svoo, svo+to+o, svo+for+o/svoo, sv(o)+prep+o, sv+to+inf, sv+inf, sv+v-ing, svo+to+inf, svo+inf, svo+v-ing, sv+ possessive v-ing, sv(o)+that-clause, svo+to be+ complement, svo+ complement, svoo, sv(o)+adverbial, sv(o)+wh-word, s(it)vo+to+inf, sv+complement, miscellaneous as in *in fact*, prep+det+n and phrasal verbs. In contrast to the large number of patterns of collocations which are suggested by Gitsaki, Lewis & Hill (1997) introduce a small number of patterns which included only five collocation patterns: adj+n, v+n, n+v, adv+adj and v+adv. For the abbreviations which are used here, it can be referred to the table of abbreviations.

In sum, based on the presented divisions of collocations in this review, collocations can be classified according to types and patterns. According to types, they can be divided into lexical and grammatical collocations. However, it seems difficult to provide a fixed number of collocation patterns in the same simple way of classifying their types due to the wide range of variety in researchers' suggestions of the collocation patterns. Nevertheless, it would be more useful if all the suggested patterns are considered since these patterns recurrently occur in the native speakers' spoken and written text forms.

2.4.4 The importance of classifying collocations

The importance of classifying collocations can be observed in various aspects. Initially, it is thought that both the lexical and grammatical patterns of combinations in which a word is used can play a substantial role in recognising and understanding a word in a particular sense (Duan & Qin, 2012). In other words, the pattern of combination in which a word is employed, i.e., the textual context, determines to a large extent its meaning. For example, the use of the verb *enjoy* in different patterns such as "v+ v-ing" and "v+n" patterns can yield different senses. Accordingly, selecting collocates in these two patterns are not done haphazardly, for example, it can be said *enjoy a dinner*, *enjoy a good income* or *enjoy one's life*, but not ?*enjoy death*, or ?*enjoy crime* (Duan & Qin, 2012, p. 1890).

Furthermore, recognising the pattern of a collocation can be useful as a tool for identifying L1 users' and L2 learners' developmental patterns of knowledge and use in corpus-based studies, identifying learners' errors, weaknesses, strength points, and their common characteristics (Chen, 2010; Paquot & Granger, 2012). These patterns can function as a parsing tool which sort phrases according to the set patterns in computational studies (Biber, 1993).

Additionally, patterns can be used as modules for extracting collocations from large corpora by using computerised software which in turn provides data that can be used as raw data in other studies (e.g., Seretan, 2011; Seretan, Nerima, & Wehrli, 2003; Seretan, Nerima, & Wehrli, 2004). Moreover, classifying collocations can also be of help in the semantic interpretation of collocations (Otero, 2008). Accordingly, it can be used to differentiate between the other types of combinations and collocations such as discriminating between collocations and idioms since they have rather different and distinctive patterns. An additional significance of classifying collocations can be seen in its use in the pedagogical contexts, for instance, learners can generate similar combinations when they acquire their general patterns or formula (Tutin, 2005).

In sum, classification of collocations can have benefits in theory and practice. It can also be utilised on NLP (Natural Language Processing) applications such as dictionaries and collocation extraction and parsing software.

2.5 Collocational knowledge

2.5.1 Introduction

Assigning collocations to vocabulary knowledge or not seems a controversial matter because some characteristics of collocations are different from those of the individual words in terms of structure, learning and teaching. While some researchers have used the term "vocabulary" to include collocations as a part or level of learners' total vocabulary knowledge (e.g., Catalán, 2002; Milton, 2009; Nation, 2013; Richards, 1976; Schmitt & Carter, 2000), some other linguists have dealt with this aspect as a separate area of linguistic knowledge (e.g., Firth, 1957). However, this aspect of knowledge appears to be of multifaceted or multidimensional, interlinked and complex nature that requires to be investigated from different angles (Henriksen, 1999; Nation, 2013; Read, 2000). As far as the current study is concerned, the term "vocabulary" is used to include collocations as a level of vocabulary knowledge since they both have

many characteristics in common particularly those related to the focus of the current study in terms of knowledge dimension classifications. Accordingly, the discussions and knowledge classifications which apply to vocabulary are taken to apply also to collocations.

Researchers have emphasised the significance of vocabulary knowledge in learning a language to the extent to consider learning a language as an issue of learning vocabulary (Levelt, 1993; Meara, 1996, 2002; Nation, 2013; Read, 2000). Generally, vocabulary knowledge is assumed to encompass numerous aspects of knowledge such as grammatical configurations, meaning, pronunciation, orthography, morphology, collocability, syntactic restrictions, semantic and syntactic relations, underlying form and derivations, connotations, and the pragmatic use of words (Catalán, 2002; Richards, 1976). Chapelle (1998) emphasises four aspects of vocabulary knowledge: lexical access processes, word characteristics, size, lexicon, and organization whereas Henriksen (1999) identifies only three interconnected aspects of vocabulary knowledge: partial-precise knowledge, depth of knowledge, and receptive-productive dimensions. Nation (2013), on his part, identifies nine aspects of vocabulary knowledge that are encompassed by three main umbrellas of form, meaning, and use where each has receptive and productive knowledge aspects. Regarding form knowledge, Nation sub-classifies it into spoken, written, and word parts knowledge. Knowledge of meaning is divided into form and meaning, concept referents, and associations. Finally, knowledge of use is classified into knowledge of grammatical functions, collocations, and constraints on use. Hence, learners' vocabulary knowledge exceeds merely knowledge of form, meaning and pronunciations as may be thought of learners to include many other aspects.

In conclusion, despite the substantial position of vocabulary knowledge, including collocations, in the process of language learning there seems to be no consensus on a definite classification of aspects or levels of vocabulary knowledge. However, studies in this field, i.e. vocabulary, indicate the significance of certain aspects of vocabulary knowledge such as collocability which is the main concern of the current study. Collocational knowledge and its dimensions are discussed in more detail in the following sections.

2.5.2 Dimensions of collocational knowledge

Typologically and terminologically stating, vocabulary knowledge dimensions have been classified into various dimensions and they have been labelled using various terms. Linguists have generally drawn the attention to the significance of distinguishing between two aspects of vocabulary knowledge which are fundamental to learners' language proficiency: breadth and depth of vocabulary knowledge (Anderson & Freebody, 1979) and these two aspects are believed to be absorbed differently (Laufer & Goldstein, 2004). For example, the terms receptive knowledge, passive knowledge, and vocabulary size have been used to denote breadth of knowledge of vocabulary; ability to recognise words or the number of words learners know whereas terms such as productive knowledge and active knowledge have been used to refer to depth of vocabulary knowledge (e.g., Anderson & Freebody, 1979; Melka, 1997; Milton, 2009); how well learners know about words and their uses in written and spoken forms in real contexts (Qian, 2002). Generally, the use of the dichotomy receptive-productive to classify learners' vocabulary knowledge dimensions seems widely accepted (Melka, 1997).

As for the relationship between receptive and productive knowledge, some studies indicate a rather strong correlation between learners' receptive and productive knowledge of vocabulary (e.g., Akbarian, 2010; Mehrpour, Razmjoo, & Kian, 2011; Zhong, 2016), whereas some other studies (e.g., Alharthi, 2014; Melka, 1997) doubt the ability of drawing a clear-cut border between learners' receptive and productive knowledge of vocabulary. The existence of a fuzzy border between learners' receptive and productive knowledge is ascribed to that L2 learners' knowledge of vocabulary is depicted as a gradable multi-level knowledge scale, accordingly, it may not be applicable to claim that a person has either whole knowledge of vocabulary or nothing at all (Fan, 2009; Nation, 1990, 2001; Richards, 1976; Shen, 2008). According to these researchers, this is because language knowledge cannot be obtained in one shot and it is not an all-or-nothing phenomenon rather it should be perceived as a continuum that consists of various levels of knowledge. In other words, L2 learners go through various phases of vocabulary knowledge development before attaining a native-like language proficiency. Furthermore, Koizumi (2005) argues these two aspects seem to be correlated to each other and they cannot be separated or isolated or studied adequately without probing their effects on each other or

investigating their interrelationships. Along this view, Vermeer (2001) finds that there is no theoretical or abstract distinction between receptive and productive vocabulary knowledge and that knowledge of both aspects are affected by the same factors. However, these arguments should not imply that learners' receptive knowledge cannot be distinguished from their productive knowledge as this has been proved to be feasible through numerous studies and tests (e.g., Gyllstad, 2007; Laufer & Goldstein, 2004), what they really indicate is that there is no an obvious pivot point or signpost phase which indicates clearly a learner's transference from the receptive phase to the productive phase of vocabulary knowledge development due to the interactive and overlapping relationship between these two aspects of knowledge.

Tahmasebi, Ghaedrahmat, & Haqverdi (2013) believe that there are different degrees of word knowledge, and language users are assumed to be aware of what a word means prior to using it in a sentence. Additionally, they establish a significant relationship between L2 learners' receptive and productive knowledge, on the one hand, and their language proficiency, on the other hand.

However, Meara (2002) believes that the dimensional approach is not sufficiently capable of explaining the diverse nature of learners and that more investigations are needed to probe the various aspects of learners' knowledge of vocabulary. Alternatively, he suggests concentrating more on learners' productive (performance) knowledge than their receptive (competence) knowledge. This shift in emphasis is justified by that L2 learners, unlike L1 speakers, lack noticeably the ability to effortlessly recognise and retrieve L2 words, i.e., they need more emphasis on their performance than on their competence. In contrast to Meara's emphasis on one side more than the other, Akbarian (2010) argues that research indicates a positive correlation between breadth and depth of vocabulary knowledge. Accordingly, these two aspects should not be emphasised separately in a way that undermines the importance of either of these dimensions at the expense of the other specifically in pedagogical contexts.

In conclusion, despite the diverse terminologies used by researchers to describe vocabulary knowledge dimensions, the general views seem to classify vocabulary knowledge into two dimensions of which the dichotomy (receptive vs productive) appears to be widely accepted by researchers. As far as the current

study is concerned, the terminologies apply to vocabulary knowledge classifications are taken to apply similarly to collocational knowledge as well. This is because collocational knowledge is considered part of or at least very similar in characteristics to vocabulary knowledge in this respect. Accordingly, collocational knowledge is classified into receptive and productive dimensions and the data analyses of the current study will be conducted on this basis.

2.5.2.1 Receptive knowledge of collocations

The significance of the receptive knowledge in language learning is well evidenced in literature and believed to be straightforwardly connected with L2 learners' skills of using language (Gallego & Llach, 2009; Schmitt, Schmitt, & Clapham, 2001) specifically for their academic progress (Yuksel, 2013). Additionally, Milton (2010) believes that learners' receptive knowledge ought to provide a reliable indicator of learners' linguistic proficiency. In addition, Mehrpour, Razmjoo, & Kian (2011) state that, concluding from their study, learners with larger vocabulary size seem to be of deeper vocabulary knowledge. In other words, an increase in learners' receptive knowledge correlates positively with the increase in their productive knowledge. However, they argue that learners' productive knowledge appeared to be contributing more than their receptive knowledge in comprehending reading texts.

In sum, receptive knowledge of vocabulary in general and collocations in particular is a distinctive aspect of learners' language knowledge and can reliably be used as an indicator of learners' language knowledge proficiency. Additionally, this aspect of knowledge appears to correlate positively with learners' productive knowledge.

2.5.2.2 Productive knowledge of collocations

Knowledge of vocabulary exceeds the merely surface or superficial knowledge of form and meaning. Rather, it probes the other aspects of vocabulary knowledge such as pronunciation, meaning, register, spelling, frequency, morphology, syntax, and collocational properties (Qian, 2002). Moreover, it is argued that having a large or sufficient vocabulary size, though it is controversial what sufficient size is, alone does not guarantee a native-like use of the vocabulary, and that vocabulary knowledge is more than meaning and form of a word (Zhong, 2011). Hence, skilful use of vocabulary requires in addition to a reasonable size of vocabulary also depth of knowledge since there

are various aspects of vocabulary knowledge that need to be considered (Read, 2007). According to Nation (2013), the productive knowledge of collocations indicates skilful knowledge about what words should be used with the word under focus. It is believed that learners' productive knowledge plays a more important role than their receptive knowledge in performance, for instance in reading comprehension (Mehrpour, Razmjoo, & Kian, 2011). In line with, Tahmasebi, Ghaedrahmat, & Haqverdi (2013) believe that learners' productive knowledge can be a better indicator of their language proficiency than their receptive knowledge.

However, by contrasting learners' receptive and productive knowledge, there is abundance of evidence on that L2 learners differ significantly with regard to their receptive and productive knowledge of vocabulary. Studies indicate that L2 learners' productive knowledge lags far behind their receptive knowledge (e.g., Alsakran, 2011; Laufer, 1998; Milton, 2009; Torabian, Maros, & Subakir, 2014; Webb, 2008). Webb (2008) argues that the ratio of the L2 learners' productive knowledge to their receptive knowledge is from 77-93%. Similarly, Milton (2009) observes that learners' productive knowledge comprises 50-80% as compared with their receptive knowledge. In other words, learners' productive knowledge lags far behind their receptive knowledge for about 20-50%. Seemingly, this gap increases with advancing learners' levels of proficiency (Laufer, 1998).

Despite the role of these two aspects of knowledge, receptive and productive knowledge, to L2 learners' proficiency, it is thought that these two dimensions of vocabulary knowledge together are not sufficient as knowledge about these two aspects are still not enough to give a comprehensive depiction of the complex nature of vocabulary knowledge (Zhong, 2011). Accordingly, more investigations are required to identify and explore the other factors which have not been accounted for that could be revealed through more investigations into this field.

Collocationally, although some researchers believe that collocations and collocational knowledge should be measures of learners' depth of vocabulary knowledge (e.g., Milton, 2009; Nation, 2013; Read, 2000; Schmitt, 2010a), a researcher such as Gyllstad (2013) finds no convincing evidence that supports the notion that collocations and collocational knowledge should belong in the depth dimension of vocabulary knowledge. Gyllstad (2013) ascribes the

controversy of assigning collocations and collocational knowledge to the depth dimension to the ambiguous definition of the depth dimension itself and backs his argument by Milton's (2009) statement that depth has not been adequately and clearly defined to the moment. Accordingly, it is not certain whether collocations should reside in the depth dimension of vocabulary knowledge. Furthermore, collocational knowledge, similar to vocabulary knowledge, can be receptive and productive as evidenced in various studies (e.g., Anderson & Freebody, 1979; Koya, 2005; Laufer, 1998; Laufer & Waldman, 2011; Melka, 1997; Milton, 2009; Webb, 2008).

Hausmann (1984) highlights the significance of the collocational knowledge to the extent to view vocabulary learning as learning of collocations in action. Additionally, it is believed that knowledge of collocations can help develop learners' receptive and productive skills (Fan, 2009). The rationale is believed to be that learners acquire vocabulary better when pairs of words are associated as this makes remembering words easier and faster and memorising collocations help them recognise lexical restrictions and use collocations as ready-made building blocks (Yazdandoost, AmalSaleh, & Kafipour, 2014). Furthermore, studies indicate that collocations appear to be memorised as wholes rather than individual items and they are remembered and produced faster than the individual items (Conklin & Schmitt, 2008; Ellis, Simpson-Vlach, & Maynard, 2008).

Zhang's (1993) observes, based on a conducted study, a correlation between learners' knowledge of collocations and their writing quality and the use of collocations in essays and writing quality. Additionally, he notes that collocational knowledge is a source of fluency in written communication. In addition, he concludes that quality of collocations, in terms of variety and accuracy, can be an indication of the quality of learners' writing.

In conclusion, learners' collocational knowledge can be receptive and productive. Additionally, these two aspects are positively correlated with each other but they develop at different speeds. Learners' receptive knowledge develops significantly faster than their productive knowledge. In addition, learners' collocational knowledge influences their other linguistic skills such as fluency, writing quality, and comprehension.

2.5.3 How much collocational knowledge is needed?

I would argue that information about the required amount of vocabulary knowledge for L2 learners to be covered at particular levels of learning can have important theoretical and practical implications. Theoretically, it helps researchers evaluate learners' vocabulary knowledge at certain levels of language proficiency by comparing knowledge of the study sample with the expected amount of knowledge at those levels. Pedagogically, it identifies areas of strengths and weaknesses in teaching approaches and curricula settings which implies more educational procedures to overcome spotted shortcomings or deficiencies in the process of learning and teaching language. For example, knowing the required vocabulary knowledge, for instance at Grade 10, can be utilised as an indicator in setting the English curricula textbooks at this level in terms of number of words, collocations, phrases, expressions and new words that textbooks should contain. Accordingly, this information helps avoiding overloading or under loading curricula textbooks.

Vocabulary, including collocations, is not an end for itself; it is used as a means of accessing and expressing feelings, thoughts, and ideas in the target language (Laufer & Nation, 1995). However, the number of words which is a prerequisite for L2 learners' fluency is still controversial. Nevertheless, many researchers have attempted to estimate the size of vocabulary knowledge required for L2 learners specifically for academic purposes at specific levels (e.g., Goulden, Nation, & Read, 1990; Nation & Waring, 1997; Nizonkiza, van Dyk, & Louw, 2013).

English language comprises more than 54000 word families (Dupuy, 1974; Goulden, Nation, & Read, 1990) of which university graduate native speakers are anticipated to know around 20000 word families, or more (Nagy & Scott, 1990). However, due to the large number of words, similarly collocations, and the required rapid rate of vocabulary development, it is believed that it is too difficult to cover all the vocabulary required for learning a language through explicit teaching or direct instructions (Nagy & Herman, 1987). Nation (1990) asserts that the appropriate load of vocabulary for each year is crucial since overloading textbooks with vocabulary may lead to poor or unintended results, and accordingly, confusing learners rather than building up their vocabulary knowledge. Consequently, learners may become unable to keep pace with the intended educational program. This in turn can reflect negatively on the total

learning process in terms of learners' knowledge, time consumed, efforts exerted, and expenditure costs spent on the educational process.

Nation (2013) argues that for non-natives to be fluent in the target language, they need to approach the vocabulary knowledge level of the native speakers, specifically high frequency and specialised words. However, he argues that approaching this extent of vocabulary knowledge does not seem to be practical through explicit instructions due to the large number of words, diverse nature of learners and shortage of time. Hence, external parties such as teachers and learners' own efforts such as extensive readings of various genres such as stories and newspapers as different contexts can play vital roles in enriching their vocabulary, and accordingly their collocational knowledge (Hu, 2013). Additionally, watching movies and video language courses can assist in sophisticating, developing, and building up learners' vocabulary in general and raising their awareness, knowledge, and preservation of collocations in particular (Boers, Lindströmberg, & Eyckmans, 2014).

In conclusion, developing L2 learners' vocabulary knowledge is essential in learning a language. Additionally, whether to consider collocations as a part of vocabulary, or a transitional phase between words and sentences or a separate level of knowledge, what applies to vocabulary knowledge in general seems also to apply to collocations in terms of amount of knowledge needed for reaching native-like fluency in the learned language. Accordingly, learners' collocational knowledge development at these two levels, receptive and productive, is essential in producing a natural and fluent communication. Receptively, learners should have a good command of collocational knowledge to avoid misunderstanding communications (Martinez & Murphy, 2011). Productively, the development of this aspect of knowledge is important so as not to allow for producing odd or non-nativelike expressions (Barfield & Gyllstad, 2009). Accordingly, advancing learners' collocational knowledge, receptively and productively, is a prerequisite for learners' language proficiency.

2.6 The linear and nonlinear view of language knowledge and development

It is believed that humans behave habitually in a linear fashion because it saves effort, energy, and time, and it simplifies complex issues (Hase, 2014). Accordingly, humans like to find causation relationships for the phenomena in

their environment. However, investigations into language knowledge and development have focused on either the product or the process of language development (de Bot, Lowie, Thorne, & Verspoor, 2013) and these two aspects have been generally viewed either from a linear or nonlinear perspective.

Although the current study is conducted from a linear perspective, introducing the non-linear view of language is necessary as well. This is because even the non-linear aspects of language are sometimes studied linearly either due to unavailability of efficient tools and resources for conducting non-linear investigations or due to the study aims which necessitate a linear study design. Conversely, some aspects of language knowledge and development may only be profoundly explored when it is conducted from a non-linear perspective. Accordingly, this makes introducing both views preferable in this review of literature. The linear and nonlinear models of language knowledge and development are reviewed in the following sections.

2.6.1 Linear perspective of language knowledge and development

Briefly defining, a linear system is “a system in which the whole is the sum of its parts” (Verspoor, de Bot, & Lowie, 2011, p. 100). Within the linear model, each component part of the system is described at one point in time. The description of the developmental pattern of its components can be done through applying fixed relationships between two variables, in which a change in the first variable is described as a function of the other variable in the form of the formula “ $Y = f(x)$ ”. Accordingly, the linear development of the sum of all the components of the system equals the amount of development in the whole system. The underlying assumption of the linear model of development is that every action causes a fixed reaction, and every reaction triggers the next reaction. Accordingly, predicting the status of the system development at any point in time is feasible. This kind of predicted development is usually called deterministic (Lowie, Caspi, van Geert, & Steenbeek, p. 106).

Several advantages have been attributed to the linear model. One of its features is that it is of versatile use since it can be used in various aspects of the human day life such as economics, manufacturing units, social matters, military matters, linguistics, computer science, and engineering models (Gaines, 1969). Additionally, it is characterised by simplicity since it is easy to understand the implementation of the process (Roberts, 1986; Kramsch, 2012).

However, several limitations have been assigned to the linear model. Zhao (2010) argues that linear models are built on an unrealistic assumption which assumes that factors never really change, i.e., they disregard change and evolution of variables. Zhao contends that factors or variables do change over time. Another limitation is the limited ability of linear models in solving complex problems, though it is believed that linear models can be generalised to deal with multivariate problems (Roberts, 1986).

Linguistically, applying the linear model on language implies that the language system consists of independent variables, such as phonological, semantic, and structural subsystems or levels that interact with each other through linking points or interfaces that connect these levels together (Lowie, 2012). Among these levels, the syntactic structure of language, which has been the focus of most of the studies, is believed to be regular and predictable. Proponents of this orientation (e.g., Baker, 2001; Chomsky, 2005; Crain & Lillo-Martin, 1999; Hauser, Chomsky, & Fitch, 2002; O'Grady, 1997) believe in a universal grammar (UG) that can be applied to all languages. Accordingly, languages have some basic principles that can be applied to all languages and they are thought to be innately possessed by all human beings. To them, this innate competence enables people to create and comprehend novel utterances that have not been used before. In accordance, the major interest of this group is to find out universals which people have in common that can be applied to L1 acquisition and L2 learning.

However, this view of language knowledge is rejected by some other scholars (e.g., Plumert & Spencer, 2007; Thelen & Smith, 1994; Tomasello, 2004; Tomasello & Abbot-Smith, 2002). Tomasello (2004), for instance, describes the UG as a very fragile hypothesis for two reasons. First, very few precise formulations are available concerning what the UG holds. Second, it offers a very limited number of suggestions for how one can test any of the assumptions that are hypothesised by the UG. Moreover, Kramsch (2012) argues that L2 learning does not develop in the same sequential order of the chapters or units of the curricula textbooks. This is because language development is thought to be a nonlinear and complex rather than a linear process (Zhao, 2010).

Consequently, due to the limitations in the linear perspective, it is believed that the linear system is incapable of adequately reflecting the true picture of

language knowledge and development (Miller, 2011). Hence, a more flexible theory of knowledge is needed to apply to language knowledge in terms of theory, change, and acquisition. This paved the way for the rise of the dynamic view of language development, which is originally a theory that has its roots in Newtonian mechanics, physics, and mathematics, which is reviewed in more detail in the next section.

2.6.2 The nonlinear view of language knowledge and development

2.6.2.1 The Dynamic Systems Theory

Prior to viewing language from a dynamic perspective, it may be more helpful in understanding this perspective when a short review of the Dynamic Systems Theory is presented. The Dynamic Systems Theory, (henceforth DST), begins as a branch of mathematics to deal with simple systems, such as the double pendulum which consists of two variables but with a complex trajectory, then it becomes the science of complex systems (de Bot, Lowie, & Verspoor, 2007). Thelen & Smith (1996) state that the dynamic system is a kind of organisation in which its lower-level constituents, i.e., subsystems, are unstable and interact together to generate specific patterns of behaviour at various levels over time in a way that creates the general stability at the higher levels of the system. In other words, the dynamic system is imbalanced at the subsystems level but balanced at the whole system level through working all parts together as one team. Additionally, every single subsystem is related to the other subsystems and the whole system in the form of a restricted complex network. Working of these parts together creates the general stability of the whole system. The different levels of the system develop differently from each other, and they do not advance at the same pace but the whole picture of the final result of the system reflects a balanced stable output of the process.

According to de Bot, Lowie, Thorne, & Verspoor (2013), systems are defined as groups of entities that function together. Each system encompasses some other sub-systems which in turn can include some other sub-sub-systems that dynamically correlate to each other to achieve the stability of the whole system. Furthermore, they contend that dynamicity of the system indicates the changes that occur to a system in reaction to the influence of internal and external variables or forces. Additionally, the dynamicity of the system indicates also that

these changes take place expectedly at times and unexpectedly at others, and this intricate system nature applies to all the levels of the complex systems.

In terms of systems complexities, it is argued that complexity of systems arises when an explanation of a phenomenon is sought in a context combined with analysing different temporal layers or patterns and at various levels of the process (Lerner & Kaufman, 1985). Moreover, Papachristos (2012) argues that since change and continuity are thought to be a matter of time, the latter is used as a frame of reference to identify and clarify type and nature of the intended changes which are sometimes difficult to identify and explain. Additionally, continuity is employed to identify the impact of the changes that occur at a level on the other levels.

Theoretically, a number of theories seem to be compatible with the basic principles of the DST such as “cognitive linguistics, connectionist theories, emergentism, grammaticalisation theory, activation theory and usage-based L1 acquisition” (de Bot, Lowie, Thorne, & Verspoor, 2013, p. 209). However, none of them seems solely to have the potentiality of yielding a comprehensive view of the studied phenomenon since each theory concentrates on certain aspects and neglects others. It is believed that, these theories are not regarded as separate or isolated ones but rather as complementary theories to each other (Ellis, 1998; Robinson & Ellis, 2008), accordingly, they better work together to yield better results.

Like any other theory, the DST has its strengths and weak-points. One of the advantages of this theory is its assumed accuracy and appropriateness for processing changes in phenomena to probe further than the prevailing static linear perspective of language acquisition (van Geert, 2008). Another feature is its close connectedness with the observed phenomenon which enables testing and replicating its application (van Geert & Steenbeek, 2005), accordingly, it is closely connected with experimental studies (Spencer et al., 2006). Additionally, it is characterised by its ability of following up the development of a phenomenon at both the micro, i.e., the short-term, and the macro, i.e., the long-term, levels and how they interactively go together through the step-by-step of the change process time (de Bot, Lowie, Horne, & Verspoor, 2013; van Geert & Steenbeek, 2005; Thelen, 2005). An additional advantage is its ability of portraying flexibility and stability of phenomena across time (Thelen, 2005).

According to Smith (2009), it is characterised by its elasticity in borrowing concepts from and account of other theories such as its account of Piaget's A-not-B task and it also takes context in its consideration at examining a phenomenon behaviour. Furthermore, a number of other features can also be attributed to the nonlinear system such as being "dynamic, chaotic, unpredictable, sensitive to initial conditions, open, self-organising, feedback sensitive, adaptive" (Larsen-Freeman, 1997, p. 142), "completely interconnected, nonlinear in development, change through internal organisation and interaction with environment, dependent on internal and external resources, constantly changing, and iterative" (Verspoor, de Bot, & Lowie, 2011, p. 9).

On the other side, this theory has not been away from criticism. Among these, it is criticised for being used in its metaphorical sense rather than applying its actual mathematical models which may weaken the analytic and explanatory power of the model (Leeuwen, 2005). For example, according to (Spivey, 2007), it does not account for the neurophysiology sufficiently in spite of holding of the actual neural functions in terms of time and state space. More elaborated, it gives an account of the genuine neural parts that perform the cognitive functions but through artificial segmentation of the time and place of the cognitive processes while, in reality, time is continuous and not like digital computerised clock that stops to be counted. Accordingly, these simulation models look more descriptive rather than be explanatory account of the phenomenon. Furthermore, it is argued that in applying this theory to the development of language and language acquisition, there is ambiguity in the way these notions and mathematical equations can be applied which does not seem accessible for an ordinary reader, and the researcher who adopts this theory appears as if he begs to be involved in troubles (van Geert & Steenbeek, 2005). Moreover, due to its recent history (Miller, 2011), it may not have been verified, revised, or critiqued sufficiently in domains other than mathematical or computational contexts to ensure its viability. Furthermore, due to the belief in the existence of various pathways of development (Thelen, 2005), the DST seems to overstress the role of the individualism of development and does not put a sufficient emphasis on resemblance or analogies among individuals (Spencer et al., 2006). Additionally, some inconsistency is attributed to it; while Thelen (1985), claims some universal developmental principles (Spencer et al.,

2006), she also points out to the impossibility of predicting how the development course would be (Thelen, 2005).

In conclusion, despite the critiques, the DST has found its place in many disciplines and it seems to offer a promising future in explaining complex issues in many disciplines amongst which applying the principles of this theory on L2 learning appears to be one of the vivid areas. Accordingly, this perspective paved the way for likely more comprehensive depictions of language theory and language development which can be encompassed under the umbrella of the Dynamic Systems Theory.

2.6.2.2. The dynamic view of language

An increasing number of researchers (e.g., de Bot, Lowie, & Verspoor, 2005, 2007; de Bot, Lowie, Thorne, & Verspoor; 2013; Kramsch, 2002; Larsen-Freeman 1997, 2002, 2006; Larsen-Freeman & Cameron, 2008) have been calling for adopting the DST in language and learning due to its claimed appropriateness to deal with the complex nature of language and social contexts, and that language development demonstrates some of the key characteristics of the dynamic systems such as high sensitivity to initial conditions, complete interconnectedness of subsystems, and variation both in and among individuals. Accordingly, this view has been increasingly adopted in various studies (e.g., Bell, 2009; Caspi, 2010; Churchill, 2007; Spoelman & Verspoor, 2010; Verspoor, Lowie, & Dijk, 2008; Verspoor, Schmid, & Xu, 2012).

The nonlinear dynamic view of language does not necessitate possessing the innate universal grammar or a specialised language faculty to describe human beings' linguistic creativeness and the related restrictions of producing and using language but their faculty of learning language disposition (de Bot, Lowie, & Verspoor, 2007). Moreover, dynamically viewed studies show that the cause and effect relationship frame, i.e., the linear view of language, is unable to yield a satisfactory portrayal of language development. Alternatively, language knowledge change occurs within complex systems which in turn consist of many constituents that interact with each other over multiple time scales (Christiansen & Kirby, 2003). Generally, proponents of the dynamic view call for shifting researchers' focus from *what* to *how* of the developmental change (Spencer, Perone, & Buss, 2011). In other words, they call to emphasise on how, i.e., the way, the development change occurs rather than concentrating on

what developmental changes occur.

Atkinson (2002, p. 525), suggests the concept of socio-cognitive perspective of language and language acquisition and learning as a substitute for the prevailing view of cognitivism. In contrast to the cognitive orientation which emphasises solely on what occurs inside the head of individuals, Atkinson (2002) presents language and L2 learning as social and cognitive phenomena, and argues that they are concurrently structured and built up in both worlds cognitively “in the head” and socially “in the world”. To Atkinson, language is a social phenomenon since it is utilised by people to communicate their feelings, thoughts, and perform their actions in a social context with the other individuals, and out of this environment, it becomes senseless. At the same time, language is a cognitive phenomenon because human beings have developed their cognitive mental abilities and use of language in a way that mirrors the essential characteristics of the brain. Additionally, he argues that L2 learning is also a cognitive phenomenon, and cognition develop simultaneously and interdependently through the various stages of human beings’ life. Accordingly, a skilful communicative use of a language presumes the socio-cognitive competence of the user. Atkinson also believes that the external context can play a great part in developing and sophisticating L2 learning. Finally, Atkinson (2007, p. 169) argues that the dynamic, synchronised and congruent interaction between the different parts of the language and L2 learning cognitively and socially is maintained by complex means he calls “alignment”. Finally, he holds that the interaction between the various elements of the process, mind-body-world, occurs in an improvisational manner according to the emerging events of the context.

The context in the dynamic perspective of language seems to function as a filter that allows certain interpretations or actions and disapprove others. Tomasello (2003) argues that the range of the possible interpretations of a language unit is highly restricted by the social contexts in which they are utilised and the relationship between the speakers. In other words, the intended meanings of words, phrases, collocations, and idioms are constrained or determined by the social contexts they are used in and what is expected between the interactive parties, i.e., the speakers. Churchill (2007) resembles the role of context in restricting the potential interpretations of a situation to a nurse and her patient.

Informed by the context and the relationship nature that connects the speakers, the nurse and the patient, there are only limited interpretation possibilities of the language used in this context such as giving medicine, injections, taking blood samples, or checking blood pressure. In other words, the context can help L2 learners in the process of learning the target language vocabulary through minimising the number of interpretation possibilities which saves learners' time and effort, in accordance, makes learning vocabulary easier and faster.

Having applied the dynamic system on language and language learning, it can be argued that a word comprises not only a group of features such as "syntactic patterns, form, meaning, collocations, lexical relations with other words, and derivatives" (Laufer, 1997, p. 141), but also the relations that connect these features with the contexts in which they occur. In other words, a word means a system of relationships among the features and the contexts they are used within (Churchill, 2007). Additionally, applying the dynamic perspective appears to be more fitting into the process than taking solely one aspect such as social context or cognition isolated from the other elements which does not seem to provide a better account of the language learning process. However, in spite of the increasing evidence on the dynamic nature of language knowledge development, it has not become the dominant perspective among researchers up to date. Nevertheless, it is believed that the DST has the ability to contribute crucially to the cognitive development understanding which found to be promising in the field of the L1 and L2 development (Marin & Peltzer-Karpf, 2009).

According to Thelen & Smith (1996), the dynamic perspective of language implies interactive relationships among various aspects of word knowledge such as knowledge about the syntactic form classes and specific contextual properties that determine the potential interpretations of the target or novel words. This tradition views language and vocabulary learning as a non-linear development of multiple vocabulary knowledge types and processes in a dynamic system (Churchill, 2007; Larsen-Freeman, 2006). Larsen-Freeman (2006) argues that the dynamic view of language and language learning can provide a useful tool for depicting the progressing change in a L2 system as in accounting for how language use is adapted to the change in context.

However, van Geert (2008) contends that the dynamic approach does not aim at providing global straightforward answers or substituting the current approach of investigating language learning, development, or change over specific time scales. Rather, van Geert elaborates, it might be of fundamental efficiency, though it may not be straightforward and faced by various challenges, in exploring what lies beyond the linear structural links among variables or their characteristics and in understanding the development dynamics. Based on the dynamic view of language, an increasing number of studies have been conducted to the moment. In the following paragraphs, some of these studies will be reviewed.

Dynamically viewed, Verspoor, Lowie, and Dijk (2008) carried out a case study on an advanced learner to examine the potential role of investigating the intra-individual variability in L2 development in providing insights into the developmental dynamics of L2 learners. By utilising advanced visualisation techniques such as min-max moving windows, the study explored L2 systems during the occurrence of a quick progress in language acquisition. The micro-genetic analyses of the L2 development demonstrated that amount of development fluctuated by continuously progressing and regressing indicating the nonlinearity of the development patterns. The study also revealed that participants who used an abundance of advanced and various types of strategies in the initial tasks utilised more advanced strategies in the following tasks. Additionally, the results showed that subsystems i.e., levels of knowledge such as the lexicon and grammar, interacted actively with each other. For example, the appearance of a multiword sentence concurred with a lexical development in the one-word phase. In earlier phases, all resources were employed to advance the lexicon whereas in the next phase, more and different resources were necessary for developing the grammatical system that provided the functional distribution of information required for producing multiple-word utterances.

A further study was conducted by Bell (2009). The study closely accounted for the minute details of the learner's lexical knowledge development to identify the change in the learner's productive lexicon knowledge. The focus of the study was on a limited number of frequently used lexical items. To this aim, the researcher tracked learner's knowledge development of formulaic sequences

structures and the way they were incorporated into learner's lexicon. The study data consisted of 28 discursive essays in one genre. The results showed cases of inconsistency in the learner's lexical knowledge development such as fluctuations and variability in the learner's patterns of use. In other words, the use of some structures was dominating while some others were disappearing. Moreover, the use of intermediate structures and competing sub-systems were identified. The study results indicated a dynamic development of learner's inter-language and knowledge of vocabulary. In addition, it demonstrated a developmental pattern in the learner's knowledge progress of specific aspects of vocabulary. For example, it was observed that prominence or word significance had a role in acquiring or losing words. Additionally, the lexical and grammatical systems were in dynamic, constant interactions in a way that the development in a system resulted in advancing the other. In other words, development in learner's vocabulary knowledge leads to increasing learner's grammatical competence, i.e., producing more correct grammatical structure with fewer mistakes. However, learner's knowledge of these two aspects fluctuates and shows progress and regress and does not improve evenly due to the influence of various learning contextual elements such as the extent of exposure to the target language, input size, and learner's cognitive ability in terms of recalling and losing encountered items.

The dynamic relationship between the lexical and grammatical systems, but within a longitudinal study design, was also examined by Caspi (2010). The study tracked the development of lexical and grammatical variables of four female L2 learners of different L1 backgrounds; Indonesian, Mandarin Chinese, Portuguese, and Vietnamese over 36 weeks. The lexical and grammatical dimensions were investigated in terms of their complexity and accuracy. Lexical complexity was examined by adopting variation-based and frequency-based measures whereas lexical accuracy was measured as a general indicator of participants' correct use ratio of the lexical items. The identified errors included errors in word choice, collocation, and derivation. However, morphological verb errors were excluded since they were considered as indexes of syntactic accuracy. The syntactic complexity was measured by taking as indicators both of clause/sentence ratio (sentence complexity) and the ratio of subordinating conjunctions to the total clause number. The syntactic accuracy indexes were

measured by adopting local and general measures such as the ratio of correct word error per clause and correct use of subordinating conjunction per required context. The main focus was on English academic vocabulary suggested by the University Word List (UWL) (Xue & Nation, 1984) which is made up of 808 words and classified into 11 frequency bands, and the Academic Word List (AWL) (Coxhead, 2000) which contains 570 word families. These two lists were used to increase the database size to avoid or reduce the practice effects that might result from dealing with a limited number of words and minimise the frequency effects on the receptive-productive gap. Participants' vocabulary knowledge was measured by the Longitudinal Academic Vocabulary Test which comprised three levels: controlled production, recall, and recognition. Each test section consisted of 30 different words and from all frequency levels derived from the UWL and AWL lists database. Assessing their vocabulary free production was conducted by comparing the vocabulary used in the participants' four sets of 36 essays, 36 essays for each participant, with the combined academic word lists. The results indicated a strong connection between lexicon and grammar, and growth and variability patterns were alternating across these two dimensions. In other words, the study results revealed that lexicon development preceded syntax development and showed that lexicon development was a prerequisite for the syntactic growth and development. However, these two linguistic dimensions competed each other for getting the resources necessary for their development. The results also demonstrated an ordinal pattern in the relationships and development of the variables. In the first phase, learners promoted their lexical knowledge by producing more complex words, and secondly, they used these words more accurately. The same developmental pattern was identified with the grammatical aspect as learners' knowledge of syntax found to be more complex and then developed to be used more accurately. Up to the present time, this study is thought to be one of the main available modelling studies on the L2 development which investigates the interaction of varied subsystems in the L2 over time. However, recruiting a small number of participants, exclusively females, and being of four different L1 backgrounds might have over-complicated interpreting the relations among variables more than they were, accordingly, weakened the generalisability of the findings.

In line with examining the dynamic relationship between the lexical and grammatical aspects of language, Verspoor, Schmid, & Xu (2012) conducted a cross-sectional study on beginner and intermediate levels of proficiency. The study investigated 437 texts written by Dutch learners of English as a second language. The study aimed at exploring the role of a dynamic usage based perspective in yielding or founding of objective assessment measures of L2 learners' written texts and probing into the dynamic process of language development for gaining better understanding of the process. The results demonstrated that learners' knowledge and proficiency development could be assessed and distinguished by lexical and syntactic measures such as sentence length, use of tenses, dependent clauses, and errors associated with these uses. This might imply that learners' language system development mainly pivoted around the interactive relationship of their lexical and grammatical knowledge aspects. Additionally, the study results indicated a non-linear variation, change, and progress of knowledge and relationships among the variables under focus. It also, from examining the transitions between levels, showed a developmental and order pattern of learners' knowledge of the lexis and grammar. For example, the results revealed that the lexical development occurred between levels 1 and 2. The syntactic changes were noticed between levels 2 and 3, and both lexical and syntactic changes were observed between levels 3 and 4. Finally, only lexical development was identified in the transition between the levels 4 and 5 which seemed to occur mainly in terms of using particles, compounds and fixed phrases. Thus, more varied words and longer sentences were used along with increasing learners' knowledge and proficiency levels. However, the study did not differentiate among morphology, lexicon, formulaic phrases, including collocations, and constructions. This was because it was conducted on the assumption that there was no real difference among these different constructions and that all constructions were seen as constructions in a linguistic continuum. However, the acquisition and knowledge development of each of these constructions seemed to be different from each other. Nevertheless, the study results have its importance in mapping the developmental pattern of language knowledge development.

In a longitudinal single case study, Fitzpatrick (2012) investigated the developmental pattern of a learner's lexical knowledge who was studying abroad over one academic year. The study attempted to identify the influence of context on the overall vocabulary knowledge development. It also examined nature and linearity of knowledge development, and whether the context had any impacts on the acquisition of certain vocabulary knowledge aspects rather than others. The study data were elicited by utilising word association in six waves over an eight-month period to probe profoundly into the produced syntagmatic responses and accordingly into the development of the learner's productive knowledge of collocations. The study results indicated an uneven and gradual development of the learner's knowledge at certain areas over the year as in their number of collocations use and producing native-like combinations. In some other aspects such as form, form-meaning associations, and orthography, a noticeable inconsistency of development was observed. Commonly, the study indicated the non-linearity of vocabulary knowledge development. However, the small sample size of this study, one participant, limits the generalisability of the results. Nevertheless, the results have its significance in understanding nature of vocabulary knowledge development even at a small scale.

Dimensionally, some studies (e.g., de Bot, Lowie, & Verspoor, 2005, 2007; de Bot, Lowie, Thorne, & Verspoor 2013; Kramsch, 2002; Larsen-Freeman, 1997) suggest a dynamic, non-linear pattern of L2 learners' receptive and productive knowledge development. According to this view of language, levels of language knowledge, receptive vs productive, interact with each other in a complex way and they simultaneously compete for learner resources and support each other's development over time. Accordingly, the competition between both levels leads one of them to get more resources than the other and create the gap in knowledge between both knowledge levels (Caspi & Lowie, 2013).

In conclusion, it is sometimes argued whether to adopt the dynamic model or another alternative in conducting a developmental study. According to Thomas, McClelland, Richardson, Schapiro, & Baughman (2009) and Smith (2009), this might not be the appropriate option because each approach offers specific insights into the focus of the study. Accordingly, I would argue that many factors contribute to deciding on the adopted approach such as aims of the study,

available resources, time, fund, researchers' skills, or any social, economic, or political factors which may assist or hamper adopting a specific study design rather than another. For example, it may be more fitting for the current study to adopt a hybrid design that mixes between more than one perspective to make use of the advantages, techniques, facilities, and potentials they offer collectively to achieve the study aims which may not be feasible from relying on a single perspective.

2.7 Conclusion

Conceptually, the notion of collocation since its coinage as a linguistic term by Firth (1957) has been the focus of numerous linguists and researchers in both L1 and L2 studies. However, this linguistic phenomenon has been defined from various perspectives which can mainly be grouped into the frequency-based and phraseological definitions, and a third growing perspective that is the combination of the two definitions. Principally, these definitions have been based on the major features of this linguistic phenomenon. The main features which have been considered in defining collocations are frequency of co-occurrence of the collocational constituents, the syntactic structure of collocations, and the semantic transparency of the constituent items. Accordingly, one or more of these features, according to the adopted definition of the researchers, should be considered in any collocational investigation.

Typologically, studies in this field have mainly suggested two types of collocations: lexical and grammatical collocations which appear in various syntactic structures such as v+n, adj+n, prep+n, and phrasal verb patterns. Studies reveal that classifying collocations into types and patterns has its implications and significance in theory and practice. Classifying collocations can be helpful in recognising and understanding the semantic interpretation of collocations, identifying L1 users and L2 learners' developmental patterns of knowledge and use in corpus-based studies, identifying learners' errors, weaknesses, strength points, and their common characteristics. Additionally, they can be employed as tools for extracting collocations from large corpora, differentiating between collocations and the other types of combinations such as idioms and free combinations, and finally, for pedagogical purposes.

The concept of collocation can have its importance in the theoretical knowledge development of theories of language due to the implications these theories have

on the concept of collocation and vice versa such as its position regarding generative grammar, usage-based models, idiom principle, lexical priming, and the theoretical issues presented by this concept. Collocations can also be of use in a wide range of applications such as Natural Language Processing software, making dictionaries, machine translations, parsing phrases and corpus linguistic research, pedagogical implications, supporting cognitive fluency, and native-like production. In accordance, considering collocation and its typology is inevitably essential in any conducted study in this area of linguistics.

Dimensionally, knowledge of collocations is mainly divided into receptive and productive. These two aspects of knowledge are positively related to each other; L2 learners' high receptive knowledge is usually correlated positively with their high productive knowledge of collocations. However, the general notion is that L2 learners' productive knowledge of collocations lags far behind their receptive knowledge. Additionally, studies indicate that L2 learners' productive knowledge is a better indicator to their language proficiency than their receptive knowledge. Nevertheless, due to the difficulty of drawing a clear-cut border between these two aspects of knowledge, it sounds more comprehensive to include both aspects into considerations when evaluating learners' collocational knowledge because they are highly connected with each other and influence each other's development.

Finally, collocational knowledge development has been viewed as a linear and non-linear path way. However, studies indicate that language knowledge development, accordingly collocational knowledge, is not constant or changing all the way. Rather, the linear and non-linear development alternate through the different stages of language development. Though, recently, it has been increasingly believed that the different levels of language knowledge develop mostly in a non-linear way progressing and regressing all the way.

Chapter three

Factors affecting collocational knowledge and development

3.1 Introduction

Studies in the field of collocations indicate that L2 learners' collocational knowledge and their collocational knowledge development correlate at various degrees of strength to many factors. Generally, these factors include learners' personal characteristics, learning context, and features of collocations. However, due to the complex nature of collocations and the diverse views of this linguistic aspect of knowledge, researchers have varied in investigating the influence of these factors on L2 learners' collocational knowledge and development. For example, studies have been conducted according to types of collocations (e.g., frequency-based, phraseological, psycholinguistic, native speaker intuition-based, grammatical approaches), or based on types of study designs (e.g., cross-sectional vs longitudinal), or according to types of acquisition models (e.g., linear, non-linear, dynamic). In the following sections, an overview of the main conducted studies and the results they came out with along with demonstrating the major factors which are evidenced to be affecting L2 learners' collocational knowledge and development will be presented.

3.2 Learners' characteristics

The personal characteristics which may have impacts on the learners' collocational knowledge and development include gender, age, aptitude, motivation, attitude, experience, and learning strategies. Concerning the influence of gender on the L2 learners' knowledge and development of collocations, it can be generally argued that use of language is normally adjusted according to the social context, where, when, why, and with whom language is used (Jule, 2008). However, researcher varied in their perspective of the influence of gender on the language users. Studies have generally indicated no significant influence of gender on the L2 learners' collocational knowledge and development (e.g., Ashiyan & Salehi, 2017; Ganji, 2012; González Fernández & Schmitt, 2015; Reza & Tabrizi, 2016). However, very few studies have shown a significant influence of gender on the L2 learners' use of language. One of these studies is that was conducted by Mazaji's (2015). The study results indicated that males outperformed females in their collocational knowledge and development. In another study by Schwartz et al.

(2013), the results indicated that males collocated significantly the possessive 'my' with 'girlfriend' or 'wife' more than females do for 'boyfriend' or 'husband'. According to the study results, the statistically significant correlation provided quantitative evidence that males behaved linguistically differently from females. Moreover, some linguists have gone even far from this in claiming an innate male ability to outperform females in using language (e.g., Jespersen, 1922). Jespersen provides evidence for this by arguing that the greatest orators of history were mostly men rarely women, if any, due to their higher innate linguistic ability. Accordingly, if it can be well-established or proved that females use language differently from males, I would claim that they learn and develop their knowledge of language, including collocations, in a different way from male learners.

As far as motivation, attitude, and language aptitude are concerned, studies have commonly highlighted the positive influence of these factors on the L2 learners' collocational knowledge and development (e.g., Dörnyei, Durow, & Zahran 2004; Dörnyei & Skehan, 2003; Sawyer & Ranta, 2001). Schmitt, Dörnyei, Adolphs, & Durow (2004), for instance, conducted a longitudinal study on 94 course participants. The study explored the influence of age, gender, language aptitude, and motivation on the learners' receptive and productive knowledge of formulaic sequences, including collocations. Participants were of two groups. The first group consisted of 62 two-month course length participants and the second group included 32 three-month course length participants who had enrolled in an EAP course. The target formulaic sequences for the study were selected based mainly on three criteria. First, the target formulaic sequences should have occurred in language use with some degree of frequency. Second, they should have been used in academic discourse contexts. Third, they should have been pedagogically useful for students and worth teaching. Accordingly, 97 candidate academic formulaic sequences were extracted from the Biber et al. (1999) analysis of lexical bundles, and 59 candidate formulaic sequences from Nattinger & DeCarrico's (1992) functional analysis of lexical phrases. Words were taken from Hyland's (2000) list to express doubt and certainty such as *dearly* and *approximately* or used as discourse markers such as *therefore* and *finally*.

A battery of tests were used in this study in a specific order to avoid any influence from these tests on the participants in conducting the other tests as follows: productive formulaic sequences, aptitude, attitude/ motivation, Vocabulary Levels Test 3000, Vocabulary Levels Test 5000, and receptive formulaic sequences. For measuring participants' productive knowledge of formulaic sequences, a mixed format test of cloze and C-test techniques was used. For testing their receptive knowledge, the same contextualisation stories which were used in the productive test were used in the receptive test with inserting a single line in place of the target formulaic sequences, and participants were asked to choose from four options in a multiple-choice test format. The study results indicated that EAP learners at advanced levels of proficiency had a command of a considerable number of academic formulaic sequences. Additionally, the results revealed that learners' knowledge development of formulaic sequences over the course took two forms. The first one was in the form of learning new formulaic sequences while the second was in the form of promoting their mastery of the known ones to advance their receptive knowledge to the productive level. Interestingly, the study revealed no significant influence of the individual differences such as aptitude, attitude, and motivation factors on their receptive and productive knowledge development. The researchers justified this unexpected result by the non-linear nature of the relationship between the acquisition of formulaic sequences and learners' characteristics. Additionally, they argued that the influence of these factors might have been modified by other factors in the learning context. Accordingly, they suggested a study with a longer period to be able to identify the influence of the individual differences of learners on the knowledge development of formulaic sequences.

Dörnyei, Durow, & Zahran (2004) conducted, in parallel with the above reviewed study, a longitudinal qualitative case study to investigate the influence of the individual differences such as language aptitude, attitudes, motivation, in addition to the contextual factors, on the L2 learners' acquisition of formulaic sequences. The sample included seven Asian, Japanese and Chinese, students who participated in a two-month or three-month pre-session intensive language courses at the University of Nottingham. The participants were those who had gained extreme scores on the two types of formulaic

sequence tests among 24 students. While three of the participants demonstrated no development in their formulaic sequence repertoire between the pre- and post-tests, the other four participants showed significant improvement during the examined period. Data were obtained from the participants through taking pen-and-paper tests and long longitudinal interviews. Participants who took part in the two-month course were interviewed at the beginning and end of the course while those who were on the three-month course were interviewed at the beginning, middle, and end of the course. Data were analysed in an ongoing manner throughout the longitudinal data collection phase. The transcribed interviews were analysed employing content analysis to identify the potential themes which would explain the formulaic language acquisition success or failure in the observed individual participants. The study results revealed interactive influences of three elements on the learners' acquisition of formulaic sequences: learners' language aptitude, motivation, and learners' socio-cultural adaptation capability. Additionally, the last factor, learners' socio-cultural adaptation capability, seemed to be more influential than the other two factors, and it could hardly be compensated by high linguistic ability and motivation to learn the language. In other words, L2 learners' involvement in the learned social context had a significant role in the development of learners' knowledge of formulaic sequences, accordingly, collocations.

In conclusion, the general direction refers to that learners' characteristics have influence on their collocational knowledge and development. However, the degree of influence of these factors vary from a learner to another due to contributing various elements in the learning process. Additionally, the relationship between the personal characteristics and learners' collocational knowledge is not direct or straightforward but they rather interact with each other in an intricate network that may not be noticeable in short period studies.

3.3 Contextual factors

Learners' collocational knowledge and development can be affected by various contextual factors. Contextual factors can include various elements such as access to native speakers, years of study, degree of explicitness in teaching and learning collocation, topic familiarity, use of references, technological

means and social media networks, curricula textbook settings, and language of instruction.

The influence of context of learning on the learners' vocabulary knowledge has been evidenced in the L1 and L2 research literature (Hu, 2013). In line with, Webb (2008) states that characteristics of the context has a substantial impact on acquiring vocabulary meaning. Hu (2013) maintains that contextual richness has a considerable influence on developing L2 learners' information about the form-meaning connections and grammatical functions of vocabulary.

Additionally, he argues that while L1 learners acquire vocabulary and the related aspects of knowledge such as form, meaning, and contextual information in a rich context, L2 learners lack this richness in context such as sufficient exposure to the target language. Consequently, L2 learners' vocabulary knowledge lags much behind that of the L1 learners.

Apart from this, learning context appears to be to an extent pertinent to learners' knowledge. Research in this respect indicates the importance of the dynamic interactions among an array of elements that contribute to developing learners' proficiency. Segalowitz & Freed (2004) contend the central role of the learning context in developing learners' linguistic skills specifically their oral proficiency. They argue that learners who study abroad in the target language context are more fluent and efficient in terms of speed of vocabulary access, hesitation, automaticity, and proper use of expressions than those learners who study in local contexts.

Collocationally, the school environment which includes teachers, teaching methods, and educational curricula programmes can have an important role in developing L2 learners' collocational knowledge. Viewing from a formal teaching perspective, Henriksen (2013) argues that teachers' overemphasising on decontextualised individual words through the conducted tasks and assessments and unavailability of sufficient materials for raising L2 learners' awareness of the existence of collocations are thought to be among the factors which result in learners' deficiency of collocational knowledge. Similarly, Gaballa & Al-Khayri (2014) confirm the influence of the learning context on the learners' knowledge of collocations; learners who study in the ESL context show higher collocational knowledge than the EFL learners.

Likewise, Alsulayyi (2015) examined the productive collocational knowledge of

ten Saudi undergraduate students in two different contexts. Five of the participants were students at the school of English at Manchester Metropolitan University in the United Kingdom whereas the other five participants were students at Najran University in the local context. To this aim, the study utilised analysing essays written by the participants of both groups. The results showed that that learners who studied in an ESL context did better than those who studied in an EFL context. This was demonstrated in that learners appeared to make less grammatical collocation errors in an ESL environment than those who studied in an EFL context. According to the study, L1 interference had a great part in producing unacceptable collocations.

Koya (2005) concludes, from contrasting school teaching textbooks with English corpora, that collocations are either not sufficiently represented in the teaching textbooks or represented with low frequencies. Moreover, their use is usually decontextualised due to the adopted teaching approaches by educationists and deficiency in the curricula designs. Accordingly, to gain the high level of proficiency required for academic purposes, more contextually appropriate procedures are required such as incorporating the most frequently used collocations, in accordance with the intended genres, in language learning programmes and they should take the ultimate priority in teaching (Ellis, Simpson-Vlach, & Maynard, 2008).

Another factor which may influence learners' collocational knowledge is learners' years of studying collocations (González Fernández & Schmitt, 2015). There appears to be a positive correlation between the length of the study period and learners' knowledge of collocations and its development (Ganji, 2012). However, this development seems to be a slow process that lasts for many years. Maram (2009) and Ebrahimi-Bazzaz, Abd Samad, Ismail, & Noordin (2014) assert that, concluding from their studies, there is a positive relationship between years of study and learners' collocational competence, and that in order to identify a noticeable advancement or development in L2 learners' knowledge of collocations, it requires at least four academic years. Similarly, Siyanova-Chanturia (2015) suggests a period of five month as sufficient exposure for beginner learners to be able to demonstrate native-like use of collocations. Additionally, in her study, Nesselhauf (2005) highlights the

influence of length of exposure to the learned language and attributes the errors made in using collocations to the influence of the length of exposure to the L2.

Timing of introducing specific collocations to L2 learners, i.e., the time learners are expected to be familiar with specific collocations, is a further contextual factor that is suggested by Laval & Bernicot (2002). According to them, timing is important specifically when learners are expected to be familiar with certain expressions that have pragmatic functions in their social communications at specific level of language proficiency. Accordingly, Nesselhauf (2005) suggests starting teaching collocations at earlier levels of schooling to provide a rather longer exposure to the target language. Otherwise, L2 learners will be in short of the required collocational knowledge for that level, consequently, they will produce improper collocations.

In addition, explicit instruction to raise learners' awareness of collocations is found to have a significant impact on learners' knowledge of collocations (Ashouri, Arjmandi, & Rahimi, 2014; Liu, 2002; Mahvelati & Mukundan, 2012). A further factor is language of instruction which is seen as another source of increasing L2 learners' exposure to the learned language and it can positively affect learners' knowledge and development of collocations. Bergström (2008) believes that reinforcing English through using English as a language of instruction has been found to develop learners' collocational knowledge as instruction in the target language enhances their receptive collocational knowledge and accordingly leads to higher performance in learners' linguistic ability. Additionally, Kameli, Mostapha, & Baki (2012) hold that using English as a language of instruction motivates L2 learners to learn English, and it increases L2 learners' confidence, vocabulary knowledge, and receptive skills.

In line with the influence of explicit instruction on learners' knowledge and development, Reza & Tabrizi (2016) conducted a quasi-experimental study to explore the influence of explicit instruction of collocation via SMS on Iranian intermediate EFL learners' collocational knowledge. The study was conducted on 75 participants. The participants included 25 males and 25 females who were divided into two experimental groups while 25 participants of both genders were assigned as a control group. Participants were instructed about 90 English collocations which included their definitions given with examples through SMS messages for 5 weeks. The control group was given the same collocations on

paper inside the classroom. Pre-test and post-test were conducted to examine the participants' collocational knowledge before and after the experiment. The results of the post-test indicated that participants' collocational knowledge of the experimental group was higher than that of the control group. In other words, the explicit instruction of collocations resulted in increasing learners' awareness and knowledge of collocations.

In addition, Choi (2016) conducted an experimental study, using the eye-tracking paradigm, on 38 Korean college students, 14 males and 14 females. The study aimed to explore the influence of the textual enhancement, boldfaced collocations, on the participants' collocational knowledge. The study results revealed that the typographically enhanced text had a positive effect on the participants' post-test performance. However, it did not show a significant influence on their processing of familiar collocations. This study draws attention to the important role of raising learners' awareness and the explicit instruction of collocations.

However, Zheng (2014) argues that contextual factors such as the pedagogical and socio-cultural contexts such as teachers and the curriculum appeared to have impacts on the participants' learning experience. For example, while teachers' class instruction concentrate on the paradigmatic aspect of vocabulary learning, they seem to neglect the syntagmatic aspect and ignore cross-linguistic semantic differences. Consequently, due to the curriculum and teacher's requirements, the output opportunities sound poor both in terms of quantity and quality.

A further influential factor is engaging with English through using hardware and software references. These include books, references, digital means such as electronic dictionaries, and engaging in social networking websites such as Facebook, Twitter, and WhatsApp (González Fernández & Schmitt, 2015). Using digital dictionaries through various types of devices such as computers, laptops, tablets, and mobiles has been proved to be useful in acquiring new vocabulary and specifically during learners' independent reading inside and outside the classroom because they allow them to consult dictionaries anywhere and at any time (Hamilton, 2012). Additionally, it is argued that utilising electronic dictionaries such as COBUILD online are more effective than paper dictionaries in retaining vocabulary meaning and particularly collocations

(Dziemianko, 2010). Practically, bilingual dictionaries seem to be the most utilised type by L2 learners and this type of dictionaries is regarded more effective in learning individual words and collocations than monolingual dictionaries (Takahashi, 2012). Considering the contextual circumstances of learners in terms of the availability of technological facilities or not, using dictionaries is one of the widespread tools used in managing meaning, collocations, and collocability of items.

From a dynamic perspective, Churchill (2007), who functioned in this study as an author and as an American learner of Japanese, conducted a longitudinal phenomenological case study research on his three-month diary of learning Japanese vocabulary. The study examined learner's lexical knowledge development of mainly individual words and secondly collocations at the micro-level. Additionally, the micro-processes by which multiple types of word knowledge became incorporated into the lexicon of a learner of Japanese were documented through describing and explaining learner's experience of learning the different types of word knowledge such as meaning, phonology, and most common usage. The study data included author's diary of the vocabulary learning experiences over the three-month course of learning Japanese and the diary was triangulated with MD recordings and relevant documents such as lab results from his stay in the hospital. The diary consisted of 78 pages in which details of the author's exposure to a lexical item was documented through the course. Descriptions of the events were documented shortly after their occurrence or observation. The author described his concentration, as a learner, on a particular compound, "saiketsu" *draw blood* which appears in collocations such as "ketsuekikensa" *no tame* "ni saiketsusuru" *to draw blood for the purposes of a blood test* and how he experienced exposure to this lexical item, and learned its phonological, orthographic forms, and common senses. Additionally, he documented how associations were made between a lexical item and other items such as making compounds or collocations. The study indicated a dynamic nature of word interpretations and presented an order in the knowledge development. To elaborate, the learner, on one level, started to distinguish between the word "saiketsu" and the related words or terms such as "chi wo toru" and "ketsuekikensa" which were encountered in the learning process. On the orthographic level, the learner could write the compound

without copying from existing examples. Accordingly, his orthographic competence development increased his ability to identify other words formed with the same compound. The study showed a dynamic nature of vocabulary knowledge progress through incorporating various word knowledge types in the learning process. It also demonstrated the significant role that the learning context interactions could play in developing learners' intricate lexical network. However, the short time span of the study, the single case design and being the author and the learner at the same time might have resulted in data bias, consequently, affecting the generalisability power of the study findings. The study could have yielded stronger evidence if the author and the learner had been separate and conducted with larger population size and used more data collection instruments.

Finally, despite the controversial role of the context, the psycholinguistic, cognitive tradition (e.g., de Bot, Paribakht, & Wesche, 1997; Hulstijn 1997; Schmitt 2000) vs the ecological, sociocognitive tradition (e.g., Atkinson, 2002; Atkinson, Churchill, Nishino & Okada, 2007; Kramsch, 2002), it is admitted, even from a psycholinguistic perspective, that the context has a significant contribution to the learners' linguistic knowledge development. Thus, context can account for a part of the understanding of learners' knowledge and development during their effort to learn a language. Accordingly, it may be of significance for researchers to be aware of the influence of context on learners' knowledge in their investigations into this aspect of language.

3.4 Features of collocations

The influence of features of collocations on learners' collocational knowledge and development is well evidenced in research literature (e.g., Durrant & Schmitt, 2009; Gitsaki, 1999; Kellerman, 1978; Koya, 2005; Kurosaki, 2012; Nesselhauf, 2005) have provided evidence to support that L2 learners' knowledge of collocations and knowledge development are affected by these characteristics. In common, studies which have dealt with features of collocations can be divided into those researching the actual or tangible aspects of collocations such as frequency of occurrence of collocations in textual contexts, form, spelling, and phonology, and those investigating the abstract aspects of collocations such as their semantic and syntactic restrictions, culture and pragmatics, or congruency with L1 (Gledhill, 2000). Apparently, many

features have been attributed to collocations. However, for the purposes of the current study, the concentration will only be on the key features which have been evidenced to have considerable impacts on the L2 learners' collocational knowledge and development.

3.4.1 The influence of frequency

In general, the feature of frequency of occurrence of collocations has been used to perform various functions. First, it has been used as a tool in identifying and defining collocations, as was explained earlier in Chapter two (Section 2.2.1), in studies of Firth (1957), Sinclair (1991), and the others. Second, it has been used in some other studies in investigating the influence of the frequent occurrence of collocations on L2 learners' collocational knowledge and development (e.g., Alternberg & Granger, 2001; Hama, 2011; Shehata, 2008; Wolter & Gyllstad, 2013). Third, frequency of occurrence has been employed to identify any external factors such as the context in which the language is learned that may affect learners' collocational knowledge and development (e.g., Alsakran, 2011). Fourth, frequency has also been used to explore the nature of using certain patterns of collocations by L2 learners at particular levels of language proficiency (e.g., Laufer & Waldman, 2011). Finally, frequency has been utilised as a tool for making register-related or discipline-related lists that can be employed within pedagogical contexts (e.g., Durrant, 2008; Shin & Nation, 2008). Functionally, Durrant & Schmitt (2010) differentiate between two types of repetitions, i.e., frequency of occurrence of collocations. The first type is the fluency-oriented repetition which is conducted through repeating individual sentence contexts. The second type is repeating the items in different contexts which increases the cognitive burden. They note that repeating the same individual sentence contexts, fluency-oriented activity, has a greater influence on L2 learners' collocational knowledge than the recurrent exposure to the same collocations in different contexts.

The versatile uses of frequency may demonstrate the significance of this feature and the substantial role it can play in studying the various aspects of this linguistic phenomenon, i.e., collocation, and L2 learning. It may also explain its proponents' emphasis on the use of this feature in defining, identifying, and extracting collocations processes, and collocational knowledge and development investigations.

As far as L2 learners' collocational knowledge and development are concerned, frequency of occurrence of collocations seems to be one of the most influential features affecting their collocational knowledge and development (Gyllstad & Wolter, 2015). It is argued that frequency of linguistic forms has a significant effect on native speakers' and L2 learners' linguistic representation and acquisition patterns, accordingly, on their performance (Bybee & Hopper, 2001; Mueller, 2011; Myles, Hooper & Mitchell, 1998). The fundamental underlying assumption is that language users in general and L2 learners in particular respond to or are affected positively by the frequent occurrence and repeated exposure to lexical items in their written and spoken linguistic inputs.

Additionally, Ellis (2002) holds that frequency has an effective role in processing the various aspects of language such as formulaic language, phonology, spelling, lexis, language comprehension, grammaticality, sentence production, and syntax. Accordingly, through the recurrent exposure to language, L1 users and L2 learners acquire the different phonological, semantic, syntactic, and cultural aspects of language knowledge (Bybee, 2008).

Kirsner (1994) contends the significance of providing sufficient exposure to the lexical items of which L2 learners lack adequate experience, i.e., the required frequency rate of exposure to collocations is necessary to establish a native-like linguistic proficiency. In line with this, Kennedy (2003) contends that frequency of experiencing collocations increases significantly the process of storing and recalling these linguistic items. Furthermore, establishing on a meta-analysis of nineteen earlier study tests of collocations to examine the relationship between learners' knowledge and frequency data, Durrant (2014) concludes that L1 collocation learning is frequency-driven and this frequency appears to correlate, though the link strength varies across the corpora, moderately with learners' knowledge of collocations.

However, Howarth (1998) argues that the automatic quantitative analyses of the frequency-based studies, which are assisted by advances in computer technology and development of large corpora, generally concentrate on L1 language users and L2 learners' performance, i.e., collocation production, and they rather neglect competence, i.e., receptive knowledge of collocations, from their analyses. Alternatively, he asserts that it is highly important to consider the processes underlying performance such as memory storage and production.

According to him, the significance of these processes comes from that it is essential to consider the kind of the stored lexical combinations and the way they are processed in the brain to produce collocations since learners' mental lexicon may hold more abstract lexical units than that can be calculated by statistical measures. Moreover, he argues that neither L1 users nor L2 learners produce combinations relying on their frequency and probability of co-occurrence. Furthermore, reliance solely on frequency may emphasise on unproblematic combinations such as a transparent collocation like *have children* which may occur recurrently due to the characteristic of the selected text topic. Accordingly, it necessitates that the concept of the phraseological significance should take into consideration the differences between the phraseological types and the way they are produced by L1 users and L2 learners and it should not rely on frequency only. However, Howarth's (1998) claim that frequency-based studies rather disregard learners' receptive knowledge may not be quite accurate since these studies seem to pay attention to this aspect of learners' knowledge as well, for example, Durrant's (2014) study mentioned earlier was a frequency-based study to investigate learners' receptive not productive knowledge.

From a phraseological perspective, Gitsaki's (1996) examined the productive collocational knowledge of 275 Greek high school students, who were studying English as a second language, through tracing the knowledge development of 37 patterns of lexical and grammatical collocations. Data were collected from the learners at three different levels of proficiency (post-beginners, intermediate, and post-intermediate). Data collecting instruments included three tasks: essay writing, translation test, and blank-filling. The essay writing was assumed to provide evidence of accurate free production of collocations while the translation and blank-filling tests were assigned to measure accuracy of the participants' collocational knowledge in cued production tasks. Examining data was in terms of the differences in accuracy that existed between and within groups at the three different levels in the tasks performed. Determining the significance of the observed differences between the groups and within the groups was carried out by employing statistical measures. Additionally, implicational scaling analyses were utilised to identify accuracy orders in the acquisition of collocations. The study results indicated that frequency of input

had a positive influence on the development of the productive collocational knowledge; an increase in the amount of exposure to a particular collocation type in the learners' textbooks resulted in developing their knowledge of that type or pattern of collocations. Additionally, the results provided evidence for that the frequent occurrence of a particular collocation type via textbooks could have an impact on the acquisition of that particular collocation regardless its frequency in the everyday speech contexts.

In a phrasal frequency-based framework, Alternberg & Granger (2001) conducted a corpus-based study to investigate EFL learners' productive knowledge of high frequency verbs and compared their use with a native speaker learner corpus. The study concentrated mainly on eight uses of the verb MAKE: *produce* something, delexical uses, causative uses, *earn (money)*, link verb uses, *make it* (idiomatic), phrasal/prepositional uses and other conventional uses such as *make good*. An answer was sought to know whether French and Swedish L2 learners overuse or underuse high frequency verbs compared with native speakers specifically MAKE and how accurately L2 learners use these verbs in their writings. In other words, the study examined the influence of frequency on the use of the verb MAKE with its collocates. To achieve the study aims, authentic learner corpora were compared with the native learner corpora. The learner corpora consisted of two corpus samples taken from The International Corpus of Learner English database. The first contained 170000 words of essays written by advanced French-speaking learners of English and the second had the same size and text genre of essays written by advanced Swedish learners of English. Both groups of learners were 2nd and 3rd year university students. The native control corpus, English, consisted of a 170000-word sample taken from the Louvain Corpus of Native English Essays database which contained essays written by native-speaker American students. The learner essays were approximately 600 words long whereas the native-speaker essays were about 800 words long and all essays were argumentative and non-technical. The linguistics software WordSmith Tools was used as a concordance instrument to facilitate the data analysis. The study results indicated that the use of high-frequency verbs was a difficult challenge ahead of L2 learners even at high proficiency levels. Consequently, learners' usage rate of this category is significantly less than that of the natives.

This result indicated various interpretations. First, it might indicate that, assuming that frequency or length of exposure to these collocations was sufficient, frequency of occurrence did not seem to be the only factor affecting L2 learners' collocational knowledge and development such as the influence of L1-L2 differences. Second, the frequency of occurrence, i.e., length of exposure, of these collocations was not high enough to develop learners' knowledge to produce native-like collocations. Third, the frequency rate was not sufficient and the influence of some other factors were present. Accordingly, the study results implied to conduct more investigations to explain and identify the influential factors behind L2 learners' collocational knowledge deficiency because some factors other than frequency seemed to affect learners' knowledge of collocations.

In support to studies such as Gitsaki's (1996), Shehata (2008) explored the effect of the amount of exposure to the target language on the collocational knowledge of advanced L2 learners. The study results revealed a moderate correlation between learners' collocational knowledge and the amount of exposure to the learned language. In another study highlighting the importance of frequency of input to L2 learners' collocational knowledge, Durrant & Schmitt (2010) tested the non-formulaic model of L2 learning. This model claims that L2 learners learn words individually without keeping in mind the collocational information of the learned words, i.e., they adopt a non-formulaic approach to language learning (e.g. Wray 2002). However, the study findings, in contrast to the non-formulaic model, indicated L2 learners' mastery of considerable collocational knowledge about what words appear together in their input. Additionally, they argued that any lack in L2 learners' collocational knowledge could be attributed to insufficient exposure or frequency of the input and not to a non-nativelike approach to learning.

Similarly, Hama (2011) conducted a study to identify factors influencing Kurdish EFL learners' productive collocational knowledge deficiency. The study was carried out on 40 Kurdish fourth year students (24 males and 16 females) who were studying EFL in the English Language and Literature department at Koya University in Kurdistan Region. The collected data were quantitative and qualitative. The quantitative data was collected by running a multiple-choice collocation completion test which was used to spot learners' major sources of

producing incorrect collocations. The test included 75 items of three different patterns of collocations: v+n, adj+n and v+prep. Participants were asked to complete collocations with the most appropriate missing items from the options given in a sentence context. The qualitative data was gathered through think-aloud protocols which aimed to identify possible main sources of producing non-native like collocations. Six participants were interviewed to express what they were thinking about while doing the collocation completion test and the transcripts of these spoken records of mental process were analysed into patterns. The six selected students were chosen based on their scores; two top scores, two average scores, and two poor scores in the collocation completion test. The study results indicated that frequency of exposure to the targeted collocations was one of the major factors affecting their productive collocational knowledge deficiency; there was a positive relationship between frequency and learners' knowledge and use of collocations. Additionally, the study revealed that frequency of the collocational components and mutual information showed no significant contribution to produce native-like collocations.

Siyanova-Chanturia (2015) conducted a longitudinal frequency-based study on collocations in beginner learner writings. The study aimed at investigating the use of n+adj (=adj+n in English) collocations of 36 Chinese beginner learners of Italian. The results indicated a significant development in the learners' use of Italian n+adj collocations in that learners of higher proficiency levels seemed to have knowledge of collocations of higher frequency and stronger association larger than knowledge of learners of lower levels of proficiency of the same type of collocations. In other words, frequency of occurrence seemed to have a positive influence on the learners' productive collocational knowledge development both in number of collocations and quality of item associations as learners' use appeared to become more native-like by the end of the language course through using, in addition to the higher frequency items, more lower frequency collocations.

Szudarski & Carter (2014) conducted an experimental study on 41 Polish EFL learners. The study aimed at examining the impact of input frequency and input enhancement on the learners' knowledge of infrequent collocations of the patterns v+n and adj+n imbedded in stories to be read during three weeks. Participants' receptive and productive knowledge of the collocations were

examined through a battery of five delayed tests which were developed based on Laufer & Girsai (2008) and Web et al. (2013). The results revealed that adopting a mix of both input flood and input enhancement could help promote learners' collocational knowledge. However, the input plus instruction strategy resulted in assisting learners' collocational knowledge development only at the form recall and form recognition levels. According to the study results, although increasing exposure through repeating the same collocations over time resulted in developing collocational knowledge, it did not necessarily imply promoting learners' collocational knowledge at all levels. This implies that various strategies and procedures needed to be taken in order to develop all aspects of collocational knowledge. Accordingly, no single method or approach is able to enhance all aspects of learners' knowledge of collocations.

From a frequency-based perspective, Nguyen & Webb (2016) examined 100 Vietnamese EFL learners' knowledge of v+n and adj+n collocations of the first three 1000-word frequency levels. Participants' proficiency level, who were university students majoring in English language, ranged between pre-intermediate and upper-intermediate. The study investigated also the influence of node word frequency, and collocation frequency on learners' receptive knowledge of collocations. The results indicated that the learners were in short of collocational knowledge at all word frequency levels; participants did not seem to have approached the mastery level of collocational knowledge of words at any tested level of frequency or collocation type. Moreover, their collocational knowledge decreased significantly at all levels. Additionally, the node word frequency appeared to be the strongest predictor of the participations' receptive collocational knowledge.

Driven by the results of the reviewed studies (e.g., Durrant, 2008; Durrant & Schmitt, 2010; Gitsaki, 1996; Hama, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Wolter & Gyllstad, 2013), it turns out that a reasonable connection exists between the frequency of occurrence of collocations and learners' collocational knowledge and development. In accordance, the rate of frequency of occurrence of collocations can be regarded as an indicator to anticipate the mental representation of collocations, accordingly, as a dependable criterion for testing learners' knowledge of collocations (Durrant, 2008, 2014).

However, the connection between frequency and collocational knowledge and development still seems to be a fuzzy relationship. Regardless the maintained significance of frequency in learning a language, Durrant & Schmitt (2009) argue that while L1 learning is thought, according to usage-based models, to be established on frequency-based analyses of memorised phrases, it is not obvious if the same claim could be applied to L2 learners. Observing from their study, they found that L2 learners depended significantly on the use of high-frequency collocations in their writing whereas they seemed to be underusing less-frequency collocations, though these collocations might have been of strong associations and salient for native speakers. For this, they recommended taking appropriate frequency information and individual differences into account while analysing L2 learners' collocational knowledge.

A further ambiguity in the relationship between frequency and collocational knowledge and development is that influence of frequency may be more important in developing certain aspects of collocational knowledge than others since it does not appear to be developing all aspects of collocational knowledge in parallel. For example, Pigada & Schmitt (2006) claim the least influence of frequency of exposure on learners' acquisition of vocabulary forms. Additionally, Yamashita & Jiang (2010) argue that learning incongruent collocations remains difficult even with considerable amount of exposure to L2. In line with this, Laufer & Waldman (2011) argue that L2 learners do not notice collocations which consist of frequent and transparent items such as *make a decision*. Consequently, increasing learners' exposure to these collocations in the inputs may not result in increasing learners' knowledge of such collocations. In sum, this uncertainty in the role of frequency in collocational knowledge may imply involving the influence of some other elements in the L2 learners' knowledge and development. Accordingly, it needs more investigations into the influence of frequency and its relationship with collocational knowledge and development.

3.4.2 The influence of the syntactic structure

Structurally, there is evidence in the literature that the syntactic structure of collocations seems to have a significant relationship with L2 learners' collocational knowledge and development (e.g., Alsakran, 2011; Alsulayyi's, 2015; Gitsaki, 1996; Koya, 2005; Shehata, 2008; Shokouhi & Mirsalari, 2010). In other words, collocations of certain syntactic structures appear to be

more/less challenging than others. Accordingly, L2 learners hold mastery of collocations of certain grammatical structures prior to collocations of other structures.

A seminal work on the influence of the syntactic structure of collocations on the learners' collocational knowledge is that was conducted by Gitsaki (1996). The study results indicated that learners' collocational knowledge development was under the effects of more than one factor of which the syntactic structure of collocations seemed to be one of the essential influencing factors. She found that collocation types which were structurally different from learners' L1 were more challenging to translate than those had similar syntactic structures. The influence of the syntactic structure of collocations on the learners' knowledge of collocations was also evidenced in that it revealed a developmental pattern in the learners' productive knowledge of collocations across and within the different levels of proficiency. The study revealed a sequential acquisition of the collocations according to the complexity of their syntactic structures in that grammatical collocations of simple structures seemed to be easier to learn than lexical collocations. Gitsaki's (1996) study is considered as one of the leading studies in suggesting the acquisitional and developmental order in learning collocations which can have significant theoretical and pedagogical implications in learning and teaching collocations. The study revealed that learners' development of collocational knowledge underwent three phases. Initially, learners began to learn or acquire collocations as unanalysed lexical units where they correctly produced or used more lexical collocations than collocations of more syntactic structure complexity. At the second phase, as learners' grammatical knowledge advanced, their use of lexical collocations comparatively decreased, their use or production of grammatical collocations increased. Finally, in the third phase, learners' competence of using both lexical and grammatical collocations increased and they became able to employ both types of collocations in a more native-like manner and they advanced to the next proficiency stage. According to the study, the zigzag movement from a syntactic structure to another till learners became capable of using and producing native-like collocations indicated the influence of the syntactic structure on learners' collocational knowledge and development.

Similarly, the influence of the syntactic structure on the L2 learners' collocational knowledge was evidenced in Koya's (2005) study but it varied according to aspect of knowledge. The study results highlighted the syntactic structure of collocations as a significant, influential factor for learners' productive collocational knowledge; learners' productive knowledge of structurally simple collocations was higher than their knowledge of the collocations of complex syntactic structures. However, the results indicated no significant impact on the learners' receptive knowledge. Additionally, he argued that the multiple-choice format of the receptive collocation test minimised the influence of the syntactic structure of collocations on the learners' scores whereas its influence appeared more obviously in the productive test due to the test format used, i.e., the translation test format, which required more syntactic information from the participants. Although these results highlighted the influence of the collocational structure, it also pinpointed the impact of other factors on the knowledge and development such as context and way of eliciting learners' knowledge. Accordingly, it indicated the intricate nature of learning and assessing learners' collocational knowledge and development.

Although the impact of the syntactic structure on L2 learners' collocational knowledge was evidenced, Shokouhi & Mirsalari's (2010) study results indicated that grammatical collocations were more challenging to L2 learners than lexical collocations which was in contrast with Gitsaki's (1996) study results. The findings revealed that collocations of certain structures such as n+prep appeared to be harder to master than others such as n+v collocations. This indicated that the grammatical structure of collocations had a role in the L2 learners' collocational knowledge development of certain grammatical structures and hindering the collocational knowledge development of some other patterns of grammatical structures. This result accorded the study findings of Mahmoud's (2005) and Bahardoust & Moeini's (2012) studies which revealed that lexical collocations were less challenging and used more in the L2 learners' writings than grammatical collocations. The result indicated that the syntactic structures of the lexical collocations make less challenges for learners than the syntactic structures of grammatical collocations. Additionally, by comparing L2 learners' use of different categories of lexical collocations in their writings, the

results showed that learners' use of the v+n and adj+n collocations outnumbered their use of n+v collocations.

In line with the significant influence of the syntactic structure on learners' collocational knowledge, Shehata (2008) concluded from her study that the collocational patterns had impacts on the learners' collocational knowledge; v+n collocations were found to be less challenging than adj+n collocations. Similar to Shehata's (2008) results, Alsakran (2011) conducted a study on 68 advanced learners of English in two different contexts. Thirty-eight students were studying in L1 context, Arabic, i.e., studying English as EFL learners while thirty students were studying in a L2 context, English, i.e., studying English as ESL. The study aimed to examine the influence of patterns of collocations on the participants' receptive and productive collocational knowledge. Three gap-filling tests were utilised to measure the participants' productive collocational knowledge; a v+n and an adj+n collocation tests in which the initial letter of collocates were provided and a v+prep collocation test in which meanings of the phrasal verbs were supplied. Their receptive collocational knowledge was measured by an appropriateness judgment test in which participants had to circle the number corresponding to the underlined part of a sentence that was judged unacceptable. The results revealed that patterns of collocations seemed to have a significant influence on the learners' collocational performance as they scored higher on the v+n collocations than on the adj+n and v+prep items. These results indicated the influence of the syntactic structure of collocations on the L2 learners' collocational knowledge. However, the study could not differentiate clearly between learners' knowledge of the grammatical and lexical collocations as learners performed highly, on the v+n items, and lowly, on the adj+n items, at the same time on the lexical collocations.

In line with the previous study, Gaballa & Al-Khayri (2014) conducted a study on 68 advanced learners of English at Taif University of whom 38 were females and 30 were males. The study attempted to identify the impact of the syntactic structure on the collocational knowledge through comparing learners' knowledge of three syntactic patterns of collocations: v+n, adj+n, and v+prep. The participants' receptive knowledge was measured by an appropriateness judgement test whereas their productive collocational knowledge was examined by three gap-filling tests. Although the L2 learners looked generally in short of

collocational knowledge receptively and productively compared to native speakers, the results revealed that the syntactic structure of collocations had a significant influence on the learners' knowledge; participants appeared to perform better on the v+n collocational pattern than the adj+n and v+prep patterns.

In terms of number of collocation types and patterns and their development, Li & Schmitt (2010) employed a statistical approach to examine a longitudinal learner corpus of four female Chinese postgraduate students majoring in English at the University of Nottingham. The study focused on the learners' productive collocational knowledge development over one academic year. The data included 36 writing assignments. The study findings indicated that learners used a large number of the adj+n pattern collocations. However, the results showed an overall limited development in the learners' use of collocations compared with native speakers. The study ascribed the shortage of the development in the learners' knowledge to the short period of investigation, one academic year, which seemed to be not long enough to clearly reveal learners' collocational knowledge development.

Likewise, Alsulayyi (2015) conducted a study in which he compared the productive knowledge of grammatical collocations of Saudi students majoring in English in the KSA with those in the UK. The study results revealed that the L2 learners studying in the UK had more collocational knowledge than those were studying in the KSA. Additionally, for both groups, certain syntactic structures appeared to be more challenging than others as learners made most of the errors in producing the grammatical collocations patterns, n+prep, adj+prep, and n+prep. Moreover, learners seemed to avoid using more complex grammatical collocation patterns such as adj+ that-clause and n+ that-clause. The avoidance and making errors were due, according to the study, to the participants' lack of sufficient knowledge of the grammatical collocations and L1 interference.

In sum, findings of the reviewed studies imply the involvement of further factors in the learners' collocational knowledge development such as their semantics, frequency of occurrence, cultural and pragmatic features, and L1-L2 differences. Moreover, merely having the grammatical skill does not seem to be sufficient to guarantee appropriate adherence to the collocation restrictions or

ensure producing correct collocations. Furthermore, the influence of the syntactic feature is still in need of more investigations due to the foggy correlation between this feature and learners' collocational knowledge and development. This can be clearly noticed in cases where other reasons lie behind the influence of the syntactic structure on the collocational knowledge such as the impact of the L1-L2 differences. Hence, this aspect of collocation is still in need of more investigations into the other factors that can be disguised in the syntactic structure gown.

3.4.3 The influence of the semantic features

The influence of the semantic features is dealt with in this section from two aspects: commutability and degree of opacity of the collocational constituents. Commutability (or substitutability) in collocations can be defined as "the extent to which the elements in the expression can be replaced or moved" (Howarth, 1996, p. 36). This feature is also called arbitrariness (Fan, 2009), which indicates that substituting a synonym for one of the collocation items may produce an odd or unacceptable lexical combination. For example, the collocation *make an effort* is conventional whereas *?make an exertion* is unacceptable (McKeown & Radev, 2000, pp. 3-4). However, collocations vary in their degrees of flexibility of replacement with other synonyms; some collocations allow a space for one or more substitutes or alternatives of their components whereas others may not allow at all (Koya, 2005). For example, in a collocation such as *make a decision*, a number of de-lexical verbs can substitute the verb *make* such as *reach* or *take* whereas in a collocation such as *shrug one's shoulders* no appropriate substitute can be found for the verb *shrug*.

Accordingly, it has been argued that this feature can be employed to differentiate collocations from the other combinations such as free combinations and idioms (e.g., Cowie, 1981; Howarth, 1996; Nesselhauf, 2005). This is evidenced in that free combinations, collocations, and idioms are different from each other in their degrees of commutability. For example, items of free combinations can be flexibly replaced by other items that fit in the same slot such as substituting a noun with another noun or a verb with another verb, collocational constituents are limitedly substitutable whereas constituents of idioms are fixed (Aisenstadt, 1979).

One of the restrictions of commutability is the semantic features of the collocational component items (Nesselhauf, 2005). For example, one of the selectional restrictions of combining the verb *kill* with a noun, which functions as an object, is the requirement that the object should contain the semantic feature of [+ANIMATE]. Such selectional restrictions prevent generating combinations such as *?kill a chair* (Kim, 2008) unless it is used metaphorically specifically in literary contexts. Such restrictions may also imply the existence of logical relationships between the collocational items.

Additionally, Aisenstadt (1981) and Cowie (1981) suggest a paradigmatic restriction among the different elements within a collocation. In the collocation *make/take a decision*, for example, one of the elements is restricted whereas the other is unrestricted. The noun *decision* can be combined with two verbs: *make* or *take*, it is rather restricted in its combinations with verbs. Contrastively, the verbs are not used with transparent meanings and can be used in various senses and can be combined with many nouns. Accordingly, the noun is rather restricted but not to only one verb [+Restricted Combinability] whereas the verbs are unrestricted [-Restricted Combinability]. Accordingly, learners' shortage in the knowledge of the commutability characteristics of the collocational items can result in producing unacceptable collocations.

Degree of the semantic transparency of a combination is considered as one of the important features of distinguishing collocations from the other types of combinations. Transparency indicates whether one of the combination elements has a literal or non-literal meaning (Nesselhauf 2005). Some linguists (Cruse, 1986; Gitsaki, 1996; Kurosaki, 2012) have emphasised the crucial role of the semantic transparency and claimed it to be the solely or at least the most effective feature that is able to distinguish between some expressions such as idioms and collocations. Additionally, due to the attributed significance of collocational restrictions in learning collocations (Blum & Levenston, 1980; Gitsaki, 1999), Nattinger (1988) argues that through identifying the semantic range and the contexts in which collocations can be used, the increase in learners' knowledge about the syntagmatic relations and the combinatory restrictions of lexical items may assist in producing appropriate collocations.

Broadly speaking, with some collocations, meaning of the whole combination can be anticipated from meanings of their constituent parts (Jaff, 2013).

However, in some other collocations, the meaning of one of the component items of a collocation becomes restricted to the context in which it is used, i.e., a collocational meaning is constructed by adding more abstract semantic features to the collocate; the word is not employed in its usual or literal sense (Kolesnikova & Gelbukh, 2012). For example, Cruse (1986) argues that in the collocations *heavy drinker* and *heavy smoker*, the expressed meaning by the adjective *heavy* in these two collocations differs from that is expressed by the same adjective in the collocation *a heavy bag*. In the first two collocations, the adjective *heavy* has a specific meaning imposed by the second attached components *drinker* and *smoker* respectively, though the added meaning does not affect the semantic transparency of the collocations. This indicates the limited compositionality of collocations; collocations seem to be not fully compositional due to an emerging element of meaning being added to the whole meaning of the combination while maintaining the element of transparency (Manning & Schütze, 1999). In other words, the whole meaning cannot be retrieved from the meanings of the constituent parts of a collocation. This is because while at least one of the elements preserves its literal sense, an additional meaning has been added to the whole meaning which does not exist in any of the components parts, i.e., a meaning by collocation has been added to the meaning of the whole combination. However, being limitedly compositional is different from being completely non-compositional. A non-compositional chunk indicates a case where the total meaning of an expression differs from the total meaning of its individual constituents; accordingly, the overall conventional meaning of this expression cannot be perceived from interpreting the individual lexical items as it is the case with idioms (Granger & Paquot, 2008).

Zheng (2014) holds that L2 learners' semantic awareness develops slowly and unpredictably, and their semantic awareness development fluctuates by undergoing through phases of progression, stabilisation, and regression. Additionally, it is verified in many relevant studies that the semantic opacity of collocations has a magnitude influence on the L2 learners' ability to transfer the figurative meanings of words (e.g., Gitsaki, 1996; Kellerman, 1978; Koya, 2005; Kurosaki, 2012; Nesselhauf, 2003, 2005). Consequently, L2 learners are found to be unable to transfer the idiomatic or figurative meanings of the learned

language. Moreover, some researchers (e.g., Kolesnikova & Gelbukh, 2012; Laufer & Waldman, 2011) maintain that the semantic features of collocations contribute even more than the statistical features in learners' knowledge development of collocations. In other words, the semantic transparency of collocation affects learners' knowledge development more than the frequency of exposure to the target collocations. The underlying assumption of this feature is that as L2 learners are unable to transfer the figurative meaning of collocations as native speakers and the challenge increases with increasing the opacity of the component items of collocations, i.e., used in their figurative not literal senses.

Similarly, Gitsaki's (1996) study results indicated that learners' knowledge development was under the effect of more than one factor of which the semantic complexity of the collocation components was found to be one of the influential elements. Collocations which consisted of elements used in their figurative meanings seemed to be more difficult or were learned later than those were made up of elements which were used in their literal meanings.

Additionally, Gitsaki (1999) maintained that semantic opacity was one of the influential factors for the collocational acquisition. For example, she found that translating collocations which contained a preposition, e.g., adj+prep, were more challenging than the collocations which contained an infinitive due to the influence of L1 interference. Additionally, she argued that although lexical collocations such as v+n and v+adv were syntactically of simple structures they were acquired later than simple grammatical collocations due to the influence of semantic complexity factors such as arbitrariness, predictability, and idiomaticity.

In another investigation into the influence of semantic transparency on learners' collocational knowledge, Nesselhauf (2003, 2005) conducted two successive studies on the use of multi-word units and more than 2000 v+n collocations through analysing a corpus of essays of German learners of English. The studies attempted to identify the challenges learners encounter and errors they make in their learning pathway. The research results indicated that collocations were more challenging than free combinations and idioms. The results revealed that the majority of errors were made in the first place in using collocations, then, free combinations whereas idioms seemed to be the least challenging.

Accordingly, combinations with medium semantic restrictions, i.e., collocations, were the most difficult ones ahead of the learners since the highest rates of mistakes were made in choosing verbs in the collocations. More precisely, the semantic characteristics of the verbs, i.e., collocates, were the most challenging aspects. In other words, the results indicated that the complex nature of collocations was what created a challenge for the L2 learners; collocations are neither completely free combinations to be able to combine any two items together nor are they completely fixed expressions like idioms to be memorised as fixed block units. Accordingly, L2 learners need to think about the possible available options to produce native-like collocations.

While Nesselhauf (2003, 2005) shed light on the influence of the semantic characteristics of the collocational component items on L2 learners' knowledge in general, Koya (2005) investigated this influence more precisely through studying the influence of the core meanings, i.e., the central and context free meanings, of the collocational items on the L2 learners' receptive and productive collocational knowledge. The study results revealed that the core meaning of verbs influenced more learners' receptive knowledge whereas core meanings of nouns and semantic transparency had more impacts on their productive knowledge of collocations. In other words, if the component parts of collocations were employed in their core meanings, learners were likely to comprehend and acquire them easily. In contrast, if collocations involved constituents with peripheral meanings or delexical meanings, they were less likely to comprehend and acquire these collocations without noticeable challenges. This study indicated the importance of core meanings, i.e., semantic transparency, in collocation learning, accordingly, learners' collocational knowledge and development.

In a corpus-based study, Laufer & Waldman (2011) investigated the use of 220 English v+n collocations in the essay writing of 307 university native speakers of Hebrew at three proficiency levels. To achieve the study aims, a learner corpus of about 300,000 words of argumentative and descriptive essays was compiled. The corpus of young adult native speakers of English, Louvain Corpus of Native English Essays (LOCNESS), was selected as a comparison model for learners' use of collocations. The 220 v+n collocations were extracted from the most frequently occurring nouns in the LOCNESS corpus and in the

learner corpus, then concordances were created for them. Two types of comparisons were performed. Learners' knowledge was compared with native speakers' knowledge on the frequency of collocation use, and it was compared with other learners' knowledge of different L2 proficiencies on the frequency and correctness of collocations. The study indicated that native speakers used more collocations than L2 learners though learners' use of collocations inclined to increase along with increasing their proficiency levels. Furthermore, the results showed that learners made errors in using correct collocations even at the most advanced levels of proficiency. The influence of the semantic transparency of the collocational items was found to be one of the reasons behind learners' errors. Interestingly, in contrast to some other studies (e.g., Gitsaki, 1996; Gyllstad & Wolter, 2015) which indicated the positive influence of transparency of the collocational items on the learners' knowledge, Laufer & Waldman's (2011) study results highlighted the negative influence of transparency of the items on the L2 learners' knowledge development. The negative influence of transparency on the L2 learners' collocational knowledge was evidenced in that learners did not seem to pay a special attention to the collocations which consisted commonly of semantically transparent and frequent individual words, e.g., *make a decision* and *send a message* as a problematic aspect of the language when they encountered them in the input. Conversely, they argued that producing native-like collocations seemed to become problematic when collocations had equivalents in the learners' L1 and they contained at least one different constituent item which resulted in producing incorrect collocations. Moreover, they argued that as collocations are not usually noticed by L2 learners due to their semantic transparency, there would be a little chance for increasing learners' knowledge of collocations through increasing exposure to collocations. However, the study results seemed to outrange the role of semantic opacity over frequency or exposure to collocations. Moreover, the limited number of collocation types investigated by the study, v+n pattern, excluded the influence of the syntactic structure feature and the other elements and features whose impacts have been evidenced in many other studies. Though, this does not imply underestimating the significance of the results and their importance specifically within pedagogical contexts.

In line with these studies, Kurosaki (2012) compared the influence of the combinability and semantic transparency of collocations on the collocational knowledge and use of two groups of L2 learners, French and Japanese. The study results indicated that the combinability and semantic transparency of collocations appeared to have influences on the L2 learners' knowledge and use of collocations in both groups of learners. The influence was evidenced in that L2 learners were challenged by transferring the figurative meaning of the collocations whose component items were not used in their literal senses. However, Kurosaki (2012) explained that the influence of this feature was not constant in the study tasks, the translation tasks and the MCQ (multiple choice questions) tasks, due to the nature of the tasks which might have triggered the interference of the influences of some other features such as L1-L2 differences.

From a phrasal frequency-based, i.e., hybrid, perspective, Gyllstad & Wolter (2015) examined the difference in processing time between two types of word combinations: free combinations and collocations. The study was conducted in the theoretical framework of tests predictions made in Howarth's Continuum Model (1996, 1998). A visual semantic judgement task was administered to 27 advanced Swedish learners of English who were compared with 38 native speakers of English as a control group. The results revealed that native and non-native participants spent more time in processing collocations than free combinations, though non-natives spend more time in the collocational processing. According to the study, the additional processing cost stemmed from the semi-transparent characteristic of collocations, according to the phraseological tradition definition of collocations. In other words, collocations were processed slower than free combination due to the semi-transparent nature of collocations. However, the study did not show whether the semantic transparency also influenced L2 learners' collocational acquisition in the same way it influenced their processing.

In conclusion, significance of the semantic feature in the L2 learners' collocational knowledge and development is well evidenced in literature. However, as was shown in the reviewed studies, the influence is interfered and interact with some other features such as frequency, syntactic structure and L1-L2 differences and similarities. Accordingly, identifying the impact of the semantic transparency feature isolated from the influence of the other factors is

an aim has not been reached to date and requires more investigations into the field.

3.4.4 The influence of congruency with L1

Congruency with the learners' L1 is another key feature of collocations which seems to be influencing L2 learners' collocational knowledge and development (Ahmadian & Darabi, 2012). A congruent collocation means that it has the same lexical items in both L1 and L2 whereas an incongruent collocation indicates that it has different lexical items in both languages (Yamashita & Jiang, 2010). The influence of this feature on the learners' collocational knowledge and development has been found to be playing a notable role in a considerable number of studies (e.g., Granger, 1998; Nesselhauf, 2003; Shehata, 2008; Wolter & Gyllstad, 2011). In addition, the influence of this feature has been investigated in different settings; as ESL (e.g., Aghbar, 1990; Gitsaki, 1996; Zhang, 1993); in EFL contexts (e.g., Alsulayyi, 2015; Channell, 1981; Farghal & Obiedant, 1995), and with learners of different languages (e.g., Bandpay, 2012; Kurosaki, 2012; Shehata, 2008). The L1-L2 differences may explain part of the arbitrary nature of collocations, since learners with certain collocations are unable to translate them word-for-word (Smadja, 1993). From this point of view, Bartsch (2004, p. 18) argues that functional, cultural, or pragmatic features can be behind the co-occurrence of some collocation items which, consequently, make it difficult for a learner who is not well-familiarised with the culture and the pragmatic functions of the learned language to produce appropriate collocations. She exemplifies the pragmatic and cultural function of collocations as meaning carriers in that, for example, although the following three sentences, a, b, and c are semantically and syntactically correct, only a proficient native speaker or language user can observe the unnaturalness of the first and second sentences, a and b:

a ?“He was convicted for performing murder.”

b ?“He was convicted for executing murder.”

c “He was convicted for committing murder.”

While the collocation *commit murder* in “*He was convicted for committing murder*” is deemed acceptable, the combinations *perform murder* and *execute murder* are unacceptable to native speakers and they sound odd or unnatural.

Additionally, due to the fact that some collocations are culture-bound, i.e., they reflect the cultural setting of the source language, they can only be understood in the source language context (Baker, 1992). Accordingly, unawareness of this feature makes perceiving the connotational meaning of the culture-bound collocations one of the most challenging tasks to L2 learners (Dinçkan, 2010). Culture-bound collocations are made of lexical items that are closely associated with the culture of the source language and can only be understood in the source language context. Accordingly, this type of collocations usually creates a challenge ahead of non-native speakers (Ebrahimi & Toosi, 2013) such as “*went banana, girlfriend*” and “*love children*” (Badawi, 2008, p.14). Furthermore, Shooshtari & Karami (2013) argue that being deficient in the cultural competence of the source language can result in producing incorrect or odd culturally bound collocations. Thus, the cultural element can play a crucial role in transferring culture-bound collocations meanings appropriately. Badawi (2008) concluded from a study on EFL learners that the majority of the learners even at advanced levels were unable to translate cultural-bound expressions appropriately.

Gitsaki’s (1999) study results revealed that the influence of L1 on the use of collocations was found to be more than the influence of combination restrictions. In addition, she attributed part of the errors made in using collocations to the influence of the L1-L2 similarities and differences. Accordingly, she called for raising learners’ awareness of the differences exist between L1 and L2. The results also highlighted the contributions of some other collocational features such as collocation restrictions in hampering learners’ use of collocations appropriately. In line with Gitsaki’s (1999) suggestion, Alternberg & Granger (2001) emphasised the L1 transfer role and consciousness raising of learners to areas of L1-L2 differences in developing their productive collocational knowledge.

Similarly, Koya’s (2005) study results indicated that the L1 interference was an influential factor in the L1 transfer of collocations. In other words, the similarities among L1 and L2 collocations led to L1 positive transfers in the acquisition of L2 collocations, whereas the differences among them resulted in negative transfers. Accordingly, Koya recommended that L1-L2 similarities should effectively be utilised for developing L2 learners’ collocational knowledge, and

raising L2 learners' awareness of the L1-L2 differences areas to avoiding the negative transfers from L1. He also noted that raising L2 learners' awareness of the L1-L2 differences might reduce the L2 learners' acquisition burden.

In a further study, Shehata (2008) examined the influence of L1 on the L2 learners' receptive and productive collocational knowledge. The study was conducted on 97 advanced Arabic-speaking learners of English who were studying English in two contexts; ESL vs EFL. The results indicated a considerable influence of the L1 on the participants' collocational knowledge. Participants who were studying in the ESL context were of command of more collocational knowledge than those who were studying in the EFL setting. Additionally, the results showed that the influence of their L1, Arabic, on the participants' receptive and productive collocational knowledge who were EFL learners was more than its influence on the ESL learners. However, the study did not show whether learners' L1 influenced more their receptive or productive collocational knowledge.

Similarly, Fan (2009) conducted an exploratory corpus-based study of two highly comparable corpora; 120 essays of which 60 essays were of native speaker students, British, and 60 essays were of non-native, Chinese, students. The study aimed at exploring students' use of collocations and any challenges that could face them in their academic writings. The 60 Chinese essays were provided by the Hong Kong Examinations Authority. Candidates in each centre comprised learners of mixed ability to represent the average performance of Hong Kong secondary school leavers who studied English language for about 11 years. The 60 British essays were collected from 60 native Year 10 students from a comprehensive school in northern England. Participants were asked to write about reporting a crime to the police in about 300 words based on a series of four pictures. The study results drew attention to the negative influence of L1, L2, and insufficient knowledge of lexis and grammar on the L2 learners' productive collocational knowledge due to the differences existed between these two languages; English and Chinese. The negative influence of the L1 was evidenced in using inappropriate collocations which were as a result of translating word-for-word of equivalent collocations. The study revealed confusion with the L2 as another factor, for example, some of the L2 learners were confused by the pronunciation of the English language, e.g., the

collocation “curry hair” with “curly hair”. To address the negative L1 transfer, the study recommended raising learners’ awareness to the L1-L2 differences. Accordingly, the study indicated the contribution of interrelated factors, including L1-L2 differences, but could not provide a clear-cut view of the L1 impact isolated from the other factors. This could be attributed to the complex nature of collocational knowledge and the various elements that affected learners’ knowledge.

A further study was conducted by Yamashita & Jiang (2010) which investigated the influence of the L1 on the acquisition of L2 collocations. The study made a comparison of the performance on a phrase-acceptability judgment task in three contexts: native speakers of English, ESL users, and EFL learners. The test materials included both congruent and incongruent collocations. The study results revealed that EFL learners made more errors and reacted more slowly to incongruent collocations than congruent collocations. The results also indicated that ESL users made less errors and react faster than EFL learners. However, EFL learners seemed to be making more errors on incongruent collocations than on congruent ones. Interestingly, the results demonstrated no significant influence of the L1 on the ESL users’ reaction time. Accordingly, the results indicated that L1 congruency affect the acquisition of L2 collocations. Additionally, the results revealed the difficulty of developing L2 learners’ knowledge of incongruent collocations even with a considerable amount of exposure to L2. Finally, the study indicated that L2 collocations were processed separately of L1 influence once they were stored in memory. This study underscores the importance of storing collocations in memory, but this process cannot be done, according to the study, through increasing L2 learners’ exposure to the L2 collocations specifically in learning incongruent collocations. This may imply contributing other factors in developing this aspect of collocational knowledge.

However, in contrast to Yamashita & Jiang’s (2010) study result about the inactive influence of frequency on the learners’ collocational knowledge development, Wolter & Gyllstad (2013) examined the influence of frequency on the processing of congruent and incongruent collocations of advanced Swedish learners of English as a L2. To this aim, an acceptability judgment task was administered on Swedish learners of English and natives to evaluate their

response times and errors made for the test items, which consisted of a match set of unrelated items. The results indicated that frequency had a considerable impact on the advanced learners' congruent and incongruent collocational knowledge development. Additionally, the results showed a constant influence of the L1 on the time spent on collocation processing even at the advanced levels of English learning; the collocations which were congruent with learners' L1 seemed to be processed faster than incongruent collocations. These results necessitate, in accordance with the study results, taking the influence of frequency and learners' aptitude to process language into consideration beyond the single word level, i.e., to the collocational level.

Laufer & Waldman (2011) concluded from their study that the most influential factor in making erroneous collocations appeared to be that was induced by the influence of the L1. This was because, according to the study, learners seemed to have conveyed their messages without considering word use constraints. Moreover, they appeared to have over relied on the frequent and familiar individual words rather than multi-word units such as collocations. This study, through emphasising the role of the L1 transference, indicated the notable role of the semantic features in making errors and producing incorrect collocations. Accordingly, the existence of correlations in the mental lexicon between both L1 and L2 can have a significant contribution to the L2 learners' collocational knowledge (Wolter & Gyllstad, 2011).

From a phrasal frequency-based perspective, Kurosaki (2012) carried out a study to examine the French and Japanese learners' knowledge and use of collocations. The study explored the impact of L1 transfer, among other factors such as combinability and transparency, on the learners' collocational knowledge and use through testing them on four lexical patterns of collocations: v+n, delexicalised v+n, adj+n, and adv+adj collocations. The study results brought to the attention the effect of L1 background on learners' knowledge and use. However, unexpectedly, it showed that L1-L2 similarities did not, at all cases, affect positively L2 learners' productive knowledge. For example, even though French and English belong to the same Indo-European language family which implies sharing a large number of cognates, the French-English similarities did not seem to have helped French learners produce more correct collocations of all the examined patterns of collocations in the study. On the

contrary, Japanese learners, whose language does not belong to the same language family, produced and had access to more correct collocations than French learners in some of the examined types such as the adj+n pattern. This result was in contrast to many similar previous research results in the field which were mentioned earlier in this section which indicated the positive influence of the similarities between learners' L1 and L2.

A more recent research was a qualitative exploratory longitudinal multiple case study by Zheng (2014). The study was carried out on eight Chinese female university-level EFL learners to examine how their cross-linguistic semantic awareness developed over the course of an academic year. It aimed to describe, analyse, and account for the developmental patterns of the vocabulary knowledge of L2 learners. At the macro-level, while the study results indicated a constant receptive and controlled productive knowledge development, learners' vocabulary free production tended to stabilise and decrease after their first year of the academic study. At the micro-level, while learners' paradigmatic knowledge witnessed a continuous progress, individuals' syntagmatic knowledge inclined to stabilise. In other words, while learners' knowledge of individual words was developing, their collocational knowledge stabilised. Moreover, analysing learners' collocational mistakes indicated that their choices were influenced by whether the L2 expressions had exact word-to-word equivalents in the L1.

In sum, despite the attributed influence of L1 on the L2 learners' collocational knowledge and development in numerous studies, it seems from the studies themselves that some other factors work with, interfere or interact with it as well such as semantic transparency, commutability, syntactic structure, learning context and the L2 learned itself. Moreover, some studies yielded contradictory results. Accordingly, the need for more research emerge to identify, in more detail and accuracy, the influence of L1 on the L2 learners' knowledge and development of collocation.

3.4.5 The influence of the phonological features

In addition to the four features mentioned earlier, there are some other features which can be attributed to collocations, though they seem to be of less significance in their influence on the L2 learners' collocational knowledge. The phonological feature is one of these characteristics. In general, the current idea

about combinatory units or blocks of language is that they are produced and uttered within one level of intonation contour and faster than non-chunked language and that native speakers draw more on these memorised lexical phrases of language than processing their own speech structure using individual words as non-natives do (e.g., Crookes, 1989; Erman, 2007; Foster & Skehan, 1999; Foster, Tonkyn, & Wigglesworth, 2000; Leedham, 2011; Wood, 2004, 2008). Lindstromberg & Eyckmans (2014) argue that assonance as in *strong bond* and the repetition of the consonant sounds in alliterative collocations for example *green grass* and *sea salt* may facilitate recalling formulaic sequences. Accordingly, the facilitative roles of assonance and alliteration can be employed pedagogically to develop L2 learners' knowledge of collocations by raising L2 learners' awareness to these phonological phenomena. I would argue that this may imply also the use of the musical characteristics of a language in enhancing L2 learners' collocational knowledge and development. This could be supported by the general overwhelming influence and interest of music on the new generations specifically, younger generations. In contrast, Boers, Lindstromberg, & Eyckmans (2014) point out that the existence of specific phonological repetitions in the component parts of collocations, instead of assisting the acquisition of collocations, may impede learners' retrieving of these collocations such as *front row*, *popular appeal*, and *important point*. Accordingly, this necessitates explicitly working on raising learners' awareness of collocations through the set learning materials in educational contexts. In addition, the different influences of these phonological features whether facilitating or hampering knowledge development necessitate adopting different strategies that are flexible according to the influences these features have on the language learners. It also implies the need for more studies on the influence of the phonological feature on the L2 learners' collocational knowledge and development.

Finally, the factors which are mentioned above are not exhaustive, but they have been recognised as the main influential factors affecting collocational knowledge and development according to the conducted review of literature. However, the current study focuses on only four main features: frequency, syntactic structure, semantic transparency, and L1 congruency as mentioned earlier.

3.5 The current model of collocational knowledge and development

Investigations into collocations and L2 learners' collocational knowledge and development have reported corroboratory and contradictory views and findings. Though, a general outline of the collocational knowledge and development can still be elicited from the presented review of literature.

Generally, L2 learners' language knowledge develops rather slowly over time (e.g., Alharthi, 2014; Boers, Lindstromberg, & Eyckmans, 2014). However, due to the complex nature of language knowledge and development, the general view of language knowledge and development seems to have shifted from a linear static into a non-linear view of the language knowledge and development model (e.g., de Bot, Lowie, & Verspoor, 2005, 2007; de Bot, Lowie, Thorne, & Verspoor, 2013; Kramsch, 2002; Larsen-Freeman, 1997, 2000, 2006). In the dynamic view, language knowledge and development are perceived as complex systems that consist of a number of dimensions which change constantly (Larsen-Freeman & Cameron, 2008). These dimensions develop at various and often nonlinear rates accompanied by high degrees of variation between and within the performance of individual learners over time (de Bot, Lowie, & Verspoor, 2005). The various levels of learners' linguistic knowledge compete each other for getting resources in a dynamic non-linear intricate web of relationships among the different aspects of knowledge of the learned language (Caspi, 2010; Caspi & Lowie, 2013). However, L2 learners' knowledge and development fluctuates by constant progressing and regressing during the learning process which in total achieve the general stability of the system by time pass (Verspoor, Schmid, & Xu, 2012). In other words, second language knowledge and development seem to go through periods of knowledge progressing, regressing, stability, and instability, i.e., the learning process follows identifiable stages in a nonlinear developmental pattern (Lowie, Verspoor, & de Bot, 2009).

However, I would argue in line with Vercellotti (2012) that it can be claimed that language knowledge and development are not completely static and linear nor totally non-linear and dynamic at all the development stages. In other words, specific parts of knowledge at specific phases may develop in static, linear pathways then develop nonlinearly and dynamically to achieve the final and total balance of the whole system. For example, a feature of collocations, such

as frequency, may positively influence learners' knowledge and development where an increase in exposure to a collocation leads to an increase in the learners' knowledge of that collocation. This is a linear relationship between the feature and learners' knowledge. However, the influence of this feature may behave nonlinearly in that this feature can be influenced by other factors such as congruency with L1 which may not necessarily result in increasing learners' knowledge despite increasing learners' exposure to the collocation (e.g., Yamashita & Jiang, 2010).

Collocationally, a further shift in focus has moved the emphasis from highlighting the role of learners' knowledge of single words in the language knowledge and development into emphasising the role of learners' knowledge at the phrase and multi-word units level, such as collocations, in their knowledge and development. The growing body of evidence has revealed an existing positive proportional relationship between language users' proficiency and their level of collocational knowledge (e.g., Gitsaki, 1996; Keshavrz & Salimi, 2007; Lewis, 1993, 1997; Nizonkiza, 2015).

Collocational knowledge is thought to be an essential and a significant part of the general knowledge of the learned language (e.g., Channell, 1981; Fan, 2009; Marton, 1977; Nesselhauf, 2003; Nizonkiza, van Dyk, & Louw, 2013; Wu, 1996). Although learners' collocational knowledge increases with the development of the general linguistic knowledge, in contrast to Gitsaki (1996), it does not appear to develop concurrently at the same speed rate of the other aspects of language knowledge development such as spelling, phonology, grammar, and meaning or the general linguistic proficiency (e.g., Shokouhi & Mirsalari, 2010). Conversely, learners' collocational knowledge develops very slowly in comparison with the development of the other aspects of language knowledge and proficiency (e.g., Altenberg & Granger, 2001; Laufer & Waldman 2011; Nesselhauf, 2005). The slow advancement in the collocational knowledge development is attributed to that L2 learners are unable to use collocations as efficiently as native speakers. They behave differently with collocations from native speakers. This difference is attributed to that collocations are learned by non-natives in very restricted contexts and this issue continues even with learners at advanced levels of language (e.g., Arnaud & Savignon, 1997; Hoey, 2005; Nesselhauf, 2005).

Dimensionally, despite the different suggested classifications of researchers, learners' knowledge of collocations can be mainly divided into two dimensions: receptive and productive knowledge (e.g., Anderson & Freebody, 1979; Melka, 1997; Milton, 2009). In addition, a significant positive relationship appears to exist between learners' receptive and productive collocational knowledge, and their overall language proficiency (e.g., Bonk, 2000; Koizumi, 2005; Tahmasebi, Ghaedrahmat, & Haqverdi, 2013). Furthermore, abundant evidence in this respect indicates that learners' productive collocational knowledge lags much behind their receptive knowledge of collocations (e.g., Alsakran, 2011; Laufer, 1998; Milton, 2009; Torabian, Maros, & Subakir, 2014; Webb, 2008).

Developmentally, study results refer to a developmental pattern in the L2 learners' receptive and productive collocational knowledge. Accordingly, learners seem to develop receptive knowledge of collocations faster and prior to their productive knowledge (e.g., Durrant & Schmitt, 2009; Laufer & Waldman, 2011; Li & Schmitt, 2010; Yamashita & Jiang, 2010); with further exposure, they learn new collocations receptively, simultaneously, they promote their already receptively-known collocations to the productive level (e.g., Schmitt, Dörnyei, Adolphs, & Durow, 2004). Moreover, research in this aspect indicates that the relationship between the receptive and productive level is not a straightforward linear relationship (e.g., Caspi, 2013) and the transition from the receptive to productive level is not fast, abrupt, or predictable (e.g., Laufer, 1998; Schmitt & Meara, 1997). The complex interaction between these two levels of knowledge over time while they are competing for learner resources, on the one hand, and supporting each other's growth on the other hand result in a gap between L2 learners' receptive and productive knowledge (e.g., Alsakran, 2011; Caspi & Lowie, 2013; Torabian, Maros, & Subakir, 2014). Furthermore, this gap seems to increase with increasing learners' proficiency level (e.g., Laufer, 1998).

The collocational knowledge and development of L2 learners appear to be affected by various features that are attributed to the nature of collocations themselves. The influence is evidenced in a wide range of related studies (e.g., Gitsaki, 1999; Kellerman, 1978; Koya, 2005; Kurosaki, 2012; Nesselhauf, 2005). The key features which are believed to be influencing L2 learners' collocational knowledge and development include frequency of occurrence, semantic properties, syntactic features, and congruency with L1. Additionally,

some other factors have been evidenced to affect learners' knowledge of collocations such as learning context, length of exposure to L2, degree of explicitness in teaching collocation, topic familiarity, use of references, technological means, and social media networks and curriculum settings (e.g., Ashouri, Arjmandi, & Rahimi, 2014; Ganji, 2012; Liu, 2002; Pigada & Schmitt, 2006; Webb, 2008).

Among the collocational features, frequency is one of the most powerful features that can affect L2 learners' collocational knowledge and development (e.g., Bybee, 2008; Bybee & Hopper, 2001; Gyllstad & Wolter, 2015; Mueller, 2011; Myles, Hooper, & Mitchell, 1998). It is well evidenced in the literature that learners' collocational knowledge is correlated positively with the rate of frequency of collocations; higher levels of collocational knowledge are expected to be associated with collocations of higher frequency in the learning context (e.g., Alsakran, 2011; Bybee & Hopper, 2001; Bergström, 2008; Durrant, 2008, 2014; Durrant & Schmitt, 2009, 2010; Gitsaki, 1996; Howarth, 1998; Hama, 2011; Shehata, 2008; Siyanova-Chanturia, 2015; Webb, 2008; Wolter & Gyllstad, 2013).

However, it seems that the impact of frequency on the learners' collocational knowledge is not proportional and does not increase their knowledge straightforwardly with increasing their exposure to the target collocations. Rather, its influence is effective to a certain extent, and it develops specific aspects of the learners' collocational knowledge more than others. For example, it develops specific aspects of the learners' receptive collocational knowledge such as form but with probably less influence on the productive knowledge development (e.g., Durrant & Schmitt 2010; Laufer & Waldman 2011; Nguyen & Webb, 2016; Yamashita & Jiang, 2010).

Another distinctive feature that is evidenced to influence learners' collocational knowledge and development is the syntactic structure of collocations (e.g., Alsakran, 2011; Gaballa & Al-Khayri, 2014; Gitsaki, 1996; Koya, 2005; Shehata, 2008; Shokouhi & Mirsalari, 2010). Structurally, collocations are mainly divided into two lexical and grammatical collocations, and each type is subdivided into several patterns (e.g., Bahns & Eldaw 1993; Baker, 1992; Benson, Benson, & Ilson, 1986; Gitsaki, 1996; Hausmann, 1989; Smadja, 1993; Wei, 1999). Grammatical collocation patterns commonly consist of a verb, a noun, or an

adjective combined with a preposition such as adj+prep and prep+n collocations or a grammatical structure such as an infinitive or a clause (e.g., Bahns & Eldaw 1993). Lexical collocations generally comprise an array of associations of verbs, nouns, adjectives, and adverbs such as v+n and adj+n collocations (e.g., Benson, Benson, & Ilson, 1986)

However, researching the influence of the syntactic structure on L2 learners' collocational knowledge and development has resulted in contradictory findings. Some studies have indicated a developmental pattern in that lexical collocations are more difficult to learners than grammatical ones (e.g., Gitsaki, 1996); grammatical collocations of simple syntactic structures are found to be acquired or learned before lexical collocations. Some other studies have encountered these results indicating that grammatical collocations are more challenging than lexical ones (e.g., Bahardoust & Moeini, 2012; Mahmoud, 2005; Shokouhi & Mirsalari, 2010). The inconsistency in the study findings may be due to that influence of the syntactic structure varies according to the context and other factors that affect learners' collocational knowledge development. Accordingly, some patterns of collocations pose higher challenges ahead of L2 learners than others in specific contexts.

Developmentally, according to the findings of the developmental studies in this field, L2 learners' collocational knowledge development appears to follow an ordinal pattern (e.g., Caspi, 2010; Churchill, 2007; Gitsaki, 1996; Nesselhauf, 2003; Schmitt, Dörnyei, Adolphs, & Durow, 2004; Verspoor, Lowie, & van Dijk, 2008). Accordingly, first, learners acquire lexical collocations since they develop their lexicon first (e.g., Caspi, 2010). Secondly, they start using complex grammatical collocations as their grammatical competence increases over time. Third, they use both lexical and grammatical collocations efficiently. Fourth, after mastering both types of collocations, grammatical collocations seem less challenging than lexical ones, i.e., after getting familiarised with the syntactic structure of collocations, the earlier type is learned faster than the latter (e.g., Gitsaki, 1996). I would argue that this seemingly ordinal development may partly indicate the linearity of certain aspects of collocational knowledge development at certain phases. Similarly, studies refer to the possibility of coinciding the collocational knowledge development with the knowledge

development of individual words (e.g., Verspoor, Lowie, & Dijk, 2008). This in its turn can indicate the non-linear nature of knowledge development.

Additionally, learners' language knowledge development seems to pivot mainly around the interactive relationship of their lexical and grammatical knowledge aspects. A developmental order pattern of learners' knowledge of the lexis and grammar can also be observed from examining learners' transitions between the knowledge levels. According to the studies in this aspect of linguistics (e.g., Gitsaki, 1996; Verspoor, Schmid, & Xu, 2012), if knowledge could be divided into five levels, the development in the learners' lexical knowledge usually occurs between levels 1 and 2. However, the syntactic development can be noticed between levels 2 and 3, and both lexical and syntactic development can be observed between levels 3 and 4. Finally, only lexical development can be observed in the transition between the levels 4 and 5 which is mainly a development in terms of using particles, compounds, collocations, and fixed phrases. This indicates that in the early phases of knowledge development, all available resources are employed to develop learners' lexicon. In the next phase, more resources are required for developing learners' grammatical system that provides the functional distribution of information necessary for producing multiple-word phrases or combinations (Verspoor, Lowie, & Dijk, 2008).

Studies also indicate inconsistency in the learners' lexical knowledge development and patterns of use as they fluctuate and vary during the learning process. For example, while using some structures dominate at a certain stage, use of some others disappear. In the middle of this, intermediate structures are utilised and sub-systems compete each other for getting developmental resources. The lexical and grammar systems seem to be in continuous interactions and collaboration in a way that the development in a system results in developing the other in a dynamical rather than a static, linear relationship between them. Learner's vocabulary knowledge development leads to increasing learner's grammatical competence which in return results in producing more correct grammatical structures. However, learner's lexical and grammatical knowledge fluctuate showing progress and regress and they do not improve concurrently due to the influence of various learning contextual factors (Bell, 2009; Fitzpatrick, 2012).

Studies also indicate that, at the macro-level, at an advanced level of proficiency, learners' receptive knowledge develops constantly whereas their productive knowledge development is controlled. Accordingly, learners' vocabulary free production tends to stabilise and decrease gradually. At the micro-level, while learners' paradigmatic knowledge witnesses a constant progress, individuals' syntagmatic knowledge tends to stabilise (e.g., Zheng, 2014). In other words, while learners' knowledge of individual words develops, their collocational knowledge stabilises. This indicates a strong relation between lexicon and grammar which are alternating growth and variability patterns across the process of language knowledge development.

Although these two linguistic dimensions, lexicon and grammar, compete each other for getting the resources required for their development, it seems that lexicon development precedes syntax development as a prerequisite for the syntactic knowledge development. Firstly, learners promote their lexical knowledge by producing more complex words. Secondly, they use these words more accurately. The same developmental pattern is identified in the learners' grammatical knowledge development. Learners' knowledge of syntax is found to be more complex and then develops to be used more accurately (e.g., Caspi, 2010).

Semantically, based on the collocational studies, the influence of semantics on the L2 learners' collocational knowledge and development can be viewed from two perspectives. First, collocational semantic transparency, i.e., whether the component items are used in their literal or figurative senses. Second, collocational congruency with learners' L1, i.e., whether collocations have L1 equivalents or not. Generally, it is evidenced in studies that L2 learners' semantic awareness develops gradually and arbitrarily, and it goes through phases of progression, stabilisation, and regression (e.g., Zheng, 2014).

As for the influence of collocational semantic transparency on L2 learners' knowledge and development, studies indicate that collocational transparency has generally a positive influence on learners' knowledge of collocations; opaque collocations are found to be more challenging than transparent ones (e.g., Gitsaki, 1996; Gyllstad & Wolter, 2015; Kellerman, 1978; Koya, 2005; Kurosaki, 2012; Nesselhauf, 2003, 2005) except few studies which indicate that opaque collocations are less challenging than transparent ones (e.g., Laufer &

Waldman, 2011). In other words, learners' collocational knowledge development is less challenged by collocations which consist of transparent items and more challenged by those made up of items used in their figurative senses. This is due to that L2 learners' inability to transfer, competently like native speakers, the idiomatic or figurative meanings of the learned language (e.g., Kurosaki, 2012). Additionally, it is concluded that degree of semantic transparency of collocations has a noticeable influence on L2 learners' ability to transfer the figurative meaning of words (e.g., Gitsaki, 1996; Kellerman, 1978; Koya, 2005; Kurosaki, 2013; Nesselhauf, 2003, 2005). Lastly, studies indicate that learners' lack of knowledge of specific collocational components has influence on specific aspects of learners' collocational knowledge development. Accordingly, more emphasis should be put on specific parts of collocations to develop certain aspects of collocational knowledge. For example, it is evidenced that transparency of the meaning of verbs affects L2 learners' receptive collocational knowledge whereas transparency of the core meaning of nouns has impacts on the development of their productive knowledge of collocations (e.g., Koya, 2005; Nesselhauf, 2003).

Finally, the influence of L1-L2 differences on the L2 learners' collocational knowledge and development is well evidenced in the body of literature of collocational studies (e.g., Bandpay, 2012; Farghal & Obiedant, 1995; Gitsaki, 1996; Kellerman, 1978; Koya, 2005; Nesselhauf, 2003; Shehata, 2008; Schmitt, Dörnyei, Adolphs, & Durow, 2004; Wolter & Gyllstad, 2011). Studies in this respect reveal that L2 learners rely, when they lack sufficient knowledge of the L2, heavily on their L1 through frequently using those have direct translation equivalents and lexically congruent collocations (e.g., Bahns, 1993), avoiding, as much as possible, incongruent collocations (e.g., Granger, 1998; Granger & Meunier, 2008), overusing high-frequency collocations and underusing those with less frequent collocations (Durrant & Schmitt, 2009).

Despite the difficulty of providing an accurate portray due to the contradictory and varying perspectives and findings, above is an approximate outline of the general existing view of language and L2 learners' collocational knowledge and development which can be elicited from previous studies. Additionally, it shows the main findings of the studies exploring the influence of the major factors on learners' knowledge of collocations.

3.6 Concluding and identifying the gap

The importance of collocations in L2 learning has led to investigating L2 learners' collocational knowledge and development. However, research in learners' collocational knowledge and development have not resulted, so far, in a consistent and clear outline of the collocational knowledge, developmental pattern of collocational knowledge, and how collocational features affect learners' collocational knowledge and development. This can generally be attributed to two factors. First, it is due to the complex nature of collocations and collocational knowledge. Second, it is a result of the diversities, and at times contradictions, in the researchers' perspectives of collocations in terms of definitions and results of the conducted studies on collocations.

Consequently, this situation highlights the need for more investigations to set more comprehensive criteria for defining collocations and verify the wide spread influence of certain features on the learning and knowledge development of collocations. Additionally, it underscores the need for more studies to stand on and explain the mechanism of the influence of the features on collocational knowledge and development. It also raises the need for more investigations to identify the most influential features and how to reduce their negative effects on the L2 learners' collocational knowledge and development. Finally, it brings to light the need for explaining the contradictions in the findings of the various conducted studies and reconciling them through identifying the underpinning reasons behind the contradictions in the results which could be artificial not real. In other words, there could be reasons behind these contradictory results. Accordingly, learners' knowledge and development behave differently in a way reveals unexpected or contradictory results.

Principally, the influence of collocational features: frequency, syntactic structure, semantic transparency, and congruency with L1 on the learners' collocational knowledge and development have been found to be the major influencing factors on L2 learners' knowledge and development. Although these aspects have been investigated intensively in earlier studies, a number of characteristics could be attributed to the conducted studies. First, most of the conducted studies, except very few ones, have been conducted only from one point of view; either from frequency-based or phraseological perspectives. The reliance on frequency alone may ignore or underestimate the impact of the syntactic

structure on the learners' knowledge and development. Conversely, focusing on the syntactic structure alone may ignore the substantial and undeniable influence of frequency on the learning process. Second, previous studies have mostly investigated learners' collocational knowledge and development at one point of time, and few studies have explored these aspects longitudinally and even the longitudinal studies were conducted mostly on few participants (Siyanova-Chanturia, 2015). Consequently, the findings have not been reconfirmed by successive shots of data taken from the same study sample to verify the obtained results, i.e. conducting longitudinal studies. Additionally, some aspects of knowledge require longer periods to be observable, specifically collocational knowledge which develops slowly over a long time span. Third, the tested items have not often been those which had really been experienced by the examined participants as the tested items have not been extracted from their curricula textbooks or course books. The items of such studies have been mainly extracted from COCA (Corpus of Contemporary American English), BNC (The British National Corpus) or dictionaries. Many of these items may have not been studied or learned by participants before the study. In their best situations, most of previous studies sought learners' knowledge of collocations which might have been learned incidentally not intentionally or explicitly. Fourth, most of the studies have been carried out on learners of advanced proficiency levels of language, and few studies have been done on lower levels such as beginners, pre-intermediate, intermediate, or upper-intermediate levels. Finally, the topic of this study has rarely been explored at this level of study on Kurdish students, none to the best of my knowledge. This kind of studies can provide, in addition to the existing body of literature, locally obtained information which can be utilised in the pedagogical and academic contexts in the local setting.

From above, the current study attempts to fill a small part of the lacuna left by previous studies and make a humble contribution in terms of collocation definition, adopted design, and influence of the main collocational features on the L2 learners' collocational knowledge and development. More information about the framework of the current study is provided in the following section.

3.7 The current study framework

The present study aims at exploring English collocational knowledge and development of Kurdish high school students, Year 11, over one school year. It

also attempts to identify how learners' collocational knowledge and development are affected by the main four features of collocations: frequency, syntactic structure, semantic transparency, and congruency with the learners' L1, Kurdish.

Accordingly, this study seeks basically to find the answer to how learners' collocational knowledge and development is related to specific features of collocations. The present study adopts a framework based on the mix of the frequency-based and the phraseological approaches, i.e., a phrasal frequency-based outline, and the study is not established on explicit assumptions. The study framework includes four patterns of collocations which were adapted from the Benson, Benson, & Ilson (1986), The BBI Combinatory Dictionary of English, and Gitsaki (1996). The hybrid approach, which is considered rather new compared with the other two approaches separately, has been adopted in several studies (e.g., Alternberg & Granger, 2001; Kurosaki, 2012; Nizonkiza, van Dyk, & Louw, 2013). The aim of the mixed approach is to operationalise the notion of collocation and to explore the English collocational knowledge and development of the Kurdish learners. The tested items consist of two types of collocations, lexical and grammatical collocations, and each type includes two patterns of collocations. The four patterns of collocations are comprised of two lexical collocations: v+n and adj+n, and two grammatical collocations: phrasal verbs and prep+n collocations. These four patterns are selected on the bases of their frequency and structure in the students' curricula English textbooks. The process of patterning collocation and their extraction from the textbooks are explained in detail in Chapter four (Section 4.6.3). Operationalising the lexicality or grammaticality of the collocational patterns was mainly based on Benson, Benson, and Ilson (1986), the BBI Combinatory Dictionary of English, and Gitsaki's (1996) classifications, which will be explained in detail in Chapter four (Section 4.6.4).

I believe that the hybrid approach which is adopted for conducting this study is more appropriate than the use of either of the frequency-based or the phrasal approach separately. In terms of structure, this approach excludes specific frequent patterns such (*although he*) to be regarded as a collocation relying solely on the frequency based approach. In terms of frequency, a combination of a syntactic structure such as a v+n combination which occurs more than

once is considered a collocation. Accordingly, a combination that occurs only once is not a collocation according to this study as it could have happened by mistake or chance. Hence, on the one hand, the adopted approach within a longitudinal design is seen to provide a more comprehensive view of the term “collocation” and follow up learners’ knowledge development over the school year. On the other hand, mixing both views can fill part of the gap which each approach leaves when they are employed separately. Additionally, the longitudinal design of the study can provide a second chance to reconfirm or recheck the data obtained and the gained results from analysing participants’ data and provide deeper exploration into the learners’ knowledge and development of English collocations. A detailed description of the study will be presented in the methodology chapter, Chapter four.

Chapter four

Methodology

4.1 Introduction

The rise of collocations as a significant element in the L2 learners' language proficiency necessitated thorough explorations into this linguistic phenomenon from various aspects. However, the multifaceted nature of learners' collocational knowledge has resulted in adopting diverse approaches and viewing collocations from different, overlapping, or hybrid perspectives. This diversity has yielded positive and negative outcomes. The bright side of this diversity is investigating collocations from different aspects and at different levels. This is a desirable result due to the complexity of collocations and collocational knowledge on the one hand, and the intricate and interrelated relationships of learners' knowledge of collocations, on the other hand. However, these diverse approaches have led to two less positive consequences. Firstly, they have dispersed researchers' and scholars' endeavour to yield a commonly agreeable definition to collocations that satisfies the linguistic community. A consensus on a definition of collocation can concentrate and direct researchers' efforts on one direction. Accordingly, it may facilitate studying collocational knowledge and development from different aspects, and designing and developing more investigatory appropriate tools for this purpose. Secondly, the various approaches have not established a consensus on an approach to accurately and reliably measure and track learners' collocational knowledge and development over a considerable time span.

As far as the current study is concerned, a phrasal frequency-based approach is adopted. The phrasal frequency-based approach means to define and examine collocations from the points of view of frequency-based and phrasal approaches together. This implies taking frequency of co-occurrence, syntactic structure, semantic transparency, and combinability of the collocational items into consideration. The rationale behind adopting this approach can be explained in two points. First, it is a way to overcome the shortcomings of adopting a single view of collocation rather than the other, i.e., frequency-based or phraseological view. One of the major shortcomings in the frequency-based approach that will be avoided in the phrasal frequency-based approach is

counting recurring combinations of no grammatical relationships as collocations such as (*and the*). The exclusion of such combinations from collocations will be operationalised by depending exclusively particular syntactic structures as collocations. Similarly, one of the major weaknesses of the phrasal approach will be avoided which is highlighting combinations that are not salient to native speakers, rarely or never occur in use. This avoidance is achieved by focusing only on frequent collocations rather than on some combinations which might appear in use only once or never occur at all.

Second, it allows to adopt a more comprehensive definition that encompasses the different aspects of knowledge of collocations such as frequency of occurrence, semantic transparency, and grammatical structures. Accordingly, combining most of the advantages of the two approaches in one study design may provide a better understanding of the L2 learners' collocational knowledge and development. A more comprehensive and sophisticated view of collocations which takes more elements into consideration may yield better insightful investigations and understandings of collocations.

The present study aimed at exploring the development of Kurdish students' English collocational knowledge and development. Additionally, it attempted to identify how their collocational knowledge and development were affected by four features of collocations: frequency, syntactic structure, semantic transparency, and congruency with participants' L1, i.e., Kurdish. It also attempted to examine the influence of gender on the participants' collocational knowledge and development. Characteristics of the study context such as the social, economic, political, and cultural situation of both males and females suggest taking this element into consideration as well in the current study.

Accordingly, this study sought basically to find answers to the following questions:

- What is the influence of gender on the participants' collocational knowledge and development?
- What is the influence of the frequency of collocations in their curriculum textbooks on the participants' collocational knowledge and development?
- What is the influence of the syntactic structure on the participants' collocational knowledge and development?

- What is the influence of the semantic transparency of collocations on the participants' collocational knowledge and development?
- What is the influence of the congruency with L1 on the participants' collocational knowledge and development?
- Which features (frequency, syntactic structure, semantic transparency, or congruency with L1) are more influential on the participants' knowledge and development?
- How does participants' collocational knowledge develop over a school year?

This chapter explains the adopted design for the study, recruited participants, instruments and tests used for collecting data, and the utilised analytical tools for analysing data. Additionally, the ethical concerns and issues which faced the study and the adopted procedures in dealing with the ethical issues in complying with the BERA (British Education Research Association) ethical guidelines are presented.

4.2 The longitudinal design of the study

A longitudinal study is an observational investigation in which multiple data waves that yield repeated measurements on each single participant are taken from the same population over a period of time (Belle, Fisher, Heagerty, & Lumley, 2004). This design of study aims to identify the whole pattern of development or change in the examined phenomenon (Goldstein, 1968). Schmitt (1998a) observes that vocabulary studies which have been based on one-shot data elicited from different population samples and activities seem to have offered a little depiction of learners' vocabulary acquisition. Accordingly, a longitudinal design may provide a better understanding of the vocabulary learning process since it takes data from the same sample of population and test them on the same activity over a period of time. Hence, studying a linguistic phenomenon of the same population for a longer time span is thought to be the most reliable if not the only appropriate design to identify L2 learners' collocational knowledge development (Burr & Nesselroade, 1990; Li & Schmitt, 2010).

I decided to conduct the current study in a longitudinal design. In the studies carried out within non-longitudinal designs, data have been taken at only one point in time. Consequently, studying change and development over time,

comparing two points of the same population and phenomenon over the time scale, identifying patterns of development, the transition from one stage to another, and measuring the prevalence of a factor of interest at several points in time may not be efficiently feasible within this design since such studies can only provide access to explore differences between individuals, but not within individuals (Taplin, 2005).

Collocationally, it has been noted that information taken at one point on the time scale may yield ineffective basis for investigating learners' collocational knowledge and development (Kristen, Römmer, Müller, & Kalter, 2005; Rajulton, 2001). Hence, some studies (e.g., Dörnyei, Durow, & Zahran, 2004; Fitzpatrick, 2012; Li & Schmitt, 2010a; Schmitt, Dörnyei, Adolphs, & Durow, 2004; Verspoor, Lowie, & van Dijk, 2008; Zheng, 2014) have indicated that the only reliable design which can offer an access to identify patterns of L2 learners' collocational knowledge and development is through conducting a longitudinal study of the same population sample over a time span.

In accordance, I would argue that the longitudinal approach to collocations may offer more possibilities to identify the developmental patterns and change in learners' knowledge and use of collocations and explaining the collocational learning process. Filling part of the gap is done within this design through compensating some of the limitations in the other designs or approaches in terms of the study time span, population sample size, waves of data collection, and tracking the developmental pathways of learners' collocational knowledge and development over a rather long time span.

However, numerous advantages and disadvantages have been attributed to longitudinal studies. In general, this design of studies is recommended for the wide range of features it offers over the other designs or approaches. Arguably, the longitudinal designs have been raised as a reaction to the criticism addressed against the cross-sectional study designs in studying change in learners' knowledge. While cross-sectional studies investigate and describe the condition or position of a phenomenon at a certain point in time, i.e., deal with status, longitudinal studies are more concerned with the change or development which occur to the condition or position of the phenomenon over a long period of time (Rajulton, 2001). The first advantage might be the length of the time span of observation in longitudinal studies which is regarded as one of the

prerequisites for identifying the progress and changing in a phenomenon over time (Miller, 2000).

A further plus point of this design is the belief that it is essential specifically for identifying the causal relationships between variables and transitions in the individual behaviour (Smith & Torrey, 1996) since the long-time span gives more space and opportunities to find out any cause and effect relationships that may exist between the variables. Additionally, this design proves its appropriateness for studying complex social processes because it provides the required information for establishing measuring change, temporal order, and providing stronger explanations for the causal relationships between the variables of the study (Cohen, Manion, & Morrison, 2011; Menard, 2002).

As the longitudinal design allows measuring the status of the same phenomenon of the same individuals with the same variables at two or more different points in time, measuring the change in status between these temporal points becomes feasible. Additionally, researchers can identify which variables change first and which follow through observing transitions from a status to another. By comparing a number of status points, a developmental pattern can be concluded and a temporal order can be established. Having the developmental pattern in hand and founding the temporal order of the change process of variables, a deeper and more comprehensive interpretation of the causal relationships becomes more possible.

However, the main criticisms addressed to the longitudinal studies have been related to their expensiveness, time consuming, and efforts exerted in collecting data (Cohen, Manion & Morrison, 2011). Analytically, it has been observed that the collected data within this design, requires complex software and statistical procedures to interpret the gathered data (Menard, 2002). Moreover, the ethical considerations and issues related to the confidentiality and privacy regarding accessing its data due to the prolonged contact between the researcher and the researched and dealing with personal data rise as a further issue in conducting such studies (Farrall, 2006). Furthermore, Rajulton (2001) argues that the availability of some other designs that can investigate change such as cross-sectional designs, by incorporating suitable data collecting and analysing tools, may question the rationale behind adopting such an expensive, complex, and problematic approach.

Despite the disadvantages ascribed to the longitudinal design, its adoption in the recently conducted studies specifically in tracking change and progress in social phenomena has been increasing (e.g., Caspi, 2010; Churchill, 2007; Fitzpatrick, 2012; Li & Schmitt, 2010; Schmitt, Dörnyei, Adolphs, & Durow, 2004; Siyanova-Chanturia, 2015; Zheng, 2010, 2014). Accordingly, the longitudinal design was adopted for achieving the aims of the current study.

4.3 Procedure

After identifying the gap and setting the aims of the current study, I had to decide on the procedures I should follow concerning the tests that would be used for collecting data and examining participants' collocational knowledge, the pilot studies, length of the tests, recruiting participants, and the statistical techniques which would be used in analysing data.

The first step after setting the study aims was to design the tests that would be used for collecting data to examine participants' collocational knowledge. The study explored learners' receptive and productive knowledge through two separate tests, and each test was conducted on a separate test paper. The tests were designed based on measuring tools used for the same purposes in previous studies (e.g., Gyllstad, 2007). Since participants' language proficiency at this level of learning is intermediate, based on the curricula textbooks contents and the evaluation of the teachers who teach English language at this level, I designed the tests in a way that suited their proficiency level. This was to avoid any shortcomings that could be attributed to the content of the tests such as difficulty of the test items. Thus, the test items were chosen carefully based on learners' English textbooks. The process of eliciting the test items from participants' textbooks will be explained in detail in a separate section in this chapter later (Section 4.6.3).

The second step was conducting pilot studies. Two pilot studies were conducted prior to commencing the actual data collection and analyses of the present study. The first pilot study was carried out on native speakers of English. The second pilot study was conducted on non-native speakers of English. The pilot study on the native speakers aimed at selecting appropriate collocations which would be used to investigate participants' receptive and productive knowledge and development of collocations. This pilot study consisted of two phases. In the first phase, English native speakers were asked

to decide on the appropriateness of a large number of combinations. All the combinations were elicited exclusively from learners' English textbooks. More explanations on this part of elicitation will be provided in this chapter later (Section 4.7.1). Participants of the study were mainly English teachers who were teaching English to L2 learners in one of the English language learning centres in Exeter. Some of them apologised while others were happy to participate in the process. Their task was to decide on the appropriateness of the candidate collocations. The process resulted in a list of collocations which were employed to test participants' knowledge of collocations. There was no time limit for this part of the study and participants could make their decisions on the appropriateness of the combinations at their ease and time or return the list on the second day.

The second pilot study on the native speakers of English aimed at identifying any difficulty in deciding on the appropriateness of the items which would be used in the actual tests. It also aimed to identify any issues in the gap-filling test items in terms of clarity and filling the gaps. The process resulted in diagnosing few cases of ambiguity in the contextual sentences in the gap-filling test. These sentences were modified or replaced by clear sentences. All the sentences were taken exclusively from participants' English textbooks. Similarly, there was not a time limit in this study since the aim of this stage was not to identify or check the time they take for doing the tests. Accordingly, the time they took to do the tests had no influence on the process results.

The pilot study on the non-natives, i.e., participants other than the study sample, aimed to check any difficulties, issues, and ambiguities in the test items. Additionally, this pilot study examined the sufficient time that should be allocated for learners to conduct the actual data collection tests. The results indicated that 80 minutes would be sufficient for conducting both tests. More detailed information on the pilot studies on the native and non-native participants will be explained in later sections (Section 4.7.2).

The third step was to recruit participants for the study. According to the aims of the study, the recruited population should be high school students. I decided to involve a large number of participants, around 200 participants, in the current study. The rationale behind this rather large number was to give a generalisability power to the findings. Accordingly, I had to look for a high

school with a large number of students. Finding a high school with this large number of students was one of the challenges. Another challenge, in addition to the large number of students, was to find a co-education school, i.e., where both male and female students study in the same school to examine the potential influence of gender on the participants' collocational knowledge and development. It was not difficult to get male participants from a school and get female participants from another. However, to avoid the influence of two different learning environments on the participants' collocational knowledge and development, accordingly, on the study findings as much as possible, I found it better if all participants were from the same school. Eventually, I could find such a school after consulting the local Directorate of Education in the city of Duhok, which is one of the main cities in Kurdistan. All participants were recruited from one school who were all Kurdish. Being all from the same culture can make analysing the influence of L1 on their knowledge less complicated. The average length of studying English of the participants at the time of the study was 11 years, i.e., participants were Year 11 students. More details about participants' characteristics will be provided in later sections in this chapter (Section 4.5).

The tests were to be carried out in the learners' classrooms without the use of a dictionary under the direct administration of their own teachers. More details concerning the tests will be given in later sections. The time given for completing their tasks in the actual data collection tests was tailored based on the time was taken for the tests in the pilot study phase. The receptive and productive tests were to be carried out on the same date because it would be difficult to convince the school staff to allocate separate days for each test due to the large number of classes which would be involved in the study and the chaos and disturbance they might cause to the other classes and the whole school.

Finally, I had to decide on and identify the most appropriate statistical techniques which could be used to analyse the collected data. Choosing certain analytical tools rather others was based on a literature review of similar previous studies. Accordingly, a wide range of statistical techniques was used to analyse data and answer the study questions. The statistical and analytical software included SPSS 22 (Statistical Package for the Social Sciences), NVivo 10, CLAW 5, word and excel formulas and functions.

From SPSS software package, I utilised an independent-samples t-test to examine the influence of gender on the participants' collocational knowledge. I also used this test to explore the influence of the semantic transparency and L1 congruency on the difficulty of the items to the participants. The rationale behind its use was that in examining the influence of each of gender, semantic transparency, and L1 congruency on the participants' collocational knowledge, there are two variables. The first variable, e.g., gender, is an independent categorical variable which consists of two unrelated groups or levels (male and female), and the second variable, i.e. participants' collocational knowledge, is a dependent continuous variable. Accordingly, I decided, based on the literature review and study aims, that an independent-samples t-test would be an appropriate choice because if an independent variable has two levels, t-test is usually used.

As for investigating the correlation between frequency of occurrence and difficulty of the items to the participants, I employed a Pearson correlation coefficient. The Pearson correlation factor (r) shows the strength of the relationship between occurrences of the items in the textbooks and their levels of difficulty to the learners.

Developmentally, I utilised paired-samples t-test to examine participants' collocational knowledge development of the variables according to their features- frequency, syntactic structure, semantic transparency, and congruency with L1- such as examining participants' knowledge development of higher and lower frequency items, or congruent and incongruent collocations.

Finally, I used a one-way ANOVA test for examining the influence of the syntactic structure of collocations on the participants' collocational knowledge and development. The choice was due to that investigating this feature required comparing an independent variable which consisted of four groups, $v+n$, $adj+n$, $prep+n$, and phrasal verb collocations, with a dependent variable, participants' collocational knowledge. Conducting one ANOVA test on the four syntactic structures is better than running four independent-samples t-test, one for each syntactic structure which may result in more errors in the results.

Worth mentioning, I relied on the facility value means for examining the influence of frequency, syntactic structure, semantic transparency, and L1 congruency on the difficulty of the items, accordingly, participants' collocational

knowledge. Developmentally, I employed participants' total scores on the items for investigating the influence of gender and features of collocations on their collocational knowledge development.

Finally, the software NVivo 10 was employed to convert the pdf files into word file format to facilitate using the software CLAWS 5 in tagging parts of speech of the textbooks in the process of extracting collocations. Additionally, word and excel formulas and functions were used to find and count collocations and collocational items in the extraction phase. More detailed information on the collocation extraction phase and the use of the software will be explained in later sections (Section 4.6.3).

4.4 Design

In an endeavour to fill part of the gap in methodology and knowledge regarding collocational knowledge and development, the present study adopted a phrasal frequency-based perspective of collocations. Within a longitudinal design, the study attempted to track the collocational knowledge and development of Kurdish high school students and identify the influence of gender and four features of collocations (frequency, syntactic structures, semantic transparency, and congruency with L1) on the learners' knowledge and development of collocations.

Within the framework of the adopted design for the present study, the operational definition of the current study defined a collocation as a type of combination of two words or more that occurs more than once in a certain restricted syntactic structure with restricted combinability to perform a semantic or pragmatic function according to the context in which it is used.

In other words, for a combination to be regarded as a collocation, some criteria are needed to be applicable. First, in line with Kjellmer (1982) who holds the view that if a combination occurs more than once in the natural English, it is a collocation, the combination should occur more than once in English communication texts. Second, not any recurrent co-occurrence of a combination is a collocation such as *of the* or *although he*, i.e., the combination should be in a specific grammatical structure with restricted combinability (see Section 4.6.3: Table 2). However, regarding this study, only four patterns are examined- v+n, adj+n, phrasal verbs, and prep+n- which will be explained in detail in the following sections (see Section 4.6.3). Third, the combination should convey a

communicative message within native speakers' intuition, i.e., it should be meaningful to native speakers of English.

The current study investigated learners' knowledge of collocations receptively and productively. However, due to the absence of consensus on specific tests to measure learners' collocational knowledge and development, and that no single test is able to measure all aspects of the collocational knowledge, testing learners' receptive and productive knowledge and tracking their knowledge development were conducted by two types of tests. The underpinning assumption behind choosing two different types of tests was to elicit two different types of collocational knowledge, i.e., receptive and productive knowledge of collocations. The tests included measuring learners' receptive knowledge by utilising tests modelled on Gyllstad's (2007) COLLMATCH3, and testing their productive knowledge of collocations through gap-filling tests. The selected receptive and productive tests, Gyllstad's (2007) COLLMATCH3 and gap-filling tests, have been relatively proved to reliably measure these aspects of learners' vocabulary knowledge (Pignot-Shahov, 2012) in general, and the test design of the COLLMATCH3 has been found to be reliable in testing learners' receptive and productive knowledge of collocations, in particular (Gyllstad, 2007).

4.5 Participants

The study sample, in the first wave of data collection, consisted of 281 Kurdish high school students of which 98 were males and 183 were females. However, in the second wave of data collection, 14 participants left the study and 15 new participants joined the study. In order to be able to follow up participants' knowledge and development as accurate as possible, the data of both groups, 14 leavers and 15 new joiners, were deleted from the study data base. Moreover, 15 invalid cases were discarded as they seemed to be carelessly unfinished or finished haphazardly. This was evidenced in that participants in these cases had ticked all the items (✓) or (X) in a fast ticking way which was obvious from their handwriting, or ticking one and leaving two or three items without markings, or just left all of them unmarked. Accordingly, the study analyses were conducted on 252 participants of which 87 were males and 165 were females. All the participants were from Year 11 and their age ranged between 17 and 18 years old. Enrolling students into the present study required

prior permission from the related local authorities, Ministry of Education and Directorate of Education in the Kurdistan Region and their school administration (see Appendix A). Both sexes were included in the study to consider any potential influence of gender differences on their knowledge. However, equal numbers of participants from both genders were not attainable as females were much more than males. All students were given equal chances to participate or withdraw from the study. They were asked to partake voluntarily in the study without any differentiation or discrimination in terms of ability or gender.

4.6 Data collection

4.6.1 Introduction

Collocational knowledge develops differently from individual vocabulary knowledge and does not progress in parallel with it at the same rate of development (Bardel, Lindqvist, & Laufer, 2013). Moreover, defining collocations and deciding which combination to be considered true collocations required adopting a specific definition to collocations. Furthermore, the tests relied entirely on the collocations extracted from students' English curricula course textbooks. The rationale behind this strategy was to be sure that the collocations which were employed in their tests were exclusively those they had encountered as input in their school study years, accordingly, to measure knowledge of what they had studied not what they know in general. As, collocations were not handy to select from to do the tests, i.e., the curriculum textbooks did not provide lists of collocations, it required scanning the whole curriculum textbooks from level 1 to level 11 to elicit a list of the collocations required for the study tests. Textbooks 1-10 were those they had studied before starting the current study whereas textbook 11 was the textbook they were studying during the academic year of conducting the study. A further point, selecting collocations necessitated specific criteria to decide which collocations were more appropriate to achieve the study goals.

The employed criteria and the adopted procedures for identifying appropriate collocations, process of data collection, and stages they went through will be explained in detail in the following sections.

4.6.2 Curricula textbooks

The corpora from which collocations were extracted consisted of eleven school English course textbooks (Grade 1-11). These curricula textbooks have been

prepared by McMillan Education, and they are entitled “Sunrise” which begin with “Sunrise 1” and end with “Sunrise 12”. They have been taught in the schools of Kurdistan Region since the 1990s, and they have been revised several times. The whole curricula textbooks hold approximately 3069 key words, lemma forms, (2697 key words up to Grade 11) which are distributed over 12 school years as explained in Table (1).

Table (1): Statistics of the textbooks

Textbook	New vocabulary \ year	Aggregated vocabulary size	Number of extracted collocations
Sunrise 1	44	44	3
Sunrise 2	62	106	1
Sunrise 3	100	206	5
Sunrise 4	83	289	8
Sunrise 5	108	397	13
Sunrise 6	94	491	8
Sunrise 7	634	1125	10
Sunrise 8	399	1524	20
Sunrise 9	610	2134	25
Sunrise 10	249	2383	13
Sunrise 11	314	2697	22
Sunrise 12	372	3069	Not included
Total	3069	3069	128

The information and figures about new vocabulary shown above were taken from the textbooks as provided with each book. For the purposes of the present study, information from the separate textbooks was gathered in one table to produce the total statistics. As for the fourth column (Number of extracted collocations), they were computed manually.

As can be inferred from the table above, most of the vocabulary weight (2578 words= approximately 84%) is put on the secondary stage, i.e., years 7-12. Accordingly, most of the extracted collocations for the current study tests (90 collocations= nearly 70%) were from this stage and the rest were from the primary stage, i.e., years 1-6. Textbooks Sunrise 1-11 were included in the study. However, only textbook 12 was excluded from the study, because students at this level had ministerial exams, it would be difficult to convince the teachers and school administration to allocate at least five classes to explain and run the study tests at the expenses of completing their curricula textbooks. Moreover, students in this year would be busy with doing their school work to get as high marks as possible which qualify them to join the university they were looking forward to. The language of instruction to teach English at this school was mainly in English where Kurdish, their native language, was only used to explain vague and grammatical aspects of the contents. These course textbooks were the corpora for extracting the collocations that would be utilised to examine learners' collocational knowledge development up to Year 11.

4.6.3 Collocation extraction and selection procedure

Establishing on the collocation definition adopted for this study, collocations were identified according to their frequency of occurrence, syntactic structures, combinability, semantics, and native speakers' intuition. Operationalisations of these criteria are explained in detail in the following sections.

4.6.3.1 The operationalisation of the frequency criterion

The operationalisation of the frequency criterion was of two folds: the collocation frequency in the COCA corpora and its frequency in the learners' curricula textbooks. Frequency of collocations in the COCA corpora was utilised to decide on the occurrence of the candidate combinations in native speakers' usage. In other words, frequency was employed to check naturalness of the combinations compared to native English speakers. Collocation frequency in the curricula textbooks was used to identify the relationship between frequency of occurrence in the language to which learners were exposed, textbook inputs, and learners' collocational knowledge and development. The criterion of the collocation frequency in COCA was operationalised by checking the combinations against the COCA.

For the purpose of listing the steps of operationalising this criterion in a clear simple way, the followed steps will be stated first, and then the reasons for making the decisions I came to will be presented in later paragraphs. First, I considered the different forms of a lemma as frequent occurrences of the same items, e.g., the occurrences of *makes a suggestion*, *made a suggestion* and *making suggestions* were counted as three occurrences of the same collocation. Second, I allowed four words on either side as the span between nodes and collocates. Third, I checked their mutual information scores (MI) in COCA against the study threshold, i.e., three. For lexical collocations, if the candidate MI score was three or higher, the item would stay in the list, otherwise, the item was discarded from the list of candidate collocations. However, for grammatical collocations, the MI score threshold was reduced to two.

Accordingly, operationalising this feature, i.e., frequency of occurrence, was not a straightforward process. Frequency of collocations can be counted in various ways (Schmitt, 2010a). Moreover, none of these ways have been proved to be the most representative or connected with learners' collocational knowledge (Durrant, 2014). Hence, operationalising collocation frequency is surrounded by several issues such as deciding the span between the nodes and the collocates, counting the various forms of words, and measuring the strength of word association.

The space span between the nodes and the collocates is one of the most influential elements in counting collocation frequency. However, researchers have varied in their space tolerance depending on the adopted statistical models and counting methods. Some researchers have used automatic tools, i.e., computational methods using computers, which allow searching larger corpora (e.g., Brezina, McEnery, & Wattam, 2015), or by setting a specific space span which ranges from one word to four (e.g., Sinclair, 1991) or five words or more on either side of the nodes). Durrant (2014) argues that while too narrow span of searching collocations may result in losing a considerable number of genuine collocations, the risk of including pairs with no collocational relationship increase by widening the span. Accordingly, a balance between both extremes could yield a better search span. Following Sinclair's (1991)

claim of that the widespread span between nodes and collocates is about four words to the left or right, I adopted the span 4:4.

A further issue ahead of counting frequency is whether to deal with the items as words or lemmas, i.e., how to deal with the different forms of a word.

Lemmatisation is the process of “grouping together the inflected forms of a word so they can be analysed as a single item” (Collins English Dictionary, 2017).

Some researchers support considering the different forms of a lemma recurrent occurrences of the same word (e.g., Halliday, 1966) to avoid adding unnecessary complexity to the descriptive power. However, some other researchers prefer to see them as an occurrence of a different word (e.g., Clear, 1993; Hoey, 2005; Knowles & Don, 2004; Sinclair, 1991; Stubbs, 1996; Tognini-Bonelli, 2001) so as not to risk characteristics related to collocation preferences. For example, *argue strongly* and *argued strongly* can be counted as two occurrences of the same collocation or as an occurrence for each one of them (Durrant, 2014). Commonly, the various forms of a lemma indicate similar senses. However, in some cases, forms of a lemma are not used in the same senses (Sinclair, 2004). Moreover, the word forms which belong to a particular lemma may vary according to the complementation pattern occurring with that item as the case with verbs (Hunston, 2003).

Lemmatisation has been criticised by Knowles & Don (2004, p. 71) who argue that although the lemma has traditionally been utilised to make generalisations about the words in a ‘family’, the results of corpus analyses of recent studies have questioned the assumptions underlying such a practice. This is exemplified in that some individual members of the lemma can behave independently and develop their own meanings and collocations apart from the main base word, i.e., lemma. Additionally, they argue that the past participle of the verb “*provide*” i.e., *provided*, for example, plays a completely new role when it functions in a subordinating conjunction. Formally, it can still be a member of the same lemma, but direct inferences about distribution or meaning of the lemma member is no longer straightforward. Accordingly, different forms of the same verb, adjective, and noun may have different collocational profiles. Moreover, they state that although the lemma is a fundamental concept in the processing of texts in many languages, it is not necessarily the same case with the English lemmas as they are not typical of the general category.

Consequently, linguists who establish their understanding of the lemma on English texts may obtain a distorted view of the phenomenon. Hence, they believe that generalisations about whole lemmas become less convincing. Similarly, Tognini-Bonelli (2001, pp. 92-98) challenges the view that members of a lemma can share the same meaning and differ only in their grammatical profiles. For example, she compares the use of “*facing*” which has a concrete meaning, e.g., *facing forward*, and “*faced*” which has only the metaphorical meaning, e.g., *faced with a dilemma*. Accordingly, it might be hard to gain any significant generalisations by assigning *facing* and *faced* to the lemma *face* except their similarities in spelling and pronunciation. A further issue with lemmatisation is the existence of differences between the forms such as that exists between singular and plural nouns in terms of their most common collocates (Sinclair & Renouf, 1988). This indicates that different forms of the same lemma have different collocates. In other words, lemmatisation may not allow some collocational preferences between different forms of a lemma to appear (Clear, 1993; Hoey, 2005; Sinclair, 1991; Stubbs, 1996). For instance, the English may prefer using “*catch a bus*”, a singular noun, more than “*catch buses*”, a plural noun. It has also been criticised for being rather subjective when a researcher makes judgements on the items and their uses (Sinclair, 1991). Subjectivity in this respect implies that assigning a certain use to a collocation by a particular researcher may differ from that is made by another researcher. Consequently, a collocation may belong in one meaning domain if both researchers agree on the collocation use or differ in their assignments of the collocation use, accordingly, they belong to two different domains. Hence, subjectivity has an influence on frequency count of the items in the corpus. A further point is that meaning may not be persistent across all word forms of a lemma. For example, *certain* and *certainly* are used by natives quite differently in terms of their meanings and usages (Sinclair & Renouf, 1988). An additional disadvantage of lemmatisation is that, particularly in making a frequency list, even where meaning is maintained across the different word forms, information concerning the frequency of occurrence of forms of a lemma in the corpus will not be obtainable. This is because all forms are placed under one lemma. Finally, Sinclair (1991) argues that taking the concept of lemma as being the base form of a word is doubtful because the most frequently encountered word form could equally be regarded as the lemma.

However, despite these disadvantages, I adopted the lemmatised counting due to its appropriateness to the current study, i.e., I chose to deal with collocational components as relationships between lemmas rather than word forms. The rationale behind this decision can be explained in several points. First, the decision is supported by the assumption that counting based on lemmatisation may provide more information about learners' knowledge than counting based on unlemmatised word forms (Durrant, 2014; Halliday, 1966). Accordingly, lemmatisation can provide information about learners' grammatical and spelling skills of the language. For example, it shows that whether learners were familiar with the different inflectional forms and spellings of the words. Second, taking words rather than lemmas may reduce number of instances of some collocations to the extent that making comparisons between higher and lower frequency items become difficult because there will be very few instances of each and they will be too close to each other in number. Third, a sufficient number of collocations may not be available to do the tests due to the small size of the corpus from which items are to be elicited, in this case, participants' school textbooks. Fourth, lemmatisation may reduce the number of units with the same use as it prevents repeating collocations constructed from the same lemmas but in different word forms (Nation, 2013). Finally, participants' course textbooks contained many collocations with different word forms, but indicating the same sense. Additionally, none of the extracted collocations with the same lemmas of different word forms were used in different senses.

Additionally, lemmatisation was found more appropriate for achieving the current study aims than word forms since I could deal with most of the critiques addressed to lemmatisation. First, exploring learners' collocational preferences of the different forms of a lemma was not one of the study aims. Accordingly, this disadvantage would have no negative influence on the study findings. Secondly, to reduce the influence of researcher's subjectivity as much as possible, I relied on the natives' intuition to decide on the use of a collocation, accordingly, I did not rely on only my intuition, as a researcher. Thirdly, I checked manually the different forms of words to make sure that the different forms of the lemmas in all the collocations had maintained the same senses. Finally, to avoid excluding any collocations from counting, I checked manually the frequency of all the collocations in the different word forms in the textbooks.

Another issue of counting frequency is the measure utilised in quantifying frequency of collocations. Different approaches can be used for counting frequency of collocations. One of these ways is to compute the number of occurrences of a combination. However, this approach has been criticised for underscoring very high-frequency combinations such as (*of the, and a*), which occur frequently by chance, and underestimating genuine collocations which appear to be of low-frequency such as *abject poverty* or *battering ram* (Durrant, 2014). To overcome this issue, some measures have been used by researchers amongst the most widely used measures are the “t-score” and “mutual information” (MI). According to Clear (1993, pp. 279-282), “MI is a measure of the strength of association between two words”, whereas t-score is “the confidence with which we can claim there is some association”. In these two measures, given the frequency of each individual word, the actual frequency of co-occurrence of a pair of words is compared with the expected frequency of their co-occurrence by chance alone. The expected frequency *E* is calculated by employing the following formula:

$$E = C * \frac{w1 * w2}{C^2}$$

where C= total number of word tokens in the corpus

w1, w2= frequencies of the collocational constituent words

Accordingly, t-score is calculated using the formula:

$$t = \frac{O - E}{\sqrt{O}}$$

whereas MI is calculated by the formula:

$$MI = \log_2 \frac{O}{E}$$

where O= the observed frequency of a combination (Durrant, 2014).

Clear (1993, pp. 279-282) explains the use of t-score and MI in two examples of collocations. He argues that the collocation *taste arbiters* is of a high MI but of low frequency. Accordingly, the two words are so strongly associated that the appearance of *arbiters* gives a great chance also to find *taste*. However, the generalisability of the association of the two words cannot be guaranteed due to the frequency of the collocation. This implies that either component could be encountered in other samples of language. As far as t-score is concerned, Clear (1993) explains it in the collocation *taste for which is of high t-score* as an

example. In contrast to the collocation example for MI measure, the generalisability of the association between *taste* and *for* can be confidently stated. Although the association between these two words is weaker than that between *arbiter* and *taste*, i.e., neither word is a strong predictor of the other, the combination occurs much more frequently than the combination *taste arbiters*.

As far as the study is concerned, I used MI to measure the association strength between the combination components to decide on their collocationality. The use of word associations is proved to be vital in the aspects which are related to L2 vocabulary research and measurement (Schmitt, 1998c). Accordingly, MI will be explained in more detail.

Commonly, mutual information measures the amount of information that the observation of one random variable yields about the probability of the occurrence of another random variable word (Clear, 1993; Cover & Thomas, 1991). It calculates the collocation strength, or the association strength between two words, by contrasting the actual co-occurrence of the combination items in the corpus with the anticipated co-occurrence of the items if the corpus words were to occur in a completely random order (Bestgen & Granger, 2014). Accordingly, MI calculates the non-randomness of co-occurring of two items together. Additionally, it has the ability of identifying collocations of low frequency but with strong bonding co-occurrence (González Fernández & Schmitt, 2015). The MI value is dependent on calculating number of instances of the co-occurring word in a specific span of the node word (the observed frequency), and calculating the anticipated instances of co-occurrence in the same span where the co-occurrence frequency of the word in the corpus is given (Hunston, 2002). In other words, knowledge about one of the elements of a collocation contributes to guessing the other element since it implies some data about the potential items that can be associated with, i.e., the amount of knowledge of the first element of a collocation that can be expected by calculating the second word and vice versa (Cellucci, Albano, & Rapp, 2005). For example, L2 learners' familiarity with English collocations partially helps them predict the first verb element "*make*" of the collocation "*make a decision*" if they are given the second element "*decision*" in or out of context. Mutual information measure is used in the COCA to calculate the strength of

association between items. In COCA, MI is calculated according to the following equation:

$$MI(w_1, w_2) = \log((AB * \text{sizeCorpus}) / (A * B * \text{span})) / \log(2)$$

where A=frequency of node word w1;

B=frequency of collocate w2;

AB=frequency of collocates near the node word; sizeCorpus=size of corpus;

span=span of words (e.g., three to the left and three to the right of the node word: six); $\log_{10}(2)=0.30103$ (Neuman et al., 2013, p.2).

The high MI score values indicate strong associations between items and low scores imply weak associations between them. Accordingly, approaching the value 0 means that the co-occurrence is likely by chance whereas negative values indicate that the items avoid each other (McEnery, Xiao, & Tono, 2006). The default in COCA is that an MI score of three or higher is considered as significant. In line with the COCA threshold of deciding the significance of the relationship between the collocational elements, Hunston (2002) suggests that a collocation with an MI score of three or higher is statistically significant, i.e., a strong collocation.

However, it is believed that information about the collocation strength is not always a clear-cut and reliable indication of a meaningful association (Hunston, 2002). Additionally, knowledge about how much evidence there is for it is also required, i.e., knowledge about collocation certainty and that there is enough evidence in the corpus that shows the co-occurrence of two items is more than the fancies or inclinations of a corpus. Mutual information measure has also been criticised for its biasedness towards collocations of infrequent words and its emphasis on rare words or combinations (Bouma, 2009; Kilgarriff & Kosem, 2013; Vechtomova & Wang, 2006). This raised a problem for the present study because the prep+n and phrasal verb categories necessarily involve very high frequency words (i.e., prepositions and particles). It was found that, for this category, many items which would intuitively be considered collocations had MI scores of lower than three such as *go on* (MI 2.83), *get up* (MI 2.26), *on foot* (MI 2.47), and *on a picnic* (MI 2.16). Moreover, sufficient instances of combinations of MI score values of three or above could not be found in the textbooks. For this reason, the threshold was lowered for these items, to include combinations of down to MI score value of two.

Establishing on this, in the current study, an MI score value of three or more was principally considered as the minimum for deciding on the significance of the association between two items. Accordingly, combinations of lexical words with MI scores of three or higher were regarded collocations. For combinations involving grammatical words, a slightly different approach was taken, the threshold was two or higher, as was explained earlier. Operationalising the candidate items according to their frequency in the curricula textbooks into high and low frequency items will be explained in the next sections (Section 4.6.4).

4.6.3.2 The operationalisation of the syntactic criterion

Syntactically, the syntactic patterning of the combinations was based on Benson, Benson, and Ilson (1986), the BBI Combinatory Dictionary of English, and Gitsaki's (1996) classifications of collocation types as shown in the table below.

Table (2): BBI's and Gitsaki's collocation types and extracted items

No.	Collocation patterns	Collocation types	Examples	Sources	Number of Instances
1	v+n	Lexical	launch a missile	BBI 1986/ Gitsaki 1996	1662
2	adj+n	Lexical	strong tea	BBI 1986/ Gitsaki 1996	1134
3	n+v	Lexical	bees buzz	BBI 1986/ Gitsaki 1996	144
4	n+prep+n	Lexical	a piece of advice	BBI 1986/ Gitsaki 1996	324
5	adv+adj	Lexical	deeply absorbed	BBI 1986/ Gitsaki 1996	146
6	v+adv	Lexical	affect deeply	BBI 1986/ Gitsaki 1996	245
7	n+n	Lexical	aptitude test	Gitsaki 1996	457
8	n+prep	Grammatical	argument about	Gitsaki 1996	429
9	n+to+inf	Grammatical	pleasure to do it	Gitsaki 1996	19
10	prep+det+n	Grammatical	on the contrary	Gitsaki 1996	1227
11	adj+prep	Grammatical	angry at	Gitsaki 1996	132
12	phrasal verb	Grammatical	pass on	Gitsaki 1996	1486

The syntactic structure patterns listed in Table (2) were used as filters for identifying the combination types used in the textbooks. I could download all the

textbooks from the Ministry of Education official website which were in the form of pdf files. Then, they were transformed into word file format by using a pdf to word file converter. Tagging and classifying the whole curricula textbooks were conducted manually and automatically. The purpose behind this mixed filtering process was to avoid human weakness in such tasks and to refine the automatic process as well through manually selecting exclusively targeted combinations. Automatically, the procedure started with tagging the selected text using CLAWS (The Constituent Likelihood Automatic Word-tagging System) to identify the part of speech to which each item belonged. Manually, I read through each text and checked the automatic tagging of the items. Then, I listed the identified combinations under the syntactic patterns they belonged to. Finally, the frequency with which each combination appeared was recorded. Accordingly, the filtering process showed that v+n, adj+n, phrasal verbs, and prep+n combinations were the four commonest types used in the school textbooks. Accordingly, these four patterns were used as the candidate collocational patterns for evaluating learners' collocational knowledge and development in the present study. It can be noticed from Table (2), pattern 10, that prep+n and prep+ det+n are considered as belonging to the same pattern. The rationale behind this is that collocations are seen in this study as a relationship between lemmas (e.g., Durrant, 2014; Halliday 1966) and not word forms (e.g., Clear 1993; Hoey 2005; Sinclair 1991; Stubbs 1996).

4.6.3.3 The operationalisation of the combinability criterion

As for combinability of the collocational items, the operationalisation of this criterion was established based on the conception of combinability of Cowie (1981), Howarth (1996, 1998), and Nesselhauf (2005) who define combinability to refer to degree of restrictedness of the paradigmatic substitution of the elements of a combination (see Section 3.4.3). For example, in the collocation *make/take a decision*, the noun *decision* is rather restricted in its combinations with verbs because it can be combined with two verbs: *make* or *take*. Conversely, the verbs *make* and *take* are not restricted because they can be used in various senses, and they can be combined with many nouns. Accordingly, in the current study, a combination which is restricted by selectional constraints imposed on at least one element is recognised as a collocation. Hence, collocations occupy the position between completely free

combinations and fixed idioms. Collocations were differentiated from free combinations and idioms in being more fixed than the former and less fixed and having more freedom to be substituted by other synonyms than the latter. Worth mentioning, this criterion was operationalised along with the semantic transparency in identifying collocations. Accordingly, combinations which were characterised by the following characteristics were considered collocations: [+ Restricted Combinability] / [+ Transparency], [- Restricted Combinability] / [- Transparency], [+ Restricted Combinability] / [- Transparency]. From a combinatory point of view, a lexeme in a syntagmatic relation is either replaceable with a synonymous lexeme or such replacement is highly constrained. Semantically, a lexeme could be transparent or semi-transparent or non-transparent. If a combination consists of lexemes which are both semantically transparent, replacement of the items should show a degree of restrictedness to be recognised as a collocation. In sum, identifying collocations were in terms of lexeme-replacement and semantic transparency. Accordingly, the process produced two groups of items. The first group was used for constructing the tests whereas the second group of items, non-collocations, were discarded from the list.

4.6.3.4 The operationalisation of the native intuition criterion

The intuition of native speakers was used to verify the occurrence of the selected combinations in the natural language of natives as collocations. First, the list was cut down by discarding combinations of MI score values lower than two for grammatical collocations and three for lexical collocations. Three native English speakers, who were teachers of English as a second language, agreed to go through the candidate collocations and decide on their collocationality. Each participant was given the same list of combinations that contained about 400 items. They were also informed of the study collocation definition adopted provided with an example “a type of combination of two words or more that occurs more than once in a certain restricted syntactic structure with restricted combinability to perform a semantic or pragmatic function according to the context in which it is used”, for example, English say “*heavy rain*”, but not “*massive rain*”. As the definition was rather technical, one of the three consulted natives found a little difficulty in understanding what was meant by the term “collocation”. Hence, the term was explained more for him. When two of the

three participants rejected a combination, it was discarded and when they agreed on it, it was kept. To obtain as high data quality as possible from the raters, a training procedure preceded their actual rating. For example, I asked the participants to decide on the combination “*heavy rain*” as to which of the three combination types it belonged to: free combination, collocation, or idiom. For this, I followed several steps. First, I asked them to decide whether the combination was generally acceptable or unacceptable in the English usage. Second, I asked them to decide whether it was a collocation; where one of the items has a limited degree of substitutability, or it was an idiom; where both items are strictly fixed, or it was a free combination; where both items could be freely replaced by others. The intended feedback was to produce either it was not a collocation, accordingly, discarding it from the list, or marking it as a collocation which necessitated to mark it as either transparent or opaque. Finally, the extraction process resulted in the selection of the items shown in appendix (B) which were employed in the study tests.

4.6.4 Classifying collocations

The elicited collocations had to be classified accord to the four features under focus- frequency, syntactic structure, semantic transparency, and congruency with participants’ L1- prior to dividing them into two equal groups for the data collection tests.

Classification of collocations according to the feature of frequency was based on times of occurrence in the participants’ school textbooks. The counting was made automatically and rechecked manually for verifying the obtained number of occurrences. In the automatic counting, the function “Search in Document” was employed to find and count number of occurrences. However, this method was not completely accurate since it could not count the different forms of the same lemma. To fix this issue, the different forms were searched individually then summed together. Finally, frequency was manually checked for each occurrence through going over the occurrences of each item to avoid missing any cases.

After counting the occurrence or occurrences of all collocations, the items had to be classified according to frequency based on a threshold that would be set for this purpose. However, the decision on which number of occurrence to be high and which one to be low was controversial and not straightforward to make

for two reasons. First, there is no consensus among researchers on a particular threshold of frequency as which number of occurrence to be counted as high and which one as low frequency (e.g., Coxhead, 2000; Durrant, 2008: Nation, 2013). Second, classifying collocations according to this feature was not isolated from the other features if I wanted to maintain balance between the four features in terms of number of items with each feature (see Table 3).

Accordingly, I had two options in classifying the items into high and low collocations. The first option was to set an arbitrary frequency threshold below which collocations are to be classified as low-frequency items whereas high-frequency collocations would then be those which come above the threshold. However, Durrant (2008) criticises such a threshold due to lacking any principled means of setting such a threshold. Moreover, small differences in frequencies could result in collocations falling above or below the threshold. For example, if I set the threshold at 10 occurrences, collocations with 9 occurrences would be regarded as low-frequency items whereas those with 10 as high-frequency items while the difference is only one occurrence. Accordingly, I abandoned this option.

The second option, which avoided the shortcoming of setting an arbitrary threshold of the first option, was to rank the items according to their frequencies. Accordingly, a list of items arranged according to their frequency (using log frequency instead) would be compared with their facility values to examine the influence of frequency on the difficulty of the items. In accordance, the expression "*items with lower or higher frequencies*", which did not imply adopting a particular threshold, instead of "*low-frequency or high-frequency items*", which implied adopting a definite frequency threshold, was used throughout the study in analysing the relationship between frequency of occurrence of collocations and participants' knowledge of those items. Finally, I chose the second option to avoid the disadvantages addressed to the first option.

However, to analyse the influence of frequency on the participants' collocational knowledge development, I needed to differentiate between lower and higher frequency items. Since I did not have a threshold, according to the adopted method of classifying items in terms of their frequency, I found it more appropriate to divide items into almost two equivalent groups. As number of

items was odd, 57 items, one of them would be eventually more than the other at least by one item, e.g., 28 to 29. Interestingly, the application of this criteria resulted in yielding the frequency 3 as threshold, though it was not planned to set a threshold, in both receptive and productive test items. Accordingly, items of three occurrences or less were treated as items of lower frequencies whereas those of four or above were counted as items of higher frequencies for analytical purposes. Three occurrences or less of an item meant that participants had encountered the item only three times throughout 11 school years, i.e., less than once every three years. Taking the time length, 11 years, into account, this threshold number could be considered as a low occurrence of an item in pedagogical contexts. Accordingly, the tests receptive 1 and productive 2, which consisted of the same items, were divided into 27 lower frequency and 30 higher frequency items whereas the tests receptive 2 and productive 1, which consisted of the same items, were divided into 29 lower frequency and 28 higher frequency items. These classifications of the items were the bases of analysing participants' collocational knowledge development. Syntactically, operationalising the lexicality or grammaticality of the collocational patterns was mainly based on Benson, Benson, and Ilson (1986), the BBI Combinatory Dictionary of English, and Gitsaki's (1996). Accordingly, grammatical collocations were defined and operationalised as a structure that consisted of a noun, adjective, or a verb plus a particle such as a preposition. Lexical collocations were operationalised as involving the same elements noun, verb, adjective, or adverb but without grammatical elements (see Table 2). Accordingly, the main patterns used for the current study were operationalised as follows: v+n and adj+n collocations were classified as lexical patterns whereas phrasal verb and prep+n collocations were classified as grammatical patterns.

Semantically, transparency has been utilised to differentiate between types of collocations as whether elements of collocations are fully used in their literal senses, which can be called transparent collocations, or they are used in their figurative senses, in minimum one item, which can be regarded as opaque collocations (Kurosaki, 2012). Although collocations are commonly perceived to be transparent in comparison with idioms, they vary in their degrees of transparency (Trantescu, 2010). In other words, meanings of collocations are

not necessarily derived from the compositional meanings of their constituents. In this respect, Bartsch (2004) suggests a scale of transparency in collocations that ranges from being fully transparent to superficially transparent collocations. Accordingly, fully transparent collocations refer to those whose constituent components contribute to the overall meaning of the collocations with their compositional senses. Superficially transparent collocations encompass those with an additional element of meaning which is not carried by any of the constituent components, usually culturally bound, and is only perceived from the whole meaning of the combinations. However, it seems difficult, in practice, to define a clear-cut boundary between transparent and opaque collocations.

In the present study, the semantic transparency criterion is operationalised to distinguish transparent from opaque collocations as follows. Transparent collocations [+transparent] are collocations in which elements are used in their literal senses and the overall meaning could be perceived from their constituents. Opaque collocations [-transparent] are those whose elements, at least one, are not employed to denote their lexical meanings, or the overall meaning of combinations cannot be retrieved from the lexical meanings of the constituents. The literal meaning of a word or a combination refers to its most basic meaning, i.e., the compositional meaning or sense of the word and it reflects the genuine or essential characteristic of someone or something whereas the non-literal sense of a word denotes the figurative or metaphoric (Cambridge Advanced Learner's Dictionary, 2015), technical or delexical meaning of someone or something (Howarth, 1996, 1998). Classifying the extracted candidate collocations into transparent and opaque went through several stages. First, lexically, a collocation of MI value of three or higher or a grammatical one of MI value of two or higher, was chosen. Second, the intended meaning was checked against COCA by comparing the sentential contexts in which it was used to check its frequency of use by native speakers. Third, if the constituent items of the collocation were used in their compositional meanings in the context, it was classified as a transparent collocation. If the constituent items of the collocation were not used in their compositional meanings in the context or one item completely or partially lost its compositional meaning, it was classified as an opaque collocation. Accordingly, the whole list

of collocations was classified into transparent and opaque collocations, Appendix (C).

Finally, the operationalisation of the feature of congruency with L1 was performed by classifying collocations, based on Yamashita & Jiang's (2010) definition of congruency, into congruent and incongruent items. A congruent collocation is defined in the present study as a combination which has the same lexical items in both L1 and L2, i.e., an English collocation which has a counterpart in the Kurdish language, and can be translated word for word from English to Kurdish. An incongruent collocation is defined as a combination which has different lexical items in both languages. The incongruent collocation cannot be translated directly word for word from English to Kurdish. After defining congruent and incongruent collocations, several steps were followed to classify collocations into congruent and incongruent items.

To operationalise the definition, I gave the elicited list of collocations to a group of native teachers of Kurdish language with good level of English proficiency. I asked them to translate the English collocations into Kurdish. Accordingly, if the translation of a collocation rendered literal meanings word for word it was classified as a congruent collocation. In contrast, if the translation rendered different lexical items it was classified as an incongruent collocation. In conclusion, the process resulted in classifying the list into congruent and incongruent items. However, number of items was larger than the required one. Accordingly, some items were discarded to suit the test structures and achieve the study aims.

4.6.5 The receptive knowledge tests

Two instruments were designed by Gyllstad (2007) to measure learners' receptive knowledge of collocations: COLLEX 5 and COLLMATCH 3. However, I decided to choose COLLMATCH 3 over COLLEX 5. Many reasons were behind this choice. First, COLLMATCH 3 is more practical and easier to administer than COLLEX 5. Second, it contains a larger number of items and collocation types and produces more meaningful and analysable interval data than COLLEX 5 (Gyllstad, 2007). Third, the validity and reliability of the test have been demonstrated in some previous studies (e.g., Bergström, 2008; Bueraheng & Laohawiriyanon, 2014; Gyllstad, 2007). These studies showed that scores on this test were well connected with learners' receptive knowledge

of collocations. Fourth, the multiple-choice format of the test is favourable by test-takers as it consists of two choices and they should choose the correct answer by simply ticking what they believe is the correct answer. Accordingly, it does not need as much time as writing compositional answers. Fifth, it is also quick to sit and can be easily scored. Finally, the test results can be used as proficiency and research tools to measure learners' receptive knowledge of collocation.

For the reasons and advantages mentioned above, it was decided that the COLLMATCH 3 format was convenient to achieve the study aims and seeking answers to its questions. Originally, this test is in the form of pen and paper. The test is a binary forced-choice Yes/No format that originally consists of 100 items. In this test, participants have to decide on the appropriateness of the combinations as true collocations as it is explained in the example below:

If you think a word combination exists in use in the English language, tick the 'yes' box. If you do not think a word combination exists in use in the English language, tick the 'no' box.

- 1 pick a glance yes
 No
- 2 accept differences Yes
 No

However, the test format was slightly modified to suit the study settings and aims. The modifications were mainly in terms of its number of items, collocation types, and patterns included in the test. Numerically, the one hundred items were reduced to 64 to fit the time limit and aims of the test. The rationale behind the number "64" was to have equal number of each collocation types, lexical vs grammatical, transparent vs opaque, and congruent vs incongruent. The rationale behind the time limit was that I could obtain permission from the school administration to use only two periods for each wave of data collection no more, which meant 80 minutes.

As for collocation types and patterns included in the test, they will be explained in later sections (Section 4.6.7). Additionally, for technical simplicity, the two boxes of 'yes' and 'no' were reduced to only one where to tick (✓) for the correct collocations, those exist in the English language use, or (X) for the

incorrect collocations, those do not exist in English language use as explained in the example below:

This part of the test consists of 64 items (1-64). Decide whether the following combinations are naturally and frequently used in English language. Tick (✓) for Yes or (X) for No in the box:

1 pick a glance

2 accept differences

Distractors were created from textbook collocations by substituting one of their components, mainly nodes, with inappropriate synonyms or items. Finally, to avoid misunderstanding the aim of the test, the test instructions were translated into learners' L1, Kurdish. Accordingly, the receptive tests format became as shown in Appendix (D).

4.6.6 The productive knowledge tests

Learners' productive collocational knowledge was measured by utilising a gap-filling test. The test was in the form of pen and paper. Each test consisted of 64 sentences with gaps to be filled with the missing words, collocates, to produce appropriate collocations. The contextual sentences were the actual contextual occurrences of the collocations in the participants' curricula textbooks.

However, new contextual sentences were constructed where they were necessary. In order to limit number of the potential answers or guessing untargeted collocations, the missing elements were primed by their first letters as shown in the following example:

Fill in the gaps in these sentences with the most appropriate nodes or collocates guided by the priming letters given for each blank.

Where did Ed m..... a mistake?

As for keeping balance among the items in terms of their collocational features, the same balancing criterion which was applied on the receptive tests, was also applied on the productive tests. Finally, to avoid misunderstanding aims of the tests, the test instructions were translated into learners' L1, Kurdish.

Accordingly, the productive tests format became as shown in Appendix (E).

4.6.7 Collocation types of the tests

Gyllstad's (2007) COLLMATCH 3 original version of the test includes two patterns of lexical collocations: adj+n and v+n. However, in order to be able to explore learners' knowledge of both lexical and grammatical collocations, the

test was modified to include two lexical and two grammatical patterns of collocations. The lexical patterns were v+n and adj+n collocations whereas the grammatical patterns consisted of v+prep and prep+n collocations. All the 128 items, 64 items of the first test and 64 items of the second test, were selected from those existed in the textbooks after operationalising the study criteria as was explained in the previous sections. Some of the items were slightly modified in terms of their structure such as adding an article. For example, *making suggestions* was transformed into its singular case *make a suggestion*. This was because the latter form was used in their textbooks rather than the earlier. Accordingly, matching forms of the items with the form of their existence in the curricula textbooks might assist and improve their recognition and recalling of those collocations and provide more accurate account of the learners' collocational knowledge.

Additionally, the balance between each group of collocations according to their features was maintained except with frequency since I dealt with it differently, as explained in Section (4.6.4). For example, syntactically, each test contained equal number of lexical and grammatical collocations; 32 lexical collocations and 32 grammatical collocations. The balance was also preserved in the number of items in each sub-type. For example, lexical collocations comprised 16 v+n and 16 adj+n collocations; similarly, grammatical collocations contained 16 v+prep and 16 prep+n collocations. Similarly, the balance was maintained in the numbers of items in terms of their congruency with L1; 32 congruent and 32 incongruent items. The same ratio was kept equivalent in terms of their transparency, 32 transparent vs 32 opaque collocations. In its turn, each type involved equal numbers of items in terms of their syntactic structure, semantic transparency, and L1 congruency. For example, v+n collocations were comprised of eight transparent collocations and eight opaque collocations. Transparent collocations were subdivided into four congruent and four incongruent items. In the same way, opaque collocations consisted of four congruent and four incongruent items. The balance among the different collocation types and sub-types was maintained as explained in the table below, Table (3), which is about v+n collocations items as an exemplary.

Table (3): Balance among the items

v+n collocations	Transparent		Opaque	
	Congruent	Incongruent	Congruent	Incongruent

	4	4	4	4
	4	4	4	4

The same distribution pattern was applied on all the other collocation patterns and types.

The initial process of eliciting combinations from the curricular textbooks resulted in identifying a large number of combinations. However, operationalising the criteria of defining collocations according to the definition of the study resulted in disregarding numerous combinations. Moreover, classifying collocations according to the features of frequency, patterns, types, transparency, and L1 congruency led to distributing collocations over many divisions and subdivision. Consequently, the remained collocations were not sufficient to provide enough items for two separate versions of tests. The valid collocation items, according to the study criteria, were only enough for one version test in which a balance could be maintained among all the items in terms of features and numbers of items in each feature. Hence, two options were ahead of me to choose one of them. The first option was to repeat the same test in the second wave of data collection. This option would have allowed to follow-up learners' receptive and productive collocational knowledge and development of the same items over the school year. It would have also facilitated analysing the findings since both tests would have been conducted on the same items. Accordingly, comparing learners' collocational knowledge and identifying the influence of features on their knowledge and development could be more feasible. However, this option was disregarded to avoid the influence of repetition on the test. Repeating the same items in the same test format and context could have had a negative influence on the second test results. This was because learners might have gained experience and learned the answer of some of them through discussing the possible answers with their peers after the first test. Consequently, providing inaccurate picture of their actual knowledge of the items. Moreover, repeating the same test twice could have influenced their motivations to do the test, consequently, learners might have performed the second test less enthusiastically and carelessly.

The second option was to switch collocations; using collocations of the receptive test 1 in the productive test 2 and using those of the productive test 1 in the receptive test 2. The second option was chosen for the current study for

several reasons. The first reason was to avoid repetition and its negative influence. The second reason was to make the test more interesting and avoiding their boredom and increasing their motivation for doing the test. Thirdly, swapping the items would allow measuring learners' receptive and productive knowledge of the same items. Accordingly, two types of comparison would be feasible. The first comparison could be made between learners' receptive and productive knowledge of the same items. In the second case, the influence of the context on the learners' collocational knowledge could be contrasted; guessing collocations in the receptive tests was decontextualised whereas in the productive test it was contextualised in sentences.

However, this option did not allow following-up learners' knowledge and development of the same items. Rather, the comparison had to be made between learners' knowledge of two different sets of items. Accordingly, this option might have led to more complication in the data analyses. A further anticipated limitation was the difference in difficulty of the items of the receptive and productive tests. If items of the receptive test 1 were more difficult than items of the productive test 1, assuming the same level of learners' knowledge in both tests, it might show learners doing better on the receptive 1 and productive 2 and perform worse on the receptive 2 and productive 1. This could weaken the analytical power of the obtained results.

Regardless of these limitations, I believed that this option would be more appropriate than the first option. Accordingly, the 64 items which were used in the first receptive test were used in the second productive test. Conversely, the 64 items which were used in the first productive test were used in the second receptive test as shown in Appendix (F).

4.7 Pilot tests

A pilot study was carried out to stand on the different aspects of the tests and identify the potential issues in the tests concerning their contents and time and any contextual or procedural factors that might influence the process. The pilot tests were of two types. The first test type was conducted on two native speakers of English, a man and a woman who were living in Kurdistan during the data collection period of the study. The second pilot test was conducted on about 40 male high school students. These two tests are explained in the following sections.

4.7.1 Native pilot tests

As candidate collocations were verified by native speakers in the filtration process, two more native speakers were asked to sit the test to examine the content validity of the contextual sentences of the missing words. The natives faced difficulty in guessing the missing collocational components of a few items. The first native participant could not guess four items whereas the second one failed in five of them. However, they found difficulty with different items, i.e., their mistaken items were different, except in the collocation *average temperatures*. The first native filled the blank with *air temperatures* whereas the second one filled it with *Arctic temperatures* instead. Consequently, the contextual sentences were rephrased and replaced by clear sentences.

4.7.2 Non-native pilot tests

A pilot test was conducted on 40 high school participants. The aims of the test were, first, to check the time length required for both the receptive test, i.e., the COLLMATCH3 test, and productive test, i.e., the gap-filling test. Secondly, it was intended to identify any shortcomings in the structure of the tests in terms of spelling mistakes, blank spaces for each gap, tick boxes for each collocation, and providing a meaningful context sentence for each item. Third, it aimed at identifying any unexpected issues that might prevent students from sitting both tests appropriately. Fourth, it also sought to take into consideration English teachers' and students' comments and suggestions about the tests. However, their comments, with appreciation to their ideas, did not add any new thoughts to the existent ones.

On the whole, participants' answers of the gap-filling tests were characterised by making a lot of spelling mistakes which seemed to follow a specific pattern; most of the mistakes were in the word class forms. Additionally, the majority of the students appeared to have difficulty in answering specific items; some sentences seemed to be slightly semantically ambiguous. To address this issue, those sentences were rephrased or substituted by clear ones.

The test was conducted under my personal supervision with the help of their English teachers. The teachers did not have any role in explaining the test aspects. Their role was exclusively disciplinary to keep order in the class, and they also helped in distributing the test papers. Participants were given sufficient time to finish their tests and they needed less than two class periods

(80 minutes) to do both tests. Accordingly, two class periods were taken as the time basis for conducting the actual data collection tests.

4.8 Coding and scoring the tests

Raw data cannot be interpreted if not preceded by coding and analysing. Accordingly, two phases proceeded demonstrating the results of the study. First, raw data had to be coded as an essential and prerequisite step to the second one. Secondly, coding data enabled analysis by utilising appropriate data analysing tools. Accordingly, the collected data for this research were coded and analysed as explained in the following sections.

4.8.1 Coding and scoring the receptive tests

In complying with the ethical guidelines set by BERA, participants' receptive test papers were coded and scored manually. All participants' names were coded as P1, P2, P3...P281. The Software Package for Social Sciences SPSS 22 was used for coding and analysing the data. The main bases for scoring answers were accuracy of responses and frequency of their occurrences. Depending on a binary system coding, participants' responses were scored by giving 1 to the correct answers and 0 to the wrong or no answers. The total scores of each test were 64 as there were 64 items in each test. The scoring process was rechecked by me to ensure the accuracy of the scoring process. The rechecking process resulted in finding several cases in which some correct answers were regarded wrong and vice versa. After the checking phase, all data were transferred into an excel file then they were entered into an SPSS file.

4.8.2 Coding and scoring the productive tests

The collected data were scored on the bases of frequency and accuracy with the use of SPSS 22 software. Accordingly, as frequency data, a binary coding system was employed in which the correct answers were marked as 1 whereas the wrong or no answers were marked as 0. In terms of accuracy, participants' misspellings were not considered. However, if participants responded with appropriate alternatives that began with the same suggested priming letter of the intended item, they were considered as correct answers and as a shortcoming in the test. For example, one of the participants suggested *display skills* instead of the targeted answer *develop skills* which was though infrequent in use, but suited the context in the gap-filling test. The total scores of each test

were 64 as the test consisted of 64 items. The Constituent Likelihood Automatic Word-tagging System (CLAWS), software for tagging parts of speech, was utilised at both phases, collocation extraction and data analysis phases, to check the combination elements where it was difficult to decide the grammatical function of the word class of new items created by some participants. As was the case with the receptive tests, scoring the productive tests was rechecked by me to ensure the accuracy of the scoring process. The rechecking process came out with several cases in which some correct answers were regarded wrong and vice versa. After the checking phase, all data were transferred into an excel file then they were entered into an SPSS file.

4.9 Item analysis

The item analysis of the receptive and productive scales aimed at evaluating the tests themselves. It was run to show the test performance as a whole and whether there were any strange things about the test. It also explored if there were any items that were particularly easier or more difficult than others and how the facility values of the items were, and whether all facility values in general were high or low.

4.9.1 Item analysis-receptive test 1

In the first receptive test, as shown in Table (4), among 64 items of the test, the minimum facility value was on a prep+n collocation, on the item 55 (*on a picnic*, $fv = .25$, frequency= 5) which indicated a high difficulty to the participants whereas the maximum facility value was on phrasal verb collocations, item 35 (*sit down*, $fv = .96$, frequency= 29), item 36 (*talk about*, $fv = .96$, frequency= 162), item 40 (*look at*, $fv = .96$, frequency= 228) and a prep+n collocation, item 52 (*for example*, $fv = .96$, frequency= 33) which suggested the least difficulty to the participants. The general facility value of the test .72 indicated a rather easy level to participants. The facility value for each item can be seen in Appendix (G).

Table (4): Item analysis-receptive test 1 statistics

	Number of Items	Minimum facility value	Maximum facility value	General facility value	Std. Deviation
Receptive facility value of test 1	64	.25	.96	.73	.17

4.9.2 Item analysis-receptive test 2

The minimum facility value of the test items was .36, which indicated a rather high difficulty to participants on a phrasal verb collocation, item 45 (*bump into*, fv= .36, frequency=1). The maximum facility value was .94 on v+n collocations, items 7 (*play the guitar*, fv= .94, frequency= 27) and 11 (*cook dinner*, fv= .94, frequency= 6) respectively; an adj+n collocation, item 31 (*good news*, fv= .94, frequency=4); phrasal verb collocations, items 35 (*go away*, fv= .94, frequency=7), and item 43 (*go on*, fv= .94, frequency= 7), which suggested the least difficulty to participants. We will not go into a partial analysis of the influence of frequency on the difficulty or easiness of these items as a later part of my analysis will look in detail at factors affecting difficulty. The general facility value of the test .73 indicated a rather easy level to participants as shown in Table (5). The facility value for each item can be seen in Appendix (H).

Table (5): Receptive test 2 descriptive Statistics

	Number of Items	Minimum facility value	Maximum facility value	General facility value	Std. Deviation
Receptive 2 facility value	64	.36	.94	.73	.16

4.9.3 Item analysis-productive test 1

In the first productive test, as shown in Table (6), among 64 items of the test, the minimum facility value was very low .02 which indicated a very high difficulty level and bad performance of various types of collocations; the v+n collocation item 1 (*develop his skills*, fv= .02, frequency= 2), the adj+n collocations item 17 (*prehistoric people*, fv= .02, frequency= 1), and the phrasal verb collocation item 38 (*argue against*, fv= .02, frequency= 3). However, the greatest facility value was on the prep+n collocation item 64 (*of course*, fv= .96, frequency= 47), which performed very well in the test. The general facility value of the test was .52 which demonstrated a medium performance of the test with a standard deviation of .28 which was a little farther from the expected value (0) than the receptive test was. The facility value for each item can be seen in Appendix (I).

Table (6): Item analysis-productive test 1 statistics

	Number of items	Minimum facility value	Maximum facility value	General facility value	Std. Deviation
Facility value	64	.02	.96	.52	.28

4.9.4 Item analysis-productive test 2

The second productive test yielded several results as shown in Table (7) and Appendix (J). The most and the least challenging items appeared to be from the same syntactic pattern; adj+n collocations. The item that created the biggest challenge and performed badly in the test was item 29 (*alien spaceships*, $fv = .05$, frequency=1) whereas the least challenging that performed well in the test was item 31 (*happy birthday*, $fv = .98$, frequency= 10). As it was the case with the receptive test item analysis, I will not go into a partial analysis of the influence of frequency on the difficulty or easiness of these items at this place as a later part of my analysis will look in detail at factors affecting difficulty. The general facility value of the test .56 indicted that the difficulty of the items in common was medium.

Table (7): Item analysis-productive test 2 statistics

	Number of Items	Minimum facility value	Maximum facility value	General facility value	Std. Deviation
Productive 2 facility value	64	.05	.98	.56	.24

4.10 Reliability of the constructs

Reliability tests of the scales were conducted to be sure that the constructs produce stable, consistent results, i.e., to check the degree to which the different test items that explore the same construct yield similar results, and for the fear of computing subscales.

4.10.1 Reliability of the receptive test 1

Reliability of this scale was checked through testing the internal consistency of its items. The scale appeared to have a good internal consistency, the Cronbach's alpha coefficient of the test was .73. However, the item 8 (*fly a plane*, Corrected Item-Total Correlation (henceforth CITC) = $-.06$), item 9 (*bend his knees*, CITC= -0.001), item 55 (*on a picnic*, CITC= $-.03$), and item 62 (*across the country*, CITC= -0.008) appeared not to be well correlated with the total scale, and they seemed to have negative influences on the general scale. The greatest increase in alpha would come from removing item 8 (*fly a plane*) which increased reliability by .01. These items were clearly the worst performing ones, and they had small negative correlations with the overall score. This indicated that the weaker students tended to do better on them as shown in

Appendix (K). For improving the receptive scale reliability, it was decided to remove items 8, 9, 55, and 62 from the scale. Accordingly, the reliability of the scale increased to .75. However, the second reliability test revealed a negative correlation of item 16 (*make a mistake*, CITC= -.01) with the overall score as shown in Appendix (L).

Consequently, in order to further promote the reliability of the scale, item 16 was deleted. Accordingly, the reliability Cronbach's Alpha of the scale increased to .76, and no more negative correlations with the overall score were detected. Based on the above results, the test was taken reliable for achieving the study purposes after deleting five items, i.e., the scale preserved 59 out of 64 items.

4.10.2 Reliability of the receptive test 2

The reliability test suggested a good internal consistency of the scale; the Cronbach's alpha coefficient of the 64 items on the test was .76. Although the reliability of the scale was reasonable for the study purposes, the items 38 (*argue against*, CITC= -0.01), item 64 (*by accident*, CITC= -.004), and 64 (*of course*, CITC= -.19) did not seem to be well correlated with the scale, and they had negative influences on the total scale. These items were obviously the worst performing ones as shown in Appendix (M). Consequently, to improve the reliability of the scale, I decided to remove these two items which increased the alpha value to .78 with retaining 62 items. Accordingly, the scale was considered reliable for achieving aims of the study.

4.10.3 Reliability of the productive test 1

The scale appeared to have a strong internal consistency level, the Cronbach's alpha coefficient of the 64 items on the test was .93. All items seemed to be worthy of retention. The greatest increase in alpha would come from deleting item 35 (*go away*), but removal of this item would not result in an increase in the general reliability of the construct. All items correlated positively with the overall score to a good degree (lower $r = .07$), and no items had negative influences on the scale's overall score as shown in Appendix (N). Establishing on the results above, the productive test1 scale was found to be reliable for achieving the study aims.

4.10.4 Reliability of the productive test 2

The scale reliability was checked through testing the internal consistency of its items to make sure that all the items were measuring the same underlying

scale, participants' productive collocational knowledge as a coherent, unified construct. The scale appeared to have a very strong internal consistency level. The Cronbach's alpha coefficient of the 64 items on the test was .95. No items appeared to have negative influences on the total scale and they seemed to be worthy of retention as could be seen in Appendix (O). The items were very well correlated with the total scale of the second productive collocational knowledge. Based on the results above, the second productive test scale was found to be reliable for the purposes of the study.

However, for increasing the accuracy of the data analyses in the receptive-receptive, productive-productive, and receptive-productive comparisons, we decided to have equal number of items of the same characteristics in all the tests. Accordingly, I removed the items 8, 9, 16, 38, 55, 62, and 64 from all the scales. The rationale behind selecting these items to be deleted was that the items (8, 9, 55, and 62) performed badly in the first receptive test whereas the items (38 and 64) performed badly in the second receptive test. As a result, number of items became 57 in each test. Worth mentioning, the deletion of the latter items did not result in dramatically changing the reliability levels of the constructs (before deletion: receptive 1= .76, productive 1= .93; receptive 2= .78, productive 2= .95; after deletion: receptive 1= .75, productive 1= .93; receptive 2= .76, productive 2= .95). The change occurred only with the receptive scales; receptive 1 decreased from .76 to .75 while receptive 2 decreased from .78 to .76. The productive test scales retained their previous reliability levels of .95.

In conclusion, all the tests, receptively and productively at both times, first wave and second wave of data collection, were found to have performed well after applying various criteria to check their performance. First, validity of the constructs was sought; maintaining that the tests measured what they were supposed to measure, and that the test papers contents were established on the contents of the participants' syllabus (Alias, 2005). Secondly, item analysis was conducted as a quality indicator of poor items in the tests. Accordingly, too easy and too difficult items which affected negatively on the whole constructs were discarded. Thirdly, reliability of the constructs was explored; the Cronbach's alphas of the scales ranged between .76 and .95 which have been regarded as quite acceptable in such studies to range between .70 to .95

(Tavakol & Dennick, 2011). Fourthly, supporting measures were taken to enhance the validity and reliability of the constructs such as conducting pilot tests on native and non-native participants, ensuring clarity, relevancy, and achievability by all participants and administering the tests by the researcher himself. Fifth, the tests were answering the study questions by containing items with the features required for the study objectives, i.e., balancing items in terms of number of items and features of frequency, structures, transparency, and congruency with L1. In addition, the tests were affordable; they were not costly, practical, and applicable. Additionally, the design suited the general setting of time and place of the tests. Finally, the accuracy was sought as much as possible in all aspects of data collection to increase the validity and reliability of the tests and the yielded data.

Finally, the size of effect of the significant t-tests results was also calculated in the current study. The effect size test is used to measure the strength or significance of an effect of a feature or an intervention (Becker, 2000) through quantifying the size of the difference between two groups, and it conveys the same information as a test of statistical significance, but with the emphasis on the significance of the effect, rather than the sample size (Coe, 2002). Coe (2002) argues that it can be of benefit for quantifying the effectiveness of a particular intervention or feature, relative to some comparison by focusing on the most important aspect of an intervention- the size of the effect- rather than its statistical significance, which conflates effect size and sample size.

Accordingly, the effect size test can promote a more scientific approach to the accumulation of knowledge. For these reasons, effect size is an important tool in reporting and interpreting effectiveness. I employed Cohen's d (1988) for t-test (independent samples) to assess the magnitude of the effect size.

According to Cohen's d for the independent samples t-test, Cohen's d is determined by calculating the mean difference between two groups, and then dividing the result by the pooled standard deviation. The effect size was calculated using the following formula:

$$\text{Cohen's } d = (M_2 - M_1) / SD_{\text{pooled}}$$

Where

$$SD_{\text{pooled}} = \sqrt{((SD_1^2 + SD_2^2) / 2)} \text{ (Cohen, 1988, p. 44)}$$

Cohen (1988, p. 25) suggested that $d=0.2$ be considered a 'small' effect size,

0.5 represents a 'medium' effect size and 0.8 a 'large' effect size. This means that if two groups' means do not differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant. Accordingly, the effect size was calculated only when the effect was statistically significant because there would be no sense in calculating the size of an effect, if there was no good reason to suppose there was any effect.

4.11 Waves of data collection

Data collection was on two waves over the academic year 2015-2016. I planned to leave about six months as interval between a wave and another to allow time for a recognisable development. The rationale behind this was the belief that learners' collocational knowledge is usually slow and does not develop very fast at this stage of learning. The first wave of gathering data commenced in October 2015, and it was regarded as the baseline data. The first wave of data collection was of outstanding significance to the study and the participants as the impression it left on the test-takers partly determined the destiny of the next wave of data collection. Accordingly, great attention was paid to this wave of data collection. Great care was taken in terms of time length of the test, timing, place, test format, and nature of the items included in the tests. Shortcoming in any of these aspects could be affecting the data collection process, accordingly, the study findings. The second wave of data collection was in May 2016, and it was the closing wave of data collection. The importance of the second wave of data collection lied in that it carried the final and the highest point of development and the second pole of data for making comparisons that would be used for inferring L2 learners' collocational knowledge and development across the school year.

4.12 Ethics (UK and Kurdistan)

The many ethical lapses in conducting research in various research institutions in the past decades have emerged an increasing focus on researchers' ethical responsibilities (Couch & Dodd, 2005). Accordingly, the ethical guidelines within both contexts, the United Kingdom and Kurdistan region, were highly considered as one of the major prerequisites in conducting this study. In the former context, i.e., UK, I had to obtain the ethical approval from the college ethics committee of the University of Exeter to conduct the practical part of the current study, data collection, in Kurdistan (see Appendix A). As for the latter

context, i.e., Kurdistan Region, my main responsibilities, in compliance with BERA's ethical guidelines (BERA, 2011, p. 5-10), comprised "responsibility towards, participants, research funding body, educational researchers' community, educational professionals, policy decision-makers, and the general public."

In the current study, the ethical concerns pivoted mainly on researcher's responsibility towards participants. My responsibility, as a researcher, towards participants who were mostly under 18 years old was multifaceted. Prior to commencing the study, I needed to ensure voluntary informed consent from the targeted sample. They were informed about the aims, benefits, consequences, and potential disadvantages from their participations such as the time they would spend on the tests and the effort they would exert to do the tests (see Appendix A).

Additionally, appropriate consent was sought from the local authorities in Kurdistan for conducting the pilot and data collection tests on students. A further sensitive issue was teachers' and school administrations' concerns about the confidentiality of data about their schools' levels in general and English in particular to which a special attention was paid. It was assured to them that all the data are kept highly confidential and used for research purposes only.

Openness and disclosure was another aspect of this responsibility. The recruited participants were not deceived or tricked to take part in the study. However, some procedures were adopted, with the local authority approval, to motivate their participation and reduce the attrition rate such as explaining the significance of the taken tests for the school evaluation and spotting strengths and weaknesses of the curriculum.

A further responsibility aspect towards participants was their right to withdraw from the study. Individual withdrawals from the study at any time for any reason was guaranteed without any consequences on participants. However, exploring the reasons behind their dropping and any persuasion possibility of re-engaging them into the study was sought without any pressure on their decision-making.

Incentives were offered such as useful websites of learning English, and the negative consequences of any incentive was considered in advance to avoid their effects on participants, accordingly, on the study findings. Moreover, any

expected detriments or harms that might occur as a result of their participation before or during the study were brought to their attention.

Finally, participants' privacy and their data confidentiality were of great significance in the study. Students commonly seemed to be very sensitive towards disclosing their data to others. Accordingly, the anonymity and confidentiality of their data were assured and made clear to them prior to have the research got underway. Otherwise, if some participants wished to have their data disclosed after scoring their test papers, this was also guaranteed.

Chapter five

Data analyses and results

Data analyses were conducted on two waves of data. Accordingly, both analyses were carried out separately. Analysing the first wave of data was conducted immediately after its collection at the beginning of the school year under study. Analysing the second wave of data was carried out at the end of the school year after collecting data from participants. In the same order, analyses of the first wave of data will be presented before introducing the analyses of the second wave of data in this chapter. The analyses of each data wave explore participants' characteristics, and the relationship between knowledge of collocations and each of gender of participants; frequency of collocations in students' school textbooks; syntactic structures of collocations; semantic transparency, and congruency with L1 successively. Additionally, participants' collocational knowledge development is examined. The results of the following data analyses provide answers to the study questions mentioned earlier in the methodology chapter (Section 4.1).

5.1 First wave data analyses

The first wave of data collection was conducted successfully after dealing with the shortcomings that were identified in the pilot tests. The analyses of this part of the tests dealt with examining participants' demographics, gender influence on the participants' receptive and productive collocational knowledge, and exploring the relationships between participants' collocational knowledge and the four features of collocations- frequency, syntactic structure, semantic transparency, and congruency with L1.

5.1.1 Participants' characteristics

The 252 participants were of both sexes; males were 87 who comprised about 35% of the participants whereas females were 165 who made up about 65% of the total participants.

5.1.2 Exploring relationships

The relationships were explored between various aspects of the participants' collocational receptive and productive knowledge. The probed aspects included the relationship between knowledge of collocations and

- gender of participants;

- frequency of collocations in students' school textbooks;
- syntactic structures of collocations;
- semantic transparency;
- L1 congruency.

The analyses are conducted in the order mentioned above as explained in the following sections.

5.1.2.1 Receptive test 1 analyses

5.1.2.1.1 Gender

The impact of gender on the participants' receptive collocational knowledge was carried out by a two-tailed independent-samples t-test. As Table (8) shows, males (M= 43.57) on average scored slightly higher than females (M= 41.53). Equal variances were not assumed as the Levene's test was significant. An independent-samples t-test revealed a statistically significant difference between the mean scores of both groups, males and females, $t(250)= 2.42$, $p < .05$, $d = .30$, but the effect size was small. According to Cohen (1988), effect size from .2 to .5 is small. In other words, although the difference between the means of both groups was statistically significant, the effect of the gender factor was small in its influence on the collocational knowledge of males and females.

Table (8): Receptive test 1- gender group statistics

	Participants' gender	Number of participants	Mean	Std. Deviation	Std. Error Mean
Receptive total scores 1	Male	87	43.57	6.85	.73
	Female	165	41.53	5.38	.42

According to the results above, males did better and achieved higher scores on collocations than females in the receptive test 1.

5.1.2.1.2 Frequency

The influence of frequency of collocations on the participants' collocational receptive knowledge was examined in this test. As it was explained in the methodology chapter (Section 4.6.4), collocations, according to their frequency, were not divided into high-frequency and low-frequency items, rather, items were ranked according to their frequencies in the textbooks. Accordingly, the comparison was made between the actual number of occurrences of the items in the participants' textbooks, log frequency, and the facility value of the items. The relationship between frequency of collocations and difficulty of the items

was examined by conducting a Pearson correlation analysis. The results indicated a statistically significant, positive, moderate relationship ($r = .47$, $N = 57$, $p < .05$) between log frequency and facility value of the items. In other words, higher levels of facility values were generally associated with higher levels of frequency; frequent items were less challenging than less frequent ones. Accordingly, participants' receptive knowledge of the collocations with higher frequencies was higher than their knowledge of the items with lower frequencies.

5.1.2.1.3 Syntactic structure

This aspect of the test examined the influence of the syntactic structure of collocations on the participants' collocational receptive knowledge. As it was explained in the methodology chapter (Section 4.6.4), collocations were divided according to their syntactic structure into lexical collocations, which consist of v+n and adj+n collocations, and grammatical collocations, which consist of phrasal verb and prep+n collocations. To this end, a one-way ANOVA was conducted to examine whether there were statistically significant differences among the facility values of the different syntactic patterns of collocations. Equal variances were assumed as the Levene's test was not significant.

The results revealed statistically significant differences among the syntactic groups, $F(3, 53) = 3.20$, $p < .05$. Post-hoc Games-Howell tests revealed statistically significant differences between the facility values of adj+n and phrasal verb items and between phrasal verb and prep+n items. The results indicated a significant influence of the syntactic structure of collocations on the facility values of each of these structures as shown in Table (9) and Table (10).

Table (9): Receptive test 1- syntactic structure statistics

	Number of items	Facility value mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum facility value	Maximum facility value
					Lower Bound	Upper Bound		
v+n	13	0.74	0.15	0.04	0.64	0.83	.44	.91
adj+n	16	0.70	0.15	0.04	0.62	0.79	.46	.94
phrasal verb	15	0.84	0.11	0.03	0.77	0.90	.59	.96
prep+n	13	0.68	0.17	0.05	0.58	0.78	.41	.96

Total	57	0.74	0.16	0.02	0.70	0.78	.41	.96
-------	----	------	------	------	------	------	-----	-----

Table (10): Receptive test 1- syntactic structure post-hoc Games-Howell

(I) part of speech	(J) part of speech	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
v+n	adj+n	0.03	0.06	0.95	-0.13	0.19
	phrasal verb	-0.10	0.05	0.23	-0.25	0.04
	prep+n	0.05	0.06	0.82	-0.12	0.23
adj+n	v+n	-0.03	0.06	0.95	-0.19	0.13
	phrasal verb	-.13358*	0.05	0.048	-0.27	0.00
	prep+n	0.02	0.06	0.98	-0.14	0.19
phrasal verb	v+n	0.10	0.05	0.23	-0.04	0.25
	adj+n	.13358*	0.05	0.048	0.00	0.27
	prep+n	.15656*	0.05	0.044	0.00	0.31
prep+n	v+n	-0.05	0.06	0.82	-0.23	0.12
	adj+n	-0.02	0.06	0.98	-0.19	0.14
	phrasal verb	-.15656*	0.05	0.044	-0.31	0.00

Accordingly, phrasal verb collocations appeared to be less challenging than adj+n and prep+n but not less challenging than v+n collocations. In other words, the syntactic structure of collocations seemed to have significantly affected the difficulty of the items, accordingly, participants' receptive knowledge of collocations.

5.1.2.1.4 Semantic transparency

The present test examined the impact of the semantic transparency of the collocational items on the participants' receptive knowledge of collocations. As a reminder, as it was explained in the methodology chapter (Section 4.6.4), collocations were divided according to their semantic transparency into transparent and opaque collocations. An independent-samples t-test was conducted to investigate the influence of the semantic transparency of collocations on the difficulty of the items. As the Levene's test was not statistically significant, equal variances were assumed. An independent-samples t-test did not reveal a statistically significant difference between the facility value on transparent and opaque collocations, $t(55) = -.58, p > .05$. In other words, the semantic transparency of collocations did not seem to have a

significant influence on the difficulty of the items, accordingly, participants' receptive knowledge of collocations.

5.1.2.1.5 L1 congruency

The influence of congruency with L1 on the participants' receptive knowledge of collocations was examined in this test. To refresh memory, as it was explained in the methodology chapter (Section 4.6.4), collocations were divided according to congruency with L1 into congruent and incongruent collocations. An independent-samples t-test was conducted to investigate the influence of congruency with L1 on the difficulty of the items. As the Levene's test was not statistically significant, equal variances were assumed. The test results showed that this difference was not statistically significant, $t(55) = .63, p > .05$. Accordingly, no significant impact of congruency with L1 was found on the difficulty of the items, accordingly, participants' receptive collocational knowledge in this test.

5.1.2.2 Productive test 1 analyses

5.1.2.2.1 Gender

An independent-samples t-test was also conducted to compare the productive collocational knowledge for males and females. As Table (11) shows, males ($M = 31.72$) on average scored slightly higher than females ($M = 27.04$). Equal variances were not assumed as the Levene's test was significant. An independent-samples t-test showed this difference to be statistically significant, $t(250) = 3.16, p < .05, d = .39$, but the effect size was rather small. Based on Cohen's (1988) suggestions, effect size of .39 is considered small. In other words, although the difference between the means of both groups was statistically significant, the effect of the gender factor was small in its influence on the collocational knowledge of males and females.

Table (11): Productive test 1- gender group statistics

	participant's gender	N	Mean	Std. Deviation	Std. Error Mean
Productive total scores 1	Male	87	31.72	12.10	1.30
	Female	165	27.04	9.28	.72

Accordingly, male participants seemed to be productively more competent than females in producing their textbook English collocations in the productive test 1.

5.1.2.2.2 Frequency

In order to examine the relationship between the frequency of collocations and participants' productive collocational knowledge, a Pearson correlation analysis was carried out. The results revealed a statistically significant, positive, moderate relationship ($r = .49$, $N = 57$, $p < .05$) between log frequency and productive facility value; higher levels of productive collocational knowledge were associated with higher levels of frequency. In other words, participants did better with collocations of higher frequency whereas they were more challenged with items of less frequency in the productive test 1.

5.1.2.2.3 Syntactic structure

The influence of the syntactic structure of collocations on the participants' productive collocational knowledge was examined by running a one-way ANOVA test. Equal variances were not assumed as the Levene's test was significant. The ANOVA results revealed no statistically significant differences among the syntactic groups, $F(3, 53) = 2.13$, $p > .05$ (see Table 12 and Table 13).

Table (12): Productive test 1- syntactic structure of collocations

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.46	3	.15	2.13	.11
Within Groups	3.79	53	.07		
Total	4.25	56			

Table (13): Productive test 1- syntactic structure of collocations descriptives

	Number of items	Facility value Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
v+n	13	.59	.28	.08	.42	.76	.02	.92
adj+n	16	.44	.32	.08	.26	.61	.02	.89
phrasal verb	15	.41	.24	.06	.27	.54	.07	.83
prep+n	13	.61	.20	.06	.49	.73	.13	.86
Total	57	.50	.28	.04	.43	.58	.02	.92

In other words, the syntactic structures of collocations were found to have no significant impact on the participants' productive knowledge of the tested collocations in the productive test 1.

5.1.2.2.4 Semantic transparency

The influence of the semantic transparency of collocations on the participants' productive collocational knowledge was examined by conducting an independent-samples t-test. As the Levene's test was not statistically significant, the assumption of homogeneity of variances was not violated and equal variances were assumed. An independent-samples t-test did not reveal a statistically significant difference between facility value mean of the transparent collocations and their facility value of the opaque ones, $t(55) = -.49, p > .05$. In conclusion, the semantic transparency of collocations did not seem to have a notable impact on the participants' productive knowledge in the productive test 1.

5.1.2.2.5 L1 congruency

An independent-samples t-test was carried out to examine the impact of congruency with L1 on the participants' productive collocational knowledge. As the Levene's test was not statistically significant, equal variances were assumed. An independent-samples t-test showed no statistically significant difference between participants' facility value of the congruent collocations and their facility value of the incongruent ones, $t(55) = .01, p > .05$. In sum, the results indicated no significant influence of congruency of the collocations with participants' L1 on the productive collocational knowledge in the productive test 1.

5.2 Second wave data analyses

The second wave of data collection was conducted in the last week of April 2016. In general, the procedures followed in running and analysing the second wave of the tests were the same followed in conducting the first one. In the following sections, the test results and their analyses will be presented.

5.2.1 Receptive test 2 analyses

5.2.1.1 Gender

The influence of gender on the participants' receptive collocational knowledge was carried out by a two-tailed independent-samples t-test. As Table (14) shows, males ($M = 42.86$) on average scored slightly higher than females ($M = 40.70$). The Sig. value of Levene's test for equality of variances indicated that the assumption of equal variances was violated i.e., equal variances were not assumed. An independent-samples t-test indicated a statistically significant difference between the mean scores of both groups, males and females,

$t(250) = 2.49, p < .05, d = .31$, but the effect size was small. In other words, although the difference between the means of both groups was statistically significant, the effect of the gender factor was small in its influence on the collocational knowledge of males and females.

Table (14): Receptive test 2- gender group statistics

	participant's gender	N	Mean	Std. Deviation	Std. Error Mean
Receptive total scores 2	Male	87	42.86	6.92	.74
	Female	165	40.70	5.75	.45

This indicated a moderate effect of gender role in explaining the differences between the mean scores of both groups, and that the differences were not occurred by chance. Establishing on the results above, males were found to achieve higher scores on collocations than females in the receptive test 2.

5.2.1.2 Frequency

A Pearson correlation analysis was conducted to explore the relationship between collocations frequency and participants' receptive collocational knowledge. The results indicated a statistically significant, positive moderate relationship between log frequency and the receptive facility value of the second receptive test, $r = .41, N = 57, p < .05$; higher levels of log frequency were moderately associated with higher levels of receptive facility values. The results implied a positive relationship between the frequency of occurrence of collocations in the textbooks and participants' receptive knowledge of the items; the more frequent collocations were found to be easier to recognise than the less frequent ones in the receptive test 2. In conclusion, participants' knowledge of the higher frequency items was more than their knowledge of the lower frequency items in this test.

5.2.1.3 Syntactic structure

The influence of the syntactic structure of collocations on the participants' collocational receptive knowledge was explored. A one-way ANOVA analysis was conducted to examine whether there were statistically significant differences among collocations in different syntactic structures groups in relation to their receptive knowledge. The statistic descriptives of the test were as shown in Table (15).

Table (15): Receptive test 2- syntactic structures statistics

	Number of items	Facility value mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum facility value	Maximum facility value
					Lower Bound	Upper Bound		
v+n	13	0.72	0.19	0.05	0.61	0.83	.37	.94
adj+n	16	0.71	0.16	0.04	0.62	0.79	.38	.94
phrasal verb	15	0.77	0.17	0.04	0.68	0.86	.36	.94
prep+n	13	0.71	0.14	0.04	0.62	0.79	.42	.92
Total	57	0.73	0.16	0.02	0.68	0.77	.36	.94

The Levene's test Sig. value indicated homogeneity of variances, accordingly, equality of variances was assumed. The ANOVA test results did not reveal statistically significant differences among the four syntactic groups of collocations, $F(3, 53) = .55, p > .05$, as shown in Table (16).

Table (16): Receptive test 2- syntactic structures ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.04	3	.02	.55	.65
Within Groups	1.43	53	.03		
Total	1.47	56			

In conclusion, the grammatical structure of collocations did not appear to have a statistically significant influence on the participants' receptive collocational knowledge in the receptive test 2.

5.2.1.4 Semantic transparency

Investigating the impact of the semantic transparency of collocations on participants' receptive collocational knowledge was carried out by running an independent-samples t-test. Equal variances were assumed since the Levene's test was not statistically significant. However, an independent-samples t-test did not reveal a statistically significant difference between participants' facility value mean on the transparent collocations and their facility value average on the opaque ones, $t(55) = -1.51, p > .05$. In conclusion, the results indicated no magnitude impact of the semantic transparency of the collocational items on the participants' receptive knowledge in their second session of the receptive test.

5.2.1.5 L1 congruency

Examining the influence of congruency of English collocations with participants' L1, Kurdish, collocations on the participants' receptive collocational knowledge

was carried out by conducting an independent-samples t-test. The Levene's test Sig. value indicated that the test was not significant, accordingly, equal variances were assumed. The independent-samples t-test results revealed no statistically significant difference between participants' receptive facility value average of the congruent collocations and that of the incongruent ones, $t(55) = .81, p > .05$. Accordingly, the test results demonstrated no significant influence of the collocations congruency with L1 on the participants' receptive collocational knowledge in the second receptive test.

5.2.2 Productive test 2 analyses

5.2.2.1 Gender

An independent samples t-test was conducted to compare the productive collocational knowledge for males and females. As Table (17) shows, males mean score was ($M = 34.43$) whereas that of females was ($M = 31.45$). The Sig. value of Levene's test for equality of variances indicated that the assumption of equal variances was not violated, i.e., equal variances were assumed. An independent-samples t-test showed that the difference between both groups was not statistically significant, $t(250) = 1.80, p > .05$.

Table (17): Productive test 2- gender group statistics

	Participant's gender	N	Mean	Std. Deviation	Std. Error Mean
Productive total scores 2	Male	87	34.43	13.73	1.47
	Female	165	31.45	11.71	.91

Accordingly, the results indicated that males were not significantly more competent than females in producing their textbook English collocations in the productive test 2.

Table (18): Productive test 2- syntactic structure of collocations descriptives

	N	Facility value mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
v+n	13	0.51	0.26	0.07	0.35	0.67	0.12	0.94
adj+n	16	0.45	0.27	0.07	0.31	0.60	0.05	0.98
phrasal verb	15	0.67	0.18	0.05	0.57	0.76	0.21	0.87
prep+n	13	0.66	0.15	0.04	0.57	0.76	0.36	0.83
Total	57	0.57	0.24	0.03	0.51	0.63	0.05	0.98

5.2.2.2 Frequency

The relationship between collocations frequency and participants' productive collocational knowledge was examined by running a Pearson correlation analysis. The results revealed a statistically significant, positive, moderate relationship between log frequency and facility value of the productive test items ($r = .54$, $N = 57$, $p < .05$); higher facility value means were associated with collocations of higher log frequency values. In other words, participants did better on collocations with higher frequencies than collocations with lower frequencies in their curricula textbooks in the productive test 2.

5.2.2.3 Syntactic structure

The collocation syntactic structure influence on the participants' collocational productive knowledge was examined by running a one-way ANOVA test. It was conducted to examine whether there were statistically significant differences among collocations in different syntactic structure groups. The statistic descriptives of the four syntactic groups were as shown in Table (18). Equal variances were not assumed as the Levene's test was significant, accordingly, the homogeneity of variances assumption was violated. A one-way ANOVA test results indicated that the overall differences among the syntactic groups were statistically significant, $F(3, 53) = 3.44$, $p < .05$, Table (19).

Table (19): Productive test 2- syntactic structure of collocations ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.51	3	.17	3.44	.02
Within Groups	2.64	53	.05		
Total	3.16	56			

However, a post hoc test did not reveal statistically significant differences among the individual syntactic groups. In total, the syntactic structure seemed to have influenced the participants' productive collocational knowledge. However, at the individual level of the collocations patterns, the test did not reveal a significant difference among the individual patterns.

5.2.2.4 Semantic transparency

An independent-samples t-test was carried out to examine the semantic transparency of collocations impact on participants' productive collocational knowledge. As the Levene's test was not statistically significant, equal variances were assumed. An independent-samples t-test did not reveal a statistically significant difference between participants' facility value on the

transparent collocations and their facility value on the opaque ones, $t(55) = .25$, $p > .05$). In sum, the semantic transparency of the tested collocations did not show a statistically significant influence on the participants' productive knowledge in the productive test 2.

5.2.2.5 L1 congruency

An independent- samples t-test was carried out to examine the congruency with L1 impact on the participants' productive collocational knowledge. As the Levene's test was not statistically significant, equal variances were assumed. The difference between participants' facility value on the congruent collocations and their facility value on the incongruent ones appeared to be statistically not significant, $t(55) = -.29$, $p > .05$). In sum, congruency of English collocations with the participants' L1 did not seem to have a significant impact on the participants' productive knowledge in the second session of the productive test.

5.3 Summary

To give an overall view of the influence of the four collocational features under focus, the conducted analyses and their results and to facilitate making comparisons between the results of both waves of data collection, a summary of the test results are presented below as show in Table (20).

Table (20): Collocational knowledge analyses and results summary

Features	Time 1 (first wave data)		Time 2 (second wave data)	
	Receptive	Productive	Receptive	Productive
Gender	Significant, Male>Female $t(250) = 2.42$, $p < .05$, $d = .30$)	Significant, Male>Female, $t(250) = 3.16$, $p < .05$, $d = .39$)	Significant, Male>Female, $t(250) = 2.49$, $p < .05$, $d = .31$)	Not significant, $t(250) = 1.80$, $p > .05$
Frequency	Significant, facility value increases with frequency, $(r = .47$, $N = 57$, $p < .05$)	Significant, facility value increases with frequency, $(r = .49$, $N = 57$, $p < .05$)	Significant, facility value increases with frequency, $(r = .41$, $N = 57$, $p < .05$)	Significant, facility value increases with frequency, $(r = .54$, $N = 57$, $p < .05$)

Syntactic structure	Significant, facility value varies according to the syntactic structure, $F(3, 53) = 3.20, p < .05$	Not significant, $F(3, 53) = 3.60, p > .05$	Not significant, $F(3, 53) = .55, p > .05$	Significant, facility value varies according to the syntactic structure, $F(3, 53) = 3.44, p < .05$
Semantic transparency	Not significant, $t(55) = -.58, p > .05$	Not significant, $t(55) = -.49, p > .05$	Not significant, $t(55) = -1.51, p > .05$	Not significant, $t(55) = .25, p > .05$
L1 congruency	Not significant, $t(55) = .63, p > .05$	Not significant, $t(55) = .01, p > .05$	Not significant, $t(55) = .81, p > .05$	Not significant, $t(55) = -.29, p > .05$

5.4 Collocational knowledge development

One of the main aims of the current study was exploring Kurdish students' collocational knowledge development throughout one school year. Commonly, the results indicated that participants of the current study had a command of 63.5% of the total test items (i.e., 63.5% correct answers). However, this knowledge varied according to the receptive and productive dimensions. Receptively, the results indicated that participants had a mastery of 73.4% of the receptive test items. However, productively, the results indicated participants' knowledge of 53.6% of the productive test items (i.e., 53.6% correct answers). Accordingly, participants' receptive collocational knowledge exceeded their productive knowledge of collocations, and the difference between participants' collocational knowledge of these two dimensions was statistically significant, $t(251) = 6.17, p < .05, d = .53$, and the effect size was medium (see Table 22).

Generally, the results of the data analyses of the participants' collocational knowledge development indicated three different cases of change in knowledge: statistically significant increase in the participants' collocational knowledge, statistically significant decrease in the participants' collocational knowledge, and no statistically significant results (see Table 21).

The aspects of participants' collocational knowledge which showed statistically significant knowledge developments through the school year were total collocational knowledge (+2.7%), productive collocational knowledge (+6.7%), male productive collocational knowledge (+4.7%), female productive collocational knowledge (+7.8%), productive higher frequency collocational knowledge (+9%), productive lower frequency collocational knowledge (+20.9%), productive grammatical collocational knowledge (+16.3%), productive transparent collocational knowledge (+9.3%), productive opaque collocational knowledge (+4.1%), productive congruent collocational knowledge (+5.9%), and productive incongruent collocational knowledge (+7.7%).

In contrast, the aspects which revealed significant decreases in the participants' collocational knowledge included receptive higher frequency collocational knowledge (-3.5%), receptive grammatical collocational knowledge (-2.1%), receptive transparent collocational knowledge (-3.4%), receptive incongruent collocational knowledge (-1.7%), and productive lexical collocational knowledge (-2.5%).

However, some other aspects of the participants' collocational knowledge revealed no significant changes. These aspects included receptive collocational knowledge, male receptive knowledge, female receptive knowledge, receptive lower frequency collocational knowledge, receptive lexical collocational knowledge, receptive opaque collocational knowledge, and receptive congruent collocational knowledge.

Worth mentioning, the aspects which revealed declining or no statistically significant change in the participants' collocational knowledge were mainly in the receptive dimension whereas the aspects which developed over the school year were generally in their productive knowledge.

To provide an overall view of the statistical results and facilitate making comparisons between the different aspects of participants' collocational

knowledge, a summary of the results discussed above are presented in Table (21).

Table (21): Collocational knowledge development results summary

Variables knowledge development	Time 1	Time 2	Gain= (T2-T1)	Significance
Total collocational knowledge	M= 70.89, SE= .91	M= 73.93, SE= 1.06	M= 3.04, SE= .79	Significant, $t(251)= 3.83$, $p < .05$, $d= .19$
Receptive collocational knowledge	M= 42.24, SE= .38	M= 41.45, SE= .39	M= -.79, SE= .41	Not significant, $t(251)= -1.94$, $p > .054$
Productive collocational knowledge	M= 28.65, SE= .67	M= 32.48, SE= .79	M= 3.83, SE= .66	Significant, $t(251)= 5.84$, $p < .05$, $d= .33$
Male receptive knowledge	M= 43.57, SE= .73	M= 42.86, SE= .74	M= -.71, SE= .70	Not significant, $t(86)= -1.02$, $p > .05$
Female receptive knowledge	M= 41.53, SE= .42	M= 40.70, SE= .45	M= -.83, SE= .50	Not significant, $t(164)= -1.65$, $p > .05$
Receptive higher frequency collocational knowledge	M= 23.67, SE= .20	M= 22.87, SE= .20	M= -.80, SE= .25	Significant, $t(251)= -3.25$, $p < .05$, $d= .25$
Receptive lower frequency collocational knowledge	M= 18.57, SE= .22	M= 18.58, SE= .27	M= .01, SE= .27	Not significant, $t(251)= .05$, $p > .05$
Receptive lexical collocational knowledge	M= 20.84, SE= .21	M= 20.65, SE= .21	M= -.19, SE= .24	Not significant, $t(251)= -.81$, $p > .05$
Receptive grammatical	M= 21.40, SE= .22	M= 20.80, SE= .22	M= -.60, SE= .24	Significant, $t(251)= -2.46$, $p < .05$, $d= .17$

collocational knowledge				
Receptive transparent collocational knowledge	M= 21.15, SE= .20	M= 20.17, SE= .23	M= -.98, SE= .24	Significant, $t(251) = -4.04$, $p < .05$, $d = .29$
Receptive opaque collocational knowledge	M= 21.09, SE= .23	M= 21.28, SE= .21	M= .19, SE= .25	Not significant, $t(251) = .76$, $p > .05$
Receptive congruent collocational knowledge	M= 23.36, SE= .23	M= 23.02, SE= .22	M= -.34, SE= .26	Not significant, $t(251) = -1.31$, $p > .05$
Receptive incongruent collocational knowledge	M= 18.88, SE= .19	M= 18.43, SE= .22	M= -.45, SE= .29	Significant, $t(251) = -1.99$, $p < .05$, $d = .14$
Male productive knowledge	M= 31.72, SE= 1.3	M= 34.43, SE= 1.5	M= 2.70, SE= 1.08	Significant, $t(86) = 2.50$, $p < .05$, $d = .21$
Female productive knowledge	M= 27.04, SE= .72	M= 31.45, SE= .91	M= 4.42, SE= .82	Significant, $t(164) = 5.38$, $p < .05$, $d = .42$
Productive higher frequency collocational knowledge	M=18.01, SE= .31	M= 19.62, SE= .41	M= 1.61, SE= .38	Significant, $t(251) = 4.63$, $p < .05$, $d = .28$
Productive lower frequency collocational knowledge	M= 10.65, SE= .39	M= 12.87, SE= .41	M= 2.22, SE= .36	Significant, $t(251) = 6.10$, $p < .05$, $d = .35$
Productive lexical collocational knowledge	M= 14.63, SE= .35	M= 13.90, SE= .39	M= -.73, SE= .31	Significant, $t(251) = -2.35$, $p < .05$, $d = .13$
Productive grammatical knowledge	M= 14.02, SE= .37	M= 18.58, SE= .47	M= 4.56, SE= .47	Significant, $t(251) = 9.81$, $p < .05$, $d = .68$

collocational knowledge				
Productive transparent collocational knowledge	M= 14.06, SE= .33	M= 16.75, SE= .41	M= 2.69, SE= .34	Significant, $t(251)= 7.92$, $p < .05$, $d= .46$
Productive opaque collocational knowledge	M=14.60, SE= .36	M= 15.73, SE= .40	M= 1.14, SE= .37	Significant, $t(251)= 3.10$, $p < .05$, $d= .19$
Productive congruent collocational knowledge	M=15.61, SE= .37	M= 17.43, SE= .46	M= 1.82, SE= .38	Significant, $t(251)= 4.79$, $p < .05$, $d= .28$
Productive incongruent collocational knowledge	M=13.05, SE= .32	M= 15.05, SE= .35	M= 2.00, SE= .32	Significant, $t(251)= 6.30$, $p < .05$, $d= .38$

However, the next part of the analysis looked at how improvements in collocational knowledge were related to key variables. Specifically, it compared the gain scores achieved by males vs females, on productive vs receptive versions of the test, and for the different collocation types: higher frequency vs lower frequency, lexical vs grammatical, transparent vs opaque, and congruent vs incongruent.

From a gender perspective, the test results of the individual tests, i.e., one-shot data tests, commonly indicated a statistically significant influence of gender on the participants' collocational knowledge (see Table 20). However, the analytical results of contrasting the knowledge development of males with females indicated no statistically significant differences between male and female participants' collocational knowledge development over the school year neither receptively, $t(250) = .14$, $p > .05$, nor productively, $t(250) = -1.25$, $p > .05$ (see Table 22). Since the influence of gender on the participants' collocational knowledge was statistically significant in most of the individual tests, the non-significant results of the difference between the amount of knowledge development of males and females may imply that one school year is not sufficient to demonstrate this influence on their knowledge development, or a

change in learners' attitude towards learning occurred during this period such as females being more motivated to learn than males.

From a frequency view, the results of analysing participants' knowledge development of higher frequency and lower frequency collocations indicated that participants' knowledge of the higher frequency collocations increased significantly more than their knowledge of the lower frequency items (see Table 22). Receptively, development in the participants' knowledge of the higher frequency collocations exceeded significantly their knowledge of the lower frequency items, $t(251) = 28.72$, $p < .05$, $d = .63$, and the effect size was medium, according to Cohen (1988), effect size from .5 to .8 is considered medium. In other words, the influence of the frequency feature on the participants' collocational knowledge development was of medium strength.

Similarly, development in the participants' productive knowledge of the higher frequency collocations exceeded significantly their knowledge of the lower frequency items, $t(251) = 38.83$, $p < .05$, $d = .55$, and the effect size was medium. Interestingly, the amount of knowledge development of the lower frequency items was more than that of the higher frequency items over the school year receptively, $t(251) = -2.62$, $p < .05$, $d = .20$, and productively, $t(251) = -2.20$, $p < .05$, $d = .12$, but the effect size was small for the former and trivial for the latter, according to Cohen (1988), if two groups' means do not differ by 0.2 standard deviations or more, the difference is trivial, even if it is statistically significant. This might indicate that participants had paid more attention to lower frequency items at this level of language learning.

Syntactically, the test results of the individual tests, i.e., one-shot data tests, generally indicated a statistically significant influence, though the influence was not consistent across all the tests, of the syntactic structure of collocations on the participants' collocational knowledge (see Table 20). Developmentally, the results generally indicated that the amount of knowledge development of the grammatical collocations exceeded significantly the amount of participants' knowledge development of the lexical collocations. Receptively, the results indicated no statistically significant difference between participants' knowledge development of the grammatical and lexical collocations. However, productively, the results indicated that participants' knowledge development of the grammatical collocations exceeded significantly their knowledge development of

the lexical collocations and the difference was large (mean difference $M= 5.29$) (see Table 22).

From a semantic transparency perspective, the test results of the individual tests, i.e., one-shot data tests, indicated no statistically significant influence of the semantic transparency on the participants' collocational knowledge (see Table 20). However, the influence of the semantic transparency on the participants' collocational knowledge was clearer through analysing participants' collocational knowledge development over the school year. This may indicate the appropriateness of longitudinal designs in spotting the slow changes in learners' collocational knowledge over time which cannot be identified by one-shot data study designs. The developmental analyses of the results (see Table 22), revealed that the amount of development in the receptive knowledge of opaque collocations was significantly more than that in the receptive knowledge development of the transparent collocations (mean difference $M= 1.17$). In contrast, productively, the results revealed that the amount of development in the knowledge of transparent collocations was significantly more than that in the knowledge development of the opaque collocations (mean difference $M= 1.17$). The inconsistency in the results of knowledge development could be credited to the characteristics of the items such as their degree of difficulty to the participants or the test design; recalling opaque collocations in receptive test designs may be easier than remembering the same items in the productive tests. Additionally, opaque collocations may attract learners' attention more than transparent items due to their emphasis on more challenging items than the transparent collocations. Accordingly, opaque collocations seemed less challenging in the receptive tests and more difficult to guess in the productive tests. The anticipated reasons behind inconsistency in the results will be discussed in more detail in the discussion chapter.

From a L1 congruency perspective, the test results of the individual tests, i.e., one-shot data tests, indicated no statistically significant influence of the congruency with participants' L1 collocations on the participants' collocational knowledge (see Table 20). Similarly, developmentally, the results revealed, receptively and productively, no statistically significant difference between the amount of participants' knowledge development of congruent and incongruent collocations (see Table 22). Accordingly, the results indicated that congruency

with participants' L1 seemed to have less influence on their collocational knowledge development than frequency and syntactic structure features but more than the influence of the semantic transparency feature.

As far as participants' receptive and productive collocational knowledge development, the results generally indicated a significant difference between the amount of knowledge development of the participants' receptive and productive knowledge of collocations. While the development in the participants' receptive knowledge of collocations was not statistically significant (gain $M = -.79$), their productive knowledge developed statistically significantly (gain $M = 3.83$, + 6.7%). The development in the participants' productive collocational knowledge was more than that of their receptive knowledge by +8.1% (gain difference $M = 4.62$). In other words, the amount of productive knowledge development was larger than that of their receptive knowledge over the school year. This may imply that participants' productive knowledge develops faster than their receptive knowledge at this level of language proficiency. L2 learners may feel more confident after many years of acquiring vocabulary to concentrate more, at this level of language proficiency, on producing or using collocations rather than focusing on developing their receptive knowledge and acquiring or knowing more collocations further.

Table (22): Collocational Knowledge of the contrasted variables

Contrasted levels	Level 1 (L1)	Level 2 (L2)	Gain= (L2-L1)	Significance
Receptive (L1) vs Productive (L2) knowledge	$M = -.79$ $SE = .41$	$M = 3.83$, $SE = .66$	$M = 4.62$, $SE = .75$	Significant, $t(251) = 6.17$, $p < .05$, $d = .53$
Females vs Males, receptive knowledge	$M = -.83$, $SE = .70$	$M = -.71$, $SE = .50$	$M = .12$, $SE = .86$	Not significant, $t(250) = 1.55$, $p > .05$
Females vs Males, productive knowledge	$M = 4.42$, $SE = .82$	$M = 2.70$, $SE = 1.08$	$M = -1.72$, $SE = 1.38$	Not significant, $t(250) = -1.25$, $p > .05$
Female total knowledge vs	$M = 3.59$, $SE = .98$	$M = 1.99$, $SE = 1.34$	$M = -1.6$, $SE = 1.67$	Not significant,

Male total knowledge				$t(250) = -.96$, $p > .05$
Lower vs Higher frequency receptive average knowledge	M= 18.57, SE= .21	M= 23.27, SE= .16	M= -4.70, SE= .16	Significant, $t(251) = -28.72$, $p < .05$, $d = .63$
Lower vs Higher frequency productive average knowledge	M= 11.76, SE= .35	M= 18.81, SE= .32	M= -7.06, SE= .18	Significant, $t(251) = -38.83$, $p < .05$, $d = .55$
Lower vs Higher frequency receptive gain score	M= .01, SE= .27	M= -.80, SE= .25	M= .81, SE= .31	Significant, $t(251) = 2.62$, $p < .05$, $d = .20$
Lower vs Higher productive frequency gain score	M= 2.22, SE= .36	M= 1.61, SE= .35	M= .61, SE= .28	Significant, $t(251) = 2.20$, $p < .05$, $d = .12$
Grammatical vs Lexical receptive knowledge 1	M= 21.40, SE= .22	M= 20.84, SE= .21	M= .56, SE= .21	Significant, $t(251) = 2.60$, $p < .05$, $d = .16$
Grammatical vs Lexical receptive knowledge 2	M= 20.80, SE= .22,	M= 20.65, SE= .21,	M= .15, SE= .18,	Not significant, $t(251) = .84$, $p > .05$
Grammatical vs Lexical productive knowledge 1	M= 14.02, SE= .37,	M= 14.63, SE= .35,	M= -.62, SE= .26,	Significant, $t(251) = -2.34$, $p < .05$, $d = .11$
Grammatical vs Lexical	M= 18.58, SE= .47,	M= 13.90, SE= .39,	M= 4.67, SE= .36,	Significant, $t(251) = 12.90$,

productive knowledge 2				$p < .05$, $d = .68$
Grammatical vs Lexical receptive average knowledge	M= 21.10, SE= .18	M= 20.75, SE= .18	M= .35, SE= .15	Significant, $t(251) = 2.35$, $p < .05$, $d = .12$
Grammatical vs Lexical productive average knowledge	M= 16.30, SE= .35	M= 14.27, SE= .34	M= 2.03, SE= .23	Significant, $t(251) = 8.99$, $p < .05$, $d = .37$
Grammatical vs Lexical receptive knowledge	M= -.60, SE= .24	M= -.19, SE= .24	M= -.41, SE= .25	Not significant, $t(251) = 1.63$, $p > .05$
Grammatical vs Lexical productive knowledge	M= 4.56, SE= .47	M= -.73, SE= .31	M= 5.29, SE= .44	Significant, $t(251) = 11.91$, $p < .05$, $d = .84$
Total receptive transparent scores 1 vs Total receptive opaque scores 1	M= 21.15, SE= .20	M= 21.09, SE= .23	M= .06, SE= .21	Not significant, $t(251) = .27$, $p > .05$
Total receptive transparent scores 2 vs Total receptive opaque scores 2	M= 20.17, SE= .23	M= 21.28, SE= .21	M= -1.12, SE= .19	Significant, $t(251) = -5.95$, $p < .05$, $d = .32$
Total productive transparent scores 1 vs	M= 14.06, SE= .33	M= 14.60, SE= .36	M= -.54, SE= .19	Significant, $t(251) = -2.86$, $p < .05$, $d = .10$

Total productive opaque scores 1				
Total productive transparent scores 2 vs Total productive opaque scores 2	M= 16.75, SE= .41	M= 15.73, SE= .40	M= 1.02, SE= .19	Significant, $t(251)= 5.30$, $p < .05$, $d= .16$
Transparent vs Opaque average receptive	M= 20.66, SE= .18	M= 21.19, SE= .18	M= -.53, SE= .14	Significant, $t(251)= -3.81$, $p < .05$, $d= .18$
Transparent vs Opaque average productive	M= 15.40, SE= .33	M= 15.17, SE= .33	M= .24, SE= .14	Not significant, $t(251) = 1.76$, $p > .05$
Transparent vs Opaque receptive knowledge	M= -.98, SE= .24	M= .19, SE= .25	M= -1.17, SE= .28	Significant, $t(251)= -4.23$, $p < .05$, $d= .30$
Transparent vs Opaque productive knowledge	M= 1.14, SE= .34	M= 2.69, SE= .37	M= 1.55, SE= .27	Significant, $t(251)= 5.83$, $p < .05$, $d= .27$
Total receptive congruent scores 1 vs Total receptive incongruent scores 1	M= 23.36, SE= .23	M= 18.88, SE= .19	M= 4.48, SE= .18	Significant, $t(251)= 24.31$, $p < .05$, $d= .56$
Total receptive congruent scores 2 vs	M= 23.02, SE= .22	M= 18.43, SE= .22	M= 4.59, SE= .19	Significant, $t(251)= 24.55$, $p < .05$, $d= .55$

Total receptive incongruent scores 2				
Total productive congruent scores 1 vs Total productive incongruent scores 1	M= 15.61, SE= .37	M= 13.05, SE= .32	M= 2.56, SE= .17	Significant, $t(251)= 15.41$, $p < .05$, $d= .47$
Total productive congruent scores 2 vs Total productive incongruent scores 2	M= 17.43, SE= .46	M= 15.05, SE= .35	M= 2.38, SE= .22	Significant, $t(251)= 11.03$, $p < .05$, $d= .37$
Congruent vs Incongruent average receptive	M= 23.19, SE= .18	M= 18.66, SE= .17	M= 4.53, SE= .13	Significant, $t(251)= 35.17$, $p < .05$, $d= .63$
Congruent vs Incongruent average productive	M= 16.52, SE= .37	M= 14.05, SE= .29	M= 2.47, SE= .15	Significant, $t(251)= 16.88$, $p < .05$, $d= .47$
Congruent vs Incongruent receptive knowledge	M= -.34, SE= .26	M= -.45, SE= .23	M= .12, SE= .27	Not significant, $t(251)= .43$, $p > .05$
Congruent vs Incongruent productive knowledge	M= 1.82, SE= .38	M= 2.00, SE= .32	M= -.18, SE= .25	Not significant, $t(251)= -.73$, $p > .05$

As far as Table (22) is concerned, it should be noted that the table includes results from both paired samples and independent samples t-tests.

Finally, as it was explained in the methodology chapter (Section 4.6.7), items of the receptive test 1 were the same items of the productive test 2. Similarly, items of the productive test 1 were the same items used in the receptive test 2. In this part of analysis, I investigated whether participants' general knowledge, receptively and productively, of these two groups of items differed significantly from each other, i.e., participants' knowledge of (receptive 1+ productive 2) vs (receptive 2+ productive 1). Additionally, the analysis aimed at examining whether items of the first group were easier or more difficult than that of the second group. The results revealed a statistically significant difference between participants' knowledge of both groups of items ($t(251) = -6.17, p < .05, d = .30$; gain scores = 4.62). Participants' scores on the items of the (receptive 1 & productive 2) tests were higher than their scores on the items of the tests (receptive 2 & productive 1). This indicated that participants performed better on the items of (receptive 1 & productive 2) than the items of (receptive 2 & productive 1). In other words, participants' knowledge of the earlier test items was more than that of the latter which could imply either the items of the earlier test group were easier than the latter group items, which will be regarded as a shortcoming in the test or some other factors contributed to this such as features of collocations or any other reasons. However, this comparison might be weak since it compares knowledge of two different sets of items at different times of two different levels of collocational knowledge (receptive/ productive). This point will be discussed in detail in the discussion chapter.

5.5 Findings summary

The data analyses came out with a number of findings which answered the study questions mentioned in the methodology chapter as they are listed below.

5.5.1 The influence of gender on the participants' collocational knowledge and development

Generally, the results revealed a statistically significant difference between male and female participants' collocational knowledge in all the individual receptive and productive tests. Male participants generally achieved higher mean scores than female participants on the tested collocations. As for participants' collocational knowledge development over the school year, receptively, the results indicated no significant developments in the collocational knowledge of neither male nor female participants. Contrastively, the results of the productive

tests revealed statistically significant developments in the collocational knowledge of both male and female participants. However, the results did not reveal statistically significant differences between the gain scores of males and females neither in their receptive nor productive knowledge of collocations. In other words, the differences in the amounts of knowledge development between both sexes were not statistically significant despite the significant differences between males and females in the individual receptive and productive tests.

5.5.2 The influence of frequency on the participants' collocational knowledge and development

Generally, learners' higher knowledge of collocations was associated with items of higher frequency and lower knowledge was associated with items of lower frequency. The results, from the receptive and productive tests of collocation, indicated a statistically significant, positive, moderate correlation between frequency of the collocations in the curricula textbooks and participants' knowledge of those items. Developmentally, the results of the receptive and productive tests indicated that the amount of participants' knowledge development of the lower frequency collocations was higher than that of the higher frequency items. In other words, participants' collocational knowledge of the lower frequency items developed more than their knowledge of the items of higher frequency over the school year, which might imply a shift in the participants' focus at this level of language learning to less frequent and more challenging collocations.

5.5.3 The influence of the syntactic structure on the participants' collocational knowledge and development

Generally, the syntactic structure of collocations seemed to have a statistically significant influence on the participants' collocational knowledge, specifically their productive knowledge of the grammatical items. Results showed that participants' knowledge of the grammatical collocations exceeded significantly their knowledge of the lexical items. However, this influence of the syntactic structure on the participants' collocational knowledge was not consistent in all the tests as its influence on the collocational knowledge of the four syntactic structures varied across the receptive and productive tests. While the results showed no statistically significant difference in the participants' knowledge of the different syntactic patterns of collocations in the productive test 1 and

receptive test 2, participants' collocational knowledge of the four syntactic structures differed significantly in the receptive test 1 and productive test 2. Developmentally, the results of the receptive tests revealed no statistically significant change in the participants' knowledge of the lexical collocations whereas their collocational knowledge of the grammatical items appeared to have decreased significantly over the school year. Additionally, the results revealed that the difference in the amount of change, gain scores, between participants' knowledge of the lexical and grammatical collocations was not statistically significant. Productively, while the results revealed a significant decrease in the participants' knowledge of the lexical collocations, the results indicated that participants' collocational knowledge of the grammatical items had increased significantly during the school year. In addition, the results revealed that the difference in the amount of knowledge development between participants' knowledge of the lexical and grammatical collocations was significant. In other words, most of the development happened in the participants' knowledge of the grammatical collocation patterns.

5.5.4 The influence of the semantic transparency on the participants' collocational knowledge and development

As far as the influence of the semantic transparency feature on the participants' collocational knowledge is concerned, the results indicated no statistically significant impact of the degrees of semantic transparency on the participants' collocational knowledge in all the one-shot data receptive and productive tests. Developmentally, the results indicated that participants' receptive knowledge of opaque collocations increased significantly more than their knowledge of the transparent items. In contrast, productively, the results revealed that the amount of participants' knowledge development in the transparent collocations exceeded that of the opaque collocations. In other words, participants' productive knowledge of transparent collocations developed significantly more than their knowledge of opaque collocations over the school year. In conclusion, the feature of semantic transparency had a little influence on the participants' knowledge and development and this influence varied according to the knowledge dimension of collocations as whether it was receptive or productive dimension. In comparison with the other features, it seemed to have the least influence on the participants' collocational knowledge and development.

5.5.5 The influence of the L1 congruency on the participants' collocational knowledge and development

The results of the individual receptive and productive tests indicated no statistically significant influence of the degrees of congruency with L1 on the participants' collocational knowledge. Additionally, through analysing participants' collocational knowledge development, the results indicated that the difference in the amount of participants' knowledge development of congruent and incongruent collocations, receptively and productively, was not statistically significant. Nevertheless, the difference between the mean scores of the congruent and incongruent items in the receptive and productive tests revealed a positive influence of congruency on the participants' collocational knowledge. Participants' knowledge of the congruent collocations was statistically higher than their knowledge of the incongruent collocations. Additionally, this feature seemed to have less influence from frequency and syntactic features of collocations but more influence than the semantic transparency of the collocational items on the participants' collocational knowledge and development.

5.5.6 The ordinal influence of features of collocations on the participants' collocational knowledge and development

According to the results of the data analyses, the ordinal sequence of the collocational features influence on the participants' collocational knowledge and development from the most to the least appeared to be as follows: frequency, syntactic structure, congruency with the participants' L1, and finally semantic transparency.

5.5.7 How does participants' collocational knowledge develop over a school year?

In total, the results indicated that participants had a mastery of an amount of collocational knowledge which seemed to have developed significantly through the school year. However, knowledge development appeared to be small and slow. Moreover, participants' receptive and productive knowledge did not seem to have developed in parallel to each other. Receptively, the results did not reveal a statistically significant change in the participants' knowledge of collocations over the school year. In contrast to the receptive knowledge, participants' productive knowledge of collocations seemed to have increased

significantly over the school year. However, comparing participants' receptive knowledge with their productive knowledge of collocations revealed that the latter lagged much behind the former. Interestingly, the results indicated that participants' productive knowledge developed at a faster rate than that of the receptive knowledge in the study period. Additionally, analysing the correlation between participants' receptive and productive knowledge indicated that higher levels of participants' receptive knowledge were generally associated with higher levels of their productive knowledge. Finally, the results indicated a developmental pattern in the participants' collocational knowledge. The results indicated that learners' receptive collocational knowledge rather stabilised at this stage of language proficiency while their productive knowledge seemed to have developed significantly over the school year.

5.5.8 Influence of the test designs on the data quality

The productive knowledge test format appeared to be a better indicator of the participants' collocational knowledge, and a better tool for exploring the influence of the collocational features on the participants' collocational knowledge and development than the receptive test format. This was evidenced in the consistency of the productive data, its development over time, and participants' commitment in answering the test items which was observed from participants' performance.

Chapter six

Discussion of the findings

The current study attempted to examine the influence of four features of collocations- frequency, syntactic structure, semantic transparency, and congruency with L1- and gender on the Kurdish high school students' knowledge and development of selected English collocations they had experienced, i.e., extracted from their curriculum English textbooks (Year1 up to Year 11). It also explored learners' collocational knowledge development over one school year.

The participants' study period was seen as consisting of two parts; the first part regarded their collocational knowledge at Year 10 whereas the second part regarded their collocational knowledge development over the school Year 11. The first part of knowledge was examined by the data obtained from the first wave of data collection whereas the second part was investigated through analysing the data of the second wave in comparison with that of the first wave.

To achieve the study aims, five factors were considered as the key elements that could affect participants' collocational knowledge and development:

- gender difference; as how being male or female could have an impact on the participants' collocational knowledge and development.
- the frequency of occurrences of collocations in the participants' school textbooks
- the syntactic structure of collocations
- the semantic transparency of collocations
- collocations congruency with learners' L1

Accordingly, the study attempted to seek answers for the following questions:

- How does gender difference have impact on the participants' collocational knowledge and development?
- How does frequency of collocations in the participants' curriculum textbooks affect their collocational knowledge and development?
- How does the syntactic structure of collocations influence participants' collocational knowledge and development?
- How does the semantic transparency of collocations influence participants' collocational knowledge and development?

- How does congruency with participants' L1 equivalents have impact on the participants' collocational knowledge and development?

Additionally, the study attempted to answer the following two questions:

- Is there an ordinal pattern of the influence of the four features on the participants' collocational knowledge and development?
- How does participants' collocational knowledge develop over a school year?

In the following sections, the findings will be discussed in the order of the study questions mentioned in the methodology chapter: gender, frequency, syntactic structure, semantic transparency, L1 congruency, order of features influence, and finally participants' collocational knowledge and development over the school year.

6.1 The influence of gender on the participants' collocational knowledge and development

The present study results generally revealed significant differences between the mean scores of males and females in the individual receptive and productive tests. Male participants appeared to achieve significantly higher mean scores than female participants on the tested collocations.

Receptively, in the first test, males ($M= 43.57$) performed higher than females ($M= 41.53$) and the difference between both means was statistically significant but the effect size was small ($d= .30$). Similarly, in the second receptive test, males ($M= 42.86$) on average scored higher than females ($M= 40.70$) but the effect size was small as well ($d= .31$). Productively, in the first test, males ($M= 31.72$) performed significantly better than females ($M= 27.04$) and the effect size was small ($d= .39$). However, in the second productive test, although the mean scores of males ($M= 34.43$) seemed higher than that of the females ($M= 31.45$), the difference between the mean scores of both genders was not statistically significant. In general, males performed significantly better in the receptive and productive tests than females (see Table 5.21).

This result was different from the results of many other previous studies where some revealed females performing better than males (e.g., Gaballa & Alkhayri, 2014) or those which revealed no significant difference between males and females (e.g., Ganji, 2012; Tabrizi & Moghadam, 2016). The result lends

support to studies such as (Mazaji, 2015) which indicated that gender had a significant impact on L2 learners' vocabulary development.

However, males' outperforming females should not necessarily imply that males are genetically, as Jespersen (1922) believes, outperforming females specifically in learning language. The reasons behind this difference could be attributed to learners' characteristics such as motivation, social factors such as the tribal customs and traditions, the female status in the society, their freedom of decision making, future expectations, economic such as their financial status, political atmosphere, educational systems, or all these factors together.

Although women situation and their status in the Kurdish society are improving, women are still under considerable social, political, and economic pressures. Consequently, such a situation could have consequences on their motivation to promote their self-esteem, which in turn could affect their future expectations, consequently, influencing their desire to earn knowledge, in general, and learn a foreign language, in particular.

Developmentally, the results indicated no significant increase in the receptive knowledge of neither male nor female participants. Additionally, the data analyses of the knowledge development between Time 1 and Time 2 revealed that the difference between males' receptive gain scores ($M = -.71$) and females' receptive gain scores ($M = -.83$) was not significant. This indicated that although males did better than females in both receptive tests, the change in the receptive knowledge of the male participants was not significantly higher than the change in the females' receptive knowledge of collocations (see Table 5.22). Productively, the results revealed significant developments in the productive knowledge of both males and females. However, the results did not reveal a significant difference between the gain scores of males and females (males' gain scores: 2.70; females' gain scores= 4.42). In other words, the difference in the amounts of knowledge development of both sexes was not statistically significant over the school year (see Table 5.22). In conclusion, the results revealed a significant difference between male and female participants' knowledge of collocations; males generally held more collocational knowledge than females, but males' collocational knowledge development over the school year was not significantly higher than females' knowledge development. The non-significant difference of knowledge development between males and

females could be, receptively and productively, attributed to either improvement of the female status by time specifically their motivation, confidence, political or any contextual factors which were negatively affecting their knowledge development, or that one school year was not sufficient to reveal the influence of gender on this aspect of knowledge development. Hence, this situation may need longitudinal studies of longer periods and also more studies investigating the influence of the contextual factors on the learners' knowledge development and test these assumptions in the local context, Kurdistan.

6.2 The influence of frequency on the participants' collocational knowledge and development

The results of the current study showed frequency of occurrence of collocations in the participants' curriculum textbooks as the most powerful feature which influenced participants' collocational knowledge and development. The results revealed a statistically significant, positive, moderate correlation between the frequency of the collocations in the curricula textbooks and participants' collocational knowledge and development. This result is consistent with the growing body of evidence that indicates a positive correlation between L2 learners' collocational knowledge and the frequency of occurrence of collocations (Bybee, 2008; Bybee & Hopper, 2001; Durrant & Schmitt, 2010; González Fernández & Schmitt, 2015; Gyllstad & Wolter, 2015; Mueller, 2011; Myles, Hooper, & Mitchell, 1998; Siyanova-Chanturia et al., 2011).

The results generally revealed that participants' higher knowledge was associated with collocations of higher frequency, and lower knowledge was accompanying collocations of lower frequency. This is strongly evidenced in the studies of numerous researchers which indicated that L2 learners' collocational knowledge is correlated positively with the rate of frequency of collocations; higher levels of collocational knowledge are expected to be associated with collocations of higher frequency (e.g., Alsakran, 2011; Bergström, 2008; Bybee & Hopper, 2001; Durrant, 2008, 2014; Durrant & Schmitt, 2009, 2010; Gitsaki, 1996; Hama, 2011; Howarth, 1998; Nesselhauf, 2003; Shehata, 2008; Siyanova-Chanturia, 2015; Webb, 2008; Wolter & Gyllstad, 2013).

Receptively, the results showed a positive correlation between participants' knowledge and frequency of collocations. The correlation strength between frequency of collocations in the textbooks and participants' knowledge of the

items was moderate in both receptive tests (receptive 1: $r = .47$; receptive 2: $r = .41$) (see Table 5.20). Accordingly, the results revealed that participants' receptive knowledge of higher frequency items was significantly higher than their knowledge of lower frequency items, $t(251) = 28.72$, $p < .05$, $d = .63$, and the effect size was medium (see Table 5.22).

Developmentally, while participants' knowledge of higher frequency collocations seemed to have decreased slightly from ($M = 23.67$) in the first test to ($M = 22.87$) in the second test but the effect size was small ($d = .25$), the results did not indicate a statistically significant change in the participants' knowledge of lower frequency collocations (see Table 5.21). The interesting observation was that participants' knowledge of the lower frequency items appeared to have developed significantly over the school year more than that of the higher frequency ones receptively, $t(251) = -2.62$, $p < .05$, $d = .20$, and productively, $t(251) = -2.20$, $p < .05$, $d = .12$, but the effect size was small (see Table 5.22). Accordingly, if this result could be proved in further studies, it might indicate that learners' focus shift at this stage of language learning, taking the local context into consideration, towards collocations of less frequency. In other words, learners at this stage begin to put emphasis on the items which they are not efficiently familiar or rarely encountered in their curricula textbooks over the previous school years. In sum, the results indicated a moderate, positive influence of frequency on the participants' receptive collocational knowledge and development even though participants' receptive knowledge looked generally stabilising or decreasing in certain aspects over the school year.

Productively, the results showed a positive correlation between participants' productive knowledge and frequency of collocations. The correlation strength between frequency of collocations in the textbooks and participants' knowledge of the items was moderate in both productive tests (productive 1: $r = .49$; productive 2: $r = .54$). Accordingly, the correlation between knowledge and frequency in the productive tests looked stronger than that between knowledge and frequency in the receptive tests (see Table 5.20). Accordingly, participants' knowledge of higher frequency items was significantly higher than their knowledge of lower frequency items, $t(251) = 38.83$, $p < .05$, $d = .55$, and the effect size was medium (see Table 5.22)

Developmentally, in contrast to their receptive knowledge, participants' productive knowledge of higher frequency collocations indicated a statistically significant increase from ($M=18.01$) in the first test to ($M= 19.62$) in the second test but the effect size was small, $t(251)= 4.63$, $p < .05$, $d= .28$. Similarly, participants' productive knowledge of lower frequency collocations indicated a statistically significant increase from ($M= 10.65$) in the first test to ($M= 12.87$) in the second test and the effect size was small, $t(251)= 6.10$, $p < .05$, $d= .35$. In sum, the results indicated a moderate positive influence of frequency on the participants' productive collocational knowledge and development. Accordingly, participants' knowledge of higher frequency items was significantly higher than their knowledge of lower frequency items (see Table 5.21).

The results also indicated that the positive influence of frequency on the participants' productive knowledge was higher than its influence on their receptive knowledge since the correlation between knowledge and frequency in the former was stronger than that in the latter. This might imply that the influence of this feature varied not only according to the rate of frequency but also according to the aspect of knowledge, i.e., whether receptive or productive. Accordingly, this result supports the claim that influence of frequency on L2 learners' knowledge and development is not straightforward as it helps develop only specific aspects of L2 learners' collocational knowledge and it does not develop all aspects of learners' collocational knowledge (e.g., Szudarski & Carter, 2014). Accordingly, frequency impact varies according to the way and aim for which frequency is used (e.g., Durrant & Schmitt 2010). Furthermore, there are some other aspects which cannot be easily developed even with a considerable amount of exposure to L2 collocations (e.g., Laufer & Waldman, 2011; Yamashita & Jiang, 2010). However, the result adds to that of Szudarski & Carter (2014) which found that frequency enhances learners' receptive and productive knowledge at the form level, in that frequency enhances L2 learners' productive knowledge more than their receptive knowledge of collocations at this level of language proficiency.

In conclusion, the results highlighted frequency as one the of major influential factors affecting L2 learners' collocational knowledge and development. Additionally, frequency seemed to be more influential on the productive aspect

of participants' collocational knowledge than their receptive knowledge of collocations at this level of L2 learners' language proficiency.

6.3 The influence of the syntactic structure on the participants' collocational knowledge and development

The study explored the influence of four syntactic patterns of collocations on the participants' collocational knowledge and development of which two patterns (v+n and adj+n) were lexical collocations and the other two patterns were grammatical collocations of simple syntactic structures (phrasal verb and prep+n).

In general, the influence of the syntactic structure of collocations on the participants' collocational knowledge and development was inconsistent across the receptive and productive tests in both waves of data collection. While the differences among the facility values of the four patterns of collocations were significant in the receptive test 1, they were not significant in the productive test 1. Conversely, the differences were not significant in the receptive test 2 but significant in the productive test 2. Accordingly, the results were inconsistent between receptive 1 and receptive 2, and productive 1 and productive 2 (see Table 5.20). The inconsistency appeared to be more related to the characteristics of the items themselves. For example, the statistically significant influences were in the tests receptive 1 and productive 2 which consisted of the same items whereas the significant influences were in the tests receptive 2 and productive 1 which consisted of the same items. This might imply that participants' knowledge of the test items (receptive 1 and productive 2) was more able to reflect the influence of the syntactic structure on the participants' collocational knowledge than items of the tests (receptive 2 and productive 1). A further reason could be that a larger number of items with specific features had occurred, unintentionally, in one group rather than been balanced between both tests. The balance is not meant in terms of their frequency, syntactic structure, semantic transparency, and L1 congruency as they were generally maintained balanced in all the tests, but in terms of factors such as participants' knowledge of the test items, i.e., difficulty of the items, which could not be estimated or predicted prior conducting the tests.

In the receptive test 1 (see Table 5.9 and Table 5.10), the results indicated significant differences among the facility values of the four patterns of

collocations (v+n fv= .74, adj+n fv= .70, phrasal verb fv= .84, and prep+n fv= .68). More precisely, the statistically significant differences were between phrasal verb and adj+n collocations, and between phrasal verb and prep+n collocations. Accordingly, phrasal verb collocations seemed to be less challenging than adj+n and prep+n collocations but not less challenging than v+n collocations.

Similarly, in the productive test 2 (see Table 5.18 and Table 5.19), the results indicated that the overall differences among the facility values of the four patterns of collocations (v+n= .51, adj+n= .45, phrasal verbs= .67, and prep+n= .66) were statistically significant. However, the test results did not identify where the significant differences among the individual syntactic groups occurred.

In sum, the results of these two tests indicated a significant influence of the syntactic structure on the participants' collocational knowledge. This result supports the findings of similar studies which explored the influence of the syntactic structure of collocations on the L2 learners' knowledge and development (e.g., Alsakran, 2011; Gaballa & Al-Khayri, 2014; Gitsaki, 1996; Koya, 2005; Shehata, 2008; Shokouhi & Mirsalari, 2010).

However, in the receptive test 2 (see Table 5.15 and Table 5.16), the results indicated that the difference among the facility values of the four syntactic structures (v+n= .72, adj+n= .71, phrasal verbs= .77, and prep+n= .71) were not statistically significant. Similarly, the productive test 1 results (see Table 5.12 and Table 5.13) did not reveal significant differences among the facility values of the four patterns of collocations (v+n= .59, adj+n= .44, phrasal verbs= .41, and prep+n= .61).

Although analysing influence of the individual syntactic structure of the four patterns on the participants' knowledge showed inconsistent results, running analyses based on the dichotomy (lexical-grammatical) rather than the four patterns individually made the distinction between participants' knowledge of the different types of collocations clearer and the results more consistent (see Table 5.22). Receptively, in the receptive test 1, participants' collocational knowledge of the grammatical items (M= 21.40) was significantly higher than their knowledge of the lexical items (M= 20.84), $t(251) = 2.60$, $p < .05$, $d = .16$, but the effect size was small. Grammatically, the highest achievement was on the

phrasal verb pattern (fv= .84), and the items *sit down* (fv= .96), *talk about* (fv= .96), *look at* (fv= .96), *stand up* (fv= .95), and *pick up* (fv= .94) received the highest scores. Conversely, results of the receptive test 2 did not reveal a significant difference between participants' knowledge of the grammatical (M= 20.80) and lexical collocations (M= 20.65), $t(251) = .84, p > .05$. However, comparing participants' average knowledge of the grammatical collocations ((receptive 1 + receptive 2)/2) with their average knowledge of the lexical collocations indicated that participants' receptive knowledge of the grammatical collocations was significantly higher than their knowledge of the lexical items, $t(251) = 2.35, p < .05, d = .12$, but the effect size was trivial (see Table 5.22).

Developmentally, while participants' receptive lexical collocational knowledge revealed no statistically significant knowledge development (Time 1: M= 20.84; Time 2: M= 20.65, $t(251) = -.81, p > .05$), their receptive grammatical knowledge decreased significantly (Time 1: M= 21.40; Time 2: M= 20.80, $t(251) = -2.46, p < .05, d = .17$), but the effect size was trivial (see Table 5.21). However, the difference in the amount of the receptive knowledge change of the grammatical (gain score M= -.60) and lexical collocations (gain score M= -.19) was not significant, $t(251) = 1.63, p > .05$ (see Table 5.22). In conclusion, the results generally indicated that participants' receptive knowledge of the grammatical collocations significantly exceeded their knowledge of the lexical collocations, but amount of knowledge development of the grammatical collocations was not significantly more than amount of the receptive knowledge development of the lexical items over the school year.

Productively, in the productive test 1, participants' collocational knowledge of the lexical items (M= 14.63) was significantly higher than their knowledge of the grammatical items (M= 14.02), $t(251) = -2.34, p < .05, r = .11$, but the effect size was trivial. Lexically, the highest achievement was on the pattern v+n (fv= .59), and the collocations *play the guitar* (fv= .92), *cook dinner* (fv= .88), *climb mountains* (fv= .87), *drink tea* (fv= .87), and *catch as bus* (fv= .75) received the highest scores. Conversely, results of the productive test 2 indicated that participants' knowledge of the grammatical (M= 18.58) was significantly higher than their knowledge of the lexical collocations (M= 13.90), $t(251) = 12.90, p < .05, d = .68$. Grammatically, the highest achievement was on the phrasal verb pattern (fv= .67), and the items *sit down* (fv= .87), *pick up* (fv= .83), *stand up* (fv.

83), *talk about* ($fv = .80$), *look at* ($fv = .77$), and *wake up* ($fv = .77$) received the topmost scores.

Similar to their receptive knowledge, comparing participants' average knowledge of the grammatical collocations ($(\text{productive 1} + \text{productive 2})/2$) with their average knowledge of the lexical collocations indicated that participants' productive knowledge of the grammatical collocations was significantly higher than their knowledge of the lexical items, $t(251) = 8.99$, $p < .05$, $d = .37$, but the effect size was small (see Table 5.22).

Developmentally, while participants' productive lexical collocational knowledge revealed a statistically significant knowledge decrease (Time 1: $M = 14.63$; Time 2: $M = 13.90$, $t(251) = -2.35$, $p < .05$, $d = .13$), but the effect size was trivial, their productive grammatical knowledge increased significantly (Time 1: $M = 14.02$; Time 2: $M = 18.58$, $t(251) = 9.81$, $p < .05$, $d = .68$, over the school and the effect size was medium (see Table 5.21). Additionally, the difference in the amount of the productive knowledge change of the grammatical (gain score $M = 4.56$) and lexical collocations (gain score $M = -.73$) was significant, $t(251) = 11.91$, $p < .05$, $d = .84$, and the effect size large (see Table 5.22). In conclusion, the results generally indicated that participants' productive knowledge of the grammatical collocations exceeded their knowledge of the lexical collocations. Additionally, the results revealed that amount of knowledge development of the grammatical collocations was significantly more than amount of knowledge development of the lexical items over the school year. In other words, the results showed that the influence of the syntactic structure of collocations on the participants' productive collocational knowledge development was significantly large.

Accordingly, the results generally indicated that participants' knowledge of the grammatical collocations was significantly higher than their knowledge of the lexical collocations and that their knowledge of the grammatical collocations developed outstandingly more than their knowledge of the lexical collocations over the school year. This result lends support to the studies indicated that grammatical collocations, specifically of simple syntactic structures, are easier to acquire and create less challenges to the L2 learners than lexical collocations (e.g., Gitsaki, 1996). However, the result contrasts with those studies which indicated that grammatical collocations are more challenging to L2 learners than

lexical collocations (e.g., Bahardoust & Moeini, 2012; Mahmoud, 2005; Shokouhi & Mirsalari, 2010).

In sum, the results commonly, though inconsistent in some aspects across all the tests, revealed that the syntactic structure of the items had a significant influence on the participants' collocational knowledge and development. Grammatical collocations were generally less challenging than lexical collocations. Accordingly, participants' collocational knowledge and development of the former was higher than that of the latter.

6.4 The influence of the semantic transparency on the participants' collocational knowledge and development

The results of all the one-shot data tests, in which the analyses were based on the use of facility values of the items, revealed no significant influence of the semantic transparency feature on the participants' collocational knowledge (see Table 5.20). However, conducting the analysis by comparing participants' collocational knowledge scores of both types of collocations revealed the influence of this feature on the participants' knowledge and development more clearly (see Table 5.22).

Receptively, the results of the receptive test 1 indicated no statistically significant difference between participants' knowledge mean scores of transparent ($M= 21.15$) and opaque collocations ($M= 20.09$), $t(251) = .27$, $p > .05$. However, results of the receptive test 2 indicated that participants' knowledge of the opaque collocations ($M= 21.28$) was significantly higher than their knowledge of the transparent items ($M= 20.17$), $t(251)= 5.95$, $p < .05$, $d= .32$, but the effect size was small. The opaque items which received the highest scores were *cook dinner* ($fv= .94$), *good news* ($fv= .94$), *go on* ($fv= .94$), *climb mountains* ($fv= .92$), and *get up* ($fv= .92$).

Developmentally, the results did not show significant developments neither in the participants' knowledge of the transparent collocations which decreased significantly from (Time 1: $M= 21.15$) to (Time 2: $M= 20.17$), $t(251)= -4.04$, $p < .05$, $d= .29$, but the effect size was small, nor in the knowledge of the opaque collocations which revealed no significant change (Time 1: $M= 21.09$; Time 2: 21.28 ; $t(251)= .76$, $p > .05$) (see Table 5.21). However, the difference between the receptive gain scores of transparent ($M= -.98$) and opaque ($M= .19$) collocations was statistically significant; the gain score of the opaque

collocations was higher than the gain score of the transparent ones $t(251)=4.23$, $p < .05$, $d = .30$, but the effect size was small (see Table 5.22). In other words, the amount of the participants' receptive knowledge development of opaque collocations was statistically higher than the amount of knowledge development of their transparent collocations. However, in general, comparing participants' average scores on the transparent items ($M= 20.66$) with that on the opaque ones ($M= 21.18$) indicated that participants' knowledge of the opaque collocations was significantly higher than their knowledge of the transparent items, $t(251)= 3.81$, $p < .05$, $d = .18$, but the effect size was trivial.

Productively, similar to the results of the receptive test 2, the productive test 1 results revealed that participants' knowledge of the opaque collocations ($M= 14.60$) was significantly higher than their knowledge of the transparent items ($M= 14.06$), $t(251) = -2.86$, $p < .05$, $d = .10$, but the effect size was trivial. The opaque items which received the highest scores were *same time* ($fv = .89$), *cook dinner* ($fv = .88$), *climb mountains* ($fv = .87$), *get up* ($fv = .83$), and *catch a bus* ($fv = .75$). In contrast, the productive test 2 results indicated that participants' knowledge of the transparent items ($M= 16.75$) was significantly higher than their knowledge of the opaque ($M= 15.73$) collocations, $t(251)= 5.30$, $p < .05$, $d = .16$, but the effect size was trivial. The transparent items which received the highest scores were *ask a question* ($fv = .94$), *sit down* ($fv = .87$), *pick up* ($fv = .83$), *on the floor* ($fv = .81$), and *talk about* ($fv = .80$).

Developmentally, the results revealed a significant development in the participants' productive knowledge of the transparent collocations (Time 1: $M= 14.06$; Time 2: $M= 16.75$, $t(251)= 7.92$, $p < .05$, $d = .46$) but the effect size was rather small. Similarly, the results revealed a statistically significant development in the participants' knowledge of the opaque collocations during the school year under study (Time 1: $M= 14.60$; Time 2: $M= 15.73$, $t(251)= 3.10$, $p < .05$, $d = .19$), but the effect size was trivial, Table (5.21). However, participants' knowledge gain score of the transparent collocations ($M= 2.69$) was significantly higher than that of the opaque ones ($M= 1.14$), $t(251)= 5.83$, $p < .05$, $d = .27$, but the effect size was small. In other words, the amount of participants' productive knowledge development of the transparent collocations was significantly higher than the amount of knowledge development of the opaque collocations over the school year. However, in general, the results of

comparing participants' average score on the transparent items ($M= 15.40$) with their average scores on the opaque ones ($M= 15.17$) indicated no significant difference between participants' knowledge of the opaque and transparent items, $t(251)= 1.76$, $p> .05$, $d= .18$.

The observed inconsistency of the results could be partly attributed to the characteristics of the semantic feature itself. Although collocations were categorised in the present study into transparent and opaque, classifying items according to semantic transparency cannot be done in the same rather simple way which can be used in classifying items according to their number of occurrences or syntactic structures. For example, transparency is not graded as 1, 2,3.... etc. nor structured as in syntactic structure as v+n, adj+n...etc. into transparent, semi-transparent, opaque...etc. The area between both extremes, transparent vs opaque, is fuzzier than the area between high and low frequency items or as to which syntactic structures they belong, for example. Moreover, categorisation according to this feature is more subjective. Decisions about classifying the same items into transparent and opaque groups may differ from a researcher to another and even from a learner's comprehension of the semantic aspect of the items to another due to the subjective aspect of the decision making of this feature. Consequently, these elements may explain the difficulty of accurately categorising the items and maintaining the balance across all the tests in practice at designing the tests. Finally, a further reason for this inconsistency could be that one school year was not long enough to demonstrate a clear-cut path for the influence of semantic transparency on the participants' collocational knowledge and development.

In general, the semantic transparency of collocations seemed to have a little influence on the participants' knowledge and development of collocations. This result is consistent with the findings of the other studies which indicated the influence of the semantic transparency of collocations on the L2 learners' collocational knowledge and development (e.g., Gitsaki, 1996; Gyllstad & Wolter, 2015; Kellerman, 1978; Koya, 2005; Kurosaki, 2012; Nesselhauf, 2003, 2005). Additionally, the test results generally indicated that participants' knowledge of the opaque collocations was significantly higher than their knowledge of the transparent collocations. This result contrasts with some other studies (e.g., Gitsaki, 1996; Gyllstad & Wolter, 2015; Koya, 2005; Kurosaki,

2012; Nesselhauf, 2003, 2005) which indicate that the more transparent the items, the less challenging the items are to the L2 learners. However, the result supports the study conducted by Laufer & Waldman (2011) which indicates that transparency of collocations makes them less noticeable, accordingly, opaque collocations attract L2 learners' attention to their existence more than the transparent items do. However, this feature, semantic transparency, seemed to have less influence than the other features- frequency of occurrence, syntactic structure, and congruency with L1- on the participants' collocational knowledge and development.

6.5 The influence of the L1 congruency on the participants' collocational knowledge and development

Similar to the influence of the transparency feature, the results of all the one-shot data tests, in which the analyses were based on the use of facility values of the items, indicated no statistically significant influence of the congruency with the participants' L1 on the participants' collocational knowledge (see Table 5.20). However, comparing participants' mean scores on both congruent and incongruent items could reveal the influence of this feature, congruency with L1, on the participants' collocational knowledge and development more clearly (see Table 5.22).

Receptively, in the receptive test 1, the results indicated that participants' knowledge of congruent collocations ($M= 23.36$) was significantly higher than their knowledge of the incongruent items ($M= 18.88$), $t(251)= 24.31$, $p < .05$, $d= .56$, and the effect size was medium. The congruent items which received the highest scores were *sit down* ($fv= .96$), *talk about* ($fv= .96$), *for example* ($fv= .96$), *stand up* ($fv= .95$), *bad luck* ($fv= .94$), *cross the river* ($fv= .91$), and *ride a bicycle* ($fv= .90$). Similarly, results of the receptive test 2 indicated that participants' knowledge of the congruent items ($M= 23.02$) was significantly higher than their knowledge of the incongruent items ($M= 18.43$), $t(251)= 24.55$, $p < .05$, $d= .55$, and the effect size was medium. The highest achievements on the congruent collocations were on the items *correct answers* ($fv= .94$), *cook dinner* ($fv= .94$), *go away* ($fv= .94$), *go on* ($fv= .94$), *drink tea* ($fv= .92$), *climb mountains* ($fv= .92$), *out of control* ($fv= .92$), and *same time* ($fv= .90$).

Developmentally, the results did not show a significant development in the participants' receptive knowledge of the congruent collocations (Time 1: $M=$

23.36; Time 2: $M= 23.02$, $t(251)= -1.31$, $p> .05$) (see Table 5.21). Similarly, the results revealed no significant development in the participants' knowledge of the incongruent collocations, rather, it indicated a significant decline in their knowledge (Time 1: $M= 18.88$; Time 2: 18.43 , $t(251)= -1.99$, $p< .05$, $d= .14$), but the effect size was trivial. Additionally, the difference between the gain scores of congruent and incongruent collocations was not statistically significant. In other words, the difference in the amount of participants' receptive knowledge development of the congruent collocations and that of the incongruent collocations was not statistically significant (see Table 5.22). However, comparing participants' average scores on the congruent items with their average scores on the incongruent ones (see Table 5.22), indicated that participants' knowledge of the congruent collocations ($M= 23.19$) was significantly higher than their knowledge of the incongruent items ($M= 18.66$), $t(251)= 35.17$, $p< .05$, $d= .63$, and the effect size was medium.

The results indicated generally a decline in the participants' receptive knowledge of collocations (see Table 5.21). The general potential reasons of this decline in knowledge will be mentioned and discussed in detail in the total knowledge development section later in this chapter (Section 6.7). A further reason that could be added, which is more related to this feature of collocation, is degree of congruency with L1. Some items might not have been strong equivalents, which could be considered as a shortcoming in the test items, or stronger equivalents could not be obtained from their textbooks. A further possibility is that they might have been of different degrees of strength of congruency with their L1 equivalents, which is inevitable as gaining complete equal equivalents is a hard aim to achieve. Accordingly, I would suggest that we should take congruency as graded category rather than a binary category. As I believe, items are not either congruent or incongruent with learners' L1. Rather, they are gradable and differ from an item to another on the congruency scale. Additionally, I would argue that the same idea applies to semantic transparency which was discussed in the previous section (6.4). Items can differ from each other on the scale of semantic transparency and should not be taken as a binary category. A collocation could be completely congruent/transparent, half congruent/transparent or completely incongruent/opaque. However, in general, comparing the receptive average score of the congruent items ($M= 23.19$) with

that of the incongruent ones (18.66) indicated that participants' knowledge of the congruent collocations was significantly higher than their knowledge of the incongruent items, $t(251)= 35.17$, $p < .05$, $d = .63$, and the effect size was medium (see Table 5.22).

Productively, in the productive test 1, the results indicated that participants' knowledge of congruent collocations ($M = 15.61$) was significantly higher than their knowledge of the incongruent items ($M = 13.05$), $t(251) = 15.41$, $p < .05$, $d = .47$, but the effect size was rather small (see Table 5.22). The congruent items which received the highest scores were *same time* ($fv = .89$), *cook dinner* ($fv = .88$), *drink tea* ($fv = .87$), *climb mountains* ($fv = .87$), and *correct answers* ($fv = .87$). Similarly, results of the productive test 2 indicated that participants' knowledge of the congruent items ($M = 17.43$) was significantly higher than their knowledge of the incongruent items ($M = 15.05$), $t(251) = 11.03$, $p < .05$, $d = .37$, but the effect size was small (see Table 5.22). The highest achievements on the congruent collocations were on the items *as a question* ($fv = .94$), *sit down* ($fv = .87$), *stand up* ($fv = .83$), *talk about* ($fv = .80$), *wake up* ($fv = .77$), and *for example* ($fv = .77$).

Developmentally, the results revealed a significant development in the participants' productive knowledge of the congruent collocations (Time 1: $M = 15.61$; Time 2: $M = 17.43$, $t(251) = 4.79$, $p < .05$, $d = .28$), but the effect size was small. Similarly, the results indicated a significant development in the participants' knowledge of the incongruent collocations (Time 1: $M = 13.05$; Time 2: $M = 15.05$, $t(251) = 6.30$, $p < .05$, $d = .38$), but the effect size was small (see Table 5.21). However, the difference between the gain scores of congruent and incongruent collocations was not statistically significant. In other words, the difference between the amount of participants' productive knowledge development of the congruent collocations and that of the incongruent collocations was not statistically significant (see Table 5.22). Accordingly, no significant influence of the congruency of collocations with L1 on the knowledge development over the school year was identified. However, comparing participants' productive knowledge average scores on the congruent items with their average scores on the incongruent ones (see Table 5.22), indicated that participants' knowledge of the congruent collocations ($M = 16.52$) was

significantly higher than their knowledge of the incongruent items ($M= 14.05$), $t(251)= 16.88$, $p < .05$, $d= .47$, but the effect size was rather small.

Accordingly, congruency of the collocations with the participants' L1 appeared to have positively affected their collocational knowledge. The result lends support to the well evidenced findings of the studies which indicated the influence of congruency with L1 on the L2 learners' collocational knowledge and development (e.g., Bandpay, 2012; Farghal & Obiedant, 1995; Gitsaki, 1996; Kellerman, 1978; Koya, 2005; Nesselhauf, 2003; Shehata, 2008; Schmitt, Dörnyei, Adolphs, & Durow, 2004; Wolter & Gyllstad, 2011).

In sum, the feature of congruency with L1 appeared to have a little influence on the participants' collocational knowledge and development. However, this impact seemed to be less influential than frequency and syntactic structure but more influential than semantic transparency due to stability and consistency of its results compared with the results of the semantic transparency feature influence on the participants' collocational knowledge and development.

6.6 The ordinal influence of the collocational features on the participants' collocational knowledge and development

The results of data analyses showed that the four features of collocations which were examined in this study varied in their degrees of influence on the participants' scores on the items, accordingly, their knowledge and development. The most powerful predictor for the participants' knowledge and development seemed to be the frequency of occurrence of collocations in the participants' curricula textbooks.

The syntactic structure of collocations came second in influence, according to the results. However, the significant influence of this feature fluctuated and seemed to be associated with one group of items rather than with both groups. Namely, the influence was significant with the items of the receptive test 1 and productive test 2 which had the same items though in different test designs. The influence of congruency with L1 came third in its influence on the participants' collocational knowledge and development. Finally, the influence of the semantic transparency came last since it showed the least impact on the participants' collocational knowledge and development. Moreover, it demonstrated the least consistency in its influence results in the conducted tests receptively and productively.

Developmentally, the general findings of the current study indicated a developmental pattern in the participants' collocational knowledge. The results indicated that participants' receptive collocational knowledge stabilised at this stage of knowledge development while their productive knowledge seemed to develop significantly over the academic year.

6.7 Participants' collocational knowledge and development

In general, the study results (see Table 5.21), indicated that participants had the command of 63.5% of the collocations they had encountered in their tests which were elicited from their school textbooks. Participants' knowledge of collocations was evidenced by an average of 73.4% of correct answers in their receptive tests and 53.6% of correct answers in their productive tests. From my experience as a teacher, which is more than 20 years, this level of students' English language knowledge is widespread, though not satisfactory, in the local pedagogical settings, i.e., the local schools in Kurdistan. Pedagogically, this level of participants' collocational knowledge implies working on developing students' receptive knowledge of collocations and much more working on developing their low productive knowledge. Accordingly, some procedures can be suggested to deal with this low language proficiency in general and low collocational knowledge in particular such as developing the inputs of the curriculum textbooks, more explicit teaching of collocations in general and special focus on developing students' productive knowledge.

However, these results are not disappointing, taking their intermediate level of education into account, and it may defy the claim that non-natives even at high levels of L2 learning proficiency do not know or use much formulaic language, such as collocations (e.g., Kjellmer, 1990; Wray, 2002). On the contrary, it seems that L2 learners even at this rather low level of proficiency, low proficiency as compared to university students' level, do know a considerable number of formulaic units such as collocations. In sum, participants seemed to have the mastery of using about 63.5% of the input collocations they had studied from their textbooks. This amount of knowledge can be seen reasonable if lack of sufficient input and learning in a second language learning context are taken into considerations.

Developmentally (see Table 5.22), the results indicated that participants' collocational knowledge had developed over the school year (total knowledge at

Time 1 $M= 70.89$, i.e., 62.2% knowledge; Time 2 $M= 73.93$, i.e., 64.8% knowledge, general knowledge $M= 72.4$, i.e., 63.5%). Accordingly, the total collocational knowledge seemed to have developed significantly through the school year. However, the amount in the knowledge development appeared to be little and slowly obtained over the school year (gain $M= 3.04$, i.e., 2.7% knowledge development, $d= .21$). Normally, increase in L2 learners' knowledge during the study course is expected in educational settings because it is the aim of the study. In other words, learners' knowledge is expected to develop after studying for a while when the aim of taking a course is to develop learners' level of certain aspects of knowledge. Otherwise, if knowledge does not develop or decreases, the factors that lie behind it should be investigated and identified, and find appropriate solutions to rectify the situation and achieve the course aims. However, the small and slow development in the participants' collocational knowledge should be expected in this situation due to the difficulty of developing this aspect of knowledge and the short time span of the study, which consisted of several months of non-intensive English learning course among studying many other subjects at school (9 to 12 subjects) in an EFL context. The small knowledge development reflects the difficulty of collocations for L2 learners. The small and slow increase in the knowledge development of this aspect of language could be ascribed, as was discussed in Chapter 2, to the difficulty of collocations and the challenge they create to L2 learners at all levels of language proficiency. Additionally, lack of sufficient explicit teaching of these items can be another reason behind this slow development of collocational knowledge.

Accordingly, the results of the current study support the findings of some other research in this respect which confirm that language knowledge, in general, and collocational knowledge, in particular, is one of the daunting, challenging, and slow developing aspects for L2 learners which continues even at high levels of L2 learners' proficiency (e.g., Ahmadian & Darabi, 2012; Altenberg & Granger, 2001; Alharthi, 2014; Arnaud & Savignon, 1997; Farrokh, 2012; Lindstromberg, & Eyckmans, 2014; Hoey, 2005; Laufer & Waldman 2011; Nesselhauf, 2005). According to research conducted on comparing L2 learners' use of collocations to that of the natives', the former quantitatively and qualitatively use collocations less than the latter (e.g., González Fernández & Schmitt, 2015). Consequently,

due to the great challenge collocations create to L2 learners, participants' knowledge of this aspect of language witnessed only a little amount of knowledge development in the current study. A further anticipated reason could be attributed to shortcomings in the adopted pedagogy in teaching the curriculum. For instance, the emphasis on these collocations might have been insufficient to implant them deep in the participants' memory through recurrent uses of those items during their school years or raising students' awareness of the existence of collocations. Moreover, collocations are not presented under the term "collocations" but rather as matching activities which does not raise students' awareness explicitly of this linguistic phenomenon. Additionally, collocations are usually presented in the inputs as homework activities in secondary textbooks which may not receive as much attention as the students' main textbooks, i.e. students' book inside the class. Accordingly, these shortcomings may suggest pedagogical implications to the educators in the local setting.

Dimensionally, participants' receptive and productive knowledge of collocations did not seem to have developed at the same rate. Receptively, the results indicated that participants had a considerable amount of collocational knowledge (Time 1: $M= 42.24$, 74.1% knowledge; Time 2: $M= 41.45$, 72.7% knowledge; general receptive knowledge, $M= 41.84$, i.e., 73.4% knowledge). However, developmentally, the results did not reveal a significant change in the participants' receptive knowledge (gain $M= -.79$, -1.4% knowledge decrease). Although the decrease in the participants' receptive knowledge of collocations was not statistically significant, the slight statistically significant declining in their knowledge of some aspects of the receptive knowledge 2 as in knowledge of lower frequency collocations, receptive lexical collocations, receptive grammatical collocations, receptive transparent collocations and receptive incongruent collocations, could be attributed to many reasons.

One of the reasons could be participants' lack of motivation in doing the test, in general, and the receptive test, in particular. This claim is retrieved from my observations on the participants' receptive test papers, some participants looked less motivated in carefully answering the receptive test items than the receptive test 1 and the productive test 2. This could be attributed partly to the nature of the test itself which required the participants simply to sign either true

or false next to the items. Moreover, in the receptive tests participants were not restricted by the contextual sentences as the case with the productive test items; in the productive tests, participants were restricted by the context of the sentences to fill the gaps whereas this context did not exist in the receptive tests in which the only restriction was the node of the missing collocate. Although this influence would apply to the productive as well, the sentential context can add more restrictions to the general context, accordingly, minimises number of the potential answers in the test-takers' mind. Consequently, the test did not require as much conscious thinking about the correct answers as the productive items did, and participants could choose either options less consciously than in the productive test.

Another reason that could be thought of was that one school year might have not been sufficient for this part of the collocational knowledge, receptive knowledge, to develop. However, this assumption seems weak due to the well-established notion that receptive knowledge exceeds and/ or precedes learners' productive knowledge of collocations (e.g., Milton, 2009; Schmitt, Dörnyei, Adolphs, & Durow, 2004). Accordingly, if the participants' productive knowledge had developed, their receptive knowledge should have developed earlier or at least stabilised but not decreased.

A further reason that was thought to be behind participants' knowledge declining could be returned to that the reliability of the receptive tests in both waves (receptive 1= .75; receptive 2= .76) was lower than that of the productive ones (productive 1= .93; productive 2= .95) which could be regarded as another shortcoming in the receptive test settings. This assumption lends support to the notion that the design of the instrument may affect the elicited knowledge from participants of the study (e.g., Zhong, 2016). This should not imply that only the difference in reliability had significantly caused the knowledge drop. Rather, I would argue that its effect with the other factors mentioned in this section collectively might have resulted in the knowledge declination.

An additional reason could be attributed to specific features of the collocation items such as frequency, syntactic structure, semantic transparency, or L1 congruency which could restrict knowledge development within the receptive context more than the productive environment. This assumption could be supported by the fact that the items on which participants' receptive knowledge

decreased were mainly of low frequency and incongruent. The mean differences of the phrasal verbs were statistically significant as the mean decreased from (M=12.55) down to (M=11.59). Features such as frequency, syntactic structure, and congruency with L1 were found to be the most powerful anticipated reasons behind participants' receptive knowledge declining. Accordingly, these features may be more influential in the receptive than the productive contexts.

Another reason could simply be that items of the receptive test 1 (productive 2) were more difficult than items of the receptive test 2 (productive 1) which may also explain why the productive test 2 was higher than productive test 1. This assumption is supported by that the mean score of (receptive 1+productive 2), M= 74.72, was significantly higher than the mean score of (receptive 2+productive 1), M= 70.10, but with a small effect size ($d= .29$). However, what weakens this supposition is that the difference between the mean score of the receptive test 1 (M= 42.24) and receptive test 2 (M= 41.45) was not statistically significant whereas the difference in participants' productive knowledge of the first (M= 28.65) and the second test (M= 32.48) was statistically significant with a small effect size ($d= .31$) (see Table 5.21). The difference in difficulty between the first and second receptive items, if the claim could be proved, would be considered as another weak point in the receptive test setting.

A further justification, the comparisons were made between two tests of different items. Although the general features and characteristics of the items of the receptive test 1 and test 2 were balanced, the comparison was made between two different groups of collocations. Accordingly, participants may vary in their knowledge of these two groups due to various factors such as frequency of the items in the participants' textbooks.

Finally, even if participants' receptive collocational knowledge 2 was significantly less than that of receptive 1, it did not necessarily mean that participants' receptive knowledge had decreased through the school year. The case might have only been that learners' knowledge of the test 2 items was less than their knowledge of the test 1 items. As participants' collocational knowledge does not develop all the way at the same speed rate, participants' receptive collocational knowledge development could have been at the time of the receptive test 1 higher than that of the receptive test 2 time. However, this

assumption could be criticised for that the study period could be too short to allow identifying such a decrease in knowledge. Moreover, the learning context usually enhances rather than hampers learners' knowledge development.

In sum, the difficulty of accurately and decisively identifying the reasons or factors behind increasing or decreasing learners' receptive knowledge comes from the complex nature of language learning, collocations in particular, the challenges collocations create ahead of L2 learners, and the various complex network of elements and factors that contribute to the process of collocational knowledge and development. However, what I go for most among all the mentioned reasons is that participants' knowledge of the receptive test 1 items was more, i.e., more familiar with the items, than their knowledge of the items of the second receptive test. Accordingly, it happened unintentionally that more unfamiliar items occurred in the receptive test 2 list.

Productively, the results indicated that participants had a fair amount of collocational knowledge (Time 1 $M= 28.65$, 50.3% knowledge); Time 2 $M= 32.48$, 56.9% knowledge), general productive knowledge ($M= 30.57$, 53.6% knowledge). Additionally, the results revealed a significant increase in the participants' productive knowledge of collocations (gain $M= 3.83$, 6.7%), but the amount knowledge development was small ($d= .31$). However, comparing participants' receptive knowledge with their productive knowledge of collocations (see Table 5.22) showed obviously that participants' productive knowledge lagged much behind their receptive knowledge. The participants' productive knowledge comprised 73.1% of their receptive knowledge, i.e., the productive knowledge of the participants lagged 26.9% behind their receptive knowledge. However, the results indicated that participants' productive knowledge (gain $M= 3.83$, 6.7% knowledge increase) developed at a faster rate than that of the receptive knowledge (gain $M= -.79$, -1.4% knowledge decrease).

Although participants' productive knowledge level was lower than their receptive one, their productive knowledge seemed to have developed significantly more than their receptive knowledge during the school year under study. Additionally, the difference between the mean gain scores of the receptive and the productive knowledge (mean gain scores difference = -4.62) was significant with a rather small effect size ($d= .45$). This means that the amount of knowledge

development in the participants' productive knowledge was significantly larger than that of their receptive knowledge. The productive knowledge mean scores increased significantly from (M= 28.65) in the first wave into (M= 32.48) in the second wave (gain score= 3.83). In contrast, participants' receptive knowledge did not show a significant development over the school year (gain score= -.79); the mean score of the receptive knowledge changed from (M= 42.24) in the first wave to (M=41.45) in the second wave though the difference was not statistically significant. The results lend support to studies such as (e.g., Alsakran, 2011; Laufer, 1998; Milton, 2009; Torabian, Maros & Subakir, 2014; Webb, 2008). The results indicate that collocations are learned or acquired receptively prior to be able to use them confidently in writing or speaking, i.e., productively. Additionally, it implies that more effort and time need to be allocated for developing L2 learners' productive knowledge than that is dedicated for developing their receptive collocational skills. Accordingly, the theoretical and pedagogical implications of this aspect need to be taken into account when curricula textbooks and L2 course are set or designed.

Finally, at the individual level, analysing the correlation between participants' receptive and productive knowledge indicated that higher levels of participants' receptive knowledge were generally associated with higher levels of their productive knowledge as well ($r = .68$). Additionally, the results revealed a correlation of medium to large strength between participants' receptive and productive knowledge of the same items (receptive 1 & productive 2 $r = .48$; receptive 2 & productive 1 $r = .57$). Accordingly, the results indicated a positive, reasonable relationship between the participants' receptive and productive knowledge of collocations. This result corroborates the study findings which indicate a rather strong correlation between learners' receptive and productive knowledge (e.g., Zhong, 2016). Accordingly, higher levels of receptive knowledge were associated with higher levels of productive knowledge. This implies that a development in the learners' receptive collocational knowledge by a specific percentage results in developing their productive knowledge but with a lower percentage due to the difference in the growth rates of the learners' receptive and productive knowledge. Moreover, the relationship between the receptive and productive knowledge is not a straightforward linear relationship (e.g., Caspi, 2013). Additionally, the transition from the receptive to the

productive level is not fast, abrupt, or predictable (e.g., Laufer, 1998; Schmitt & Meara, 1997). Accordingly, these points need to be taken into account in examining or assessing L2 learners' collocational knowledge and development.

6.8 Influence of the test designs on the data quality

The impact of the test designs on the quality and quantity of the elicited data is evidenced in some studies (e.g., Koya, 2005). In contrast to some other studies such as Gyllstad (2013) which indicated the inappropriateness of the productive construct for assessment procedures, the current study results indicated that the productive knowledge test format appeared to be a better indicator to L2 learners' collocational knowledge and development, and a better tool for exploring the influence of the collocational features on the learners' collocational knowledge and development than the receptive test format. Accordingly, this result lends support to studies with similar findings (e.g., Tahmasebi, Ghaedrahmat, & Haqverdi, 2013). The influence is due to nature of the test, structure, the way knowledge is elicited, and the freedom space given to the examinees in providing knowledge and confining to the test settings.

Chapter seven

Conclusions

The present study came up with a number of findings and implications. Accordingly, this chapter presents a short summary of the study first. Next, the contributions which this study has made to the field will be viewed. Then, the pedagogical implications of the study results will be presented. After that, limitations of the study will be brought to surface to stand on the shortcomings of the current research on the hope to be avoided in future research. Finally, directions and recommendations for further research will be suggested.

7.1 Summary

The present study investigated Kurdish high school students' collocational knowledge and development. Additionally, it explored the influence of four of the collocational features- frequency of occurrence in the input, syntactic structure of collocations, semantic transparency, and congruency of collocations with L1- on the collocational knowledge and development of 252 Kurdish Year 11 high school students. Participants of the study were learning English as a foreign language as a part of their school curriculum subjects. The analyses of the students' knowledge and development were based on their knowledge of the encountered collocations in their curriculum textbooks. Accordingly, participants' knowledge of 114 collocations was measured by two tests: an appropriateness judgement test, COLLMATCH3 test, in which they decided on the appropriateness of certain combinations as collocations, and a gap-filling test in which they completed primed missing elements of collocations in sentential contexts taken from their English textbooks. The earlier test was used to measure their receptive knowledge of collocations whereas the latter was used to measure their productive knowledge of collocations. Evidence for the influence of gender on the participants' knowledge and development was also sought in this study.

The study results revealed that participants' total collocational knowledge developed significantly over the school year. However, this development was not equal for both dimensions, receptive vs productive, of collocational knowledge. Receptively, participants' knowledge of collocations did not show a significant development throughout the school year. Conversely, participants' productive knowledge of collocations increased significantly over the same

period. Although participants' productive knowledge lagged much behind their receptive knowledge, productive knowledge increased significantly and the amount of knowledge development was significantly larger than their receptive knowledge development, i.e., participants' productive knowledge developed at a faster rate than that of the receptive knowledge at this level of proficiency. Additionally, analysing the correlation between participants' receptive and productive knowledge of collocations indicated that higher levels of receptive knowledge were associated with higher levels of productive knowledge.

As for the influence of gender differences on the participants' knowledge of collocations, the results generally revealed significant differences between males' and females' collocational knowledge; males achieved higher than females. However, the results revealed no significant difference between the amount of knowledge development of males and females neither in their receptive nor in their productive knowledge of collocations over the school year.

Concerning the influence of frequency on the participants' collocational knowledge and development, the results revealed a significant, positive, moderate relationship between frequency of the collocations in the curricula textbooks and participants' knowledge of those items; higher levels of knowledge of collocations were associated with collocations of higher frequencies.

As for the influence of the syntactic structure on the participants' collocational knowledge and development, participants varied in their collocational knowledge and development of the four syntactic structures. Generally, the results indicated that the syntactic structure of collocations had a statistical significant influence on the participants' collocational knowledge and development, specifically, on their productive knowledge development. According to the results, grammatical collocations seemed to be less challenging than lexical collocations. However, the influence of this feature was less evidenced than the influence of frequency due to inconsistency of the receptive and productive results.

In terms of collocations congruency with participants' L1, congruency seemed to have a little influence on the participants' collocational knowledge and development. Semantically, the results revealed that the semantic transparency of collocations had the least influence on the participants' collocational

knowledge and development.

Finally, the study results demonstrated an ordinal sequence of the influence of the four features of collocations on the participants' knowledge and development in which frequency seemed to be the most influential feature, next was syntactic structure, then congruency with L1, and finally semantic transparency appeared to be the least influential factor.

7.2 Contributions of the study

Contributions of the present study can be perceived from various perspectives. Starting from the definition of the term collocation, it was defined from a hybrid perspective, i.e., combining the frequency-based and phraseological definitions of collocations in one combined view. The adopted definition was based on that for a combination to be considered a collocation, certain criteria should be applied to the combination. First, it is a combination that should occur more than once, but the occurrence is not by chance or misuse by native speakers. Second, it is of specific syntactic structure and restricted combinability. Third, it should be meaningful and used according to the intuition of the native speakers. This definition is a contribution to the third rather new establishing view of collocation. The definition does not rely on frequency of occurrence alone and neglect the syntactic relations and combinability of the collocational items, as with the frequency-based definition, nor does it consider the syntactic relations and combinability but neglect frequency, as with the phraseological definition of collocations. It combines both views for a better identification of this linguistic phenomenon, i.e., collocations. Accordingly, this contribution can be of use in the theoretical and practical aspects of studying collocations. I do not claim that the current research is the first study which adopts the hybrid definition, but it is one of the few studies, though increasing by time, which have appeared recently to establish the third approach to collocations. Theoretically, the hybrid definition may assist in establishing a more comprehensive and less controversial notion of collocations. Practically, it can help in increasing the sophistication of the process of collocation extraction from large corpora by avoiding the shortcomings of each approach if they are employed separately such as yielding syntactically inappropriate collocations or ignoring salient but low frequency collocations in the extraction process.

Empirically, the study employed various criteria, strategies, and instruments to

investigate learners' knowledge receptively and productively. For example, the test items were exclusively based on the items which were actually encountered by learners over their school years. This was different from many other studies in which collocations were selected or elicited from dictionaries or corpora of which some of them have never been experienced by learners. To support this point more, the sentential context of the productive test was completely based on those existed in their curriculum textbooks. Additionally, the validity and reliability criteria were taken into consideration in setting the scales. This contribution can be of significance specifically in pedagogical contexts. Teachers and educational researchers may benefit from adopting such criteria and strategies for better evaluating and assessing students' knowledge development and language proficiency. Additionally, it may provide a good tool of evaluating teaching methodologies and curricular content efficiency for teachers, educators, and curricula textbooks designers.

A further contribution was the longitudinal design of the study. The current study contributed to the field as there have been rather a few studies adopting longitudinal studies in this field. Previous studies have mostly relied on one-shot data collection designs. Since learners' collocational knowledge develops slowly over time, studying this aspect of knowledge requires long time spans to observe any real changes in learners' collocational knowledge. Though one school year may not be sufficient to observe development in some aspects of collocations, it can be enough to realise development in many other aspects of learners' collocational knowledge and development. The current study results indicated the importance of longitudinal study design in identifying knowledge change in some aspects of collocations which could not be diagnosed or identified by adopting one-shot data studies. Accordingly, I would suggest a study period that lasts from three to six years and being conducted on large numbers of students to identify a more comprehensive, considerable knowledge change, and provide a better generalisability power to the results specifically in EFL contexts.

Another contribution was the modifications which were made to the original version of COLLMATCH 3 in terms of number of the items, features, their equal numbers, and way of answering the tests by participants. As for the gap-filling tests, the contextual sentences were exclusively elicited from learners'

textbooks. The adopted strategy paves the way for examining test takers' knowledge of what they have really encountered in their study courses. This can have designative and experimental significance. Tests need to be modified according to the study aims and questions. Accordingly, the contribution can be beneficial for test designers and educators in pedagogical and research contexts.

Choosing participants of the study was another contribution. Majority of the previous studies were conducted on university-level learners or learners from language learning centres at advanced levels of language proficiency. However, the current study was carried out on high school students who were of low to intermediate levels of language proficiency. Accordingly, the study design, testing formats, analytical tools, and results can be useful in investigations at this level of education due to the rather small number of such studies compared with those carried out on the advanced L2 learners.

In terms of the findings, the study suggested a sequential pattern of the influence of the collocational features on the L2 learners' knowledge and development. The results revealed that the four features of collocations, frequency of occurrence, syntactic structure, semantic transparency, and congruency with L1 had impacts on the L2 learners' collocational knowledge and development. However, frequency of occurrence played the most influential role in learners' knowledge development. Additionally, there appeared to be a sequence in their influence significance in which frequency came first, syntactic structure came second, congruency with L1 came third whereas semantic transparency demonstrated the least influence on the learners' knowledge development. This contribution can have important pedagogical implications. Accordingly, the curricula textbooks designers, teachers, and educators should highlight the most challenging aspects of learning collocations and language in a way that facilitates the process of learning.

In contrast to many previous studies, the study results generally indicated a gender influence on the L2 learners' receptive and productive knowledge of collocations; males did better than females in mostly all the tests. However, as the study implies that this difference between both genders seems to be context-restricted rather than being genetic, it necessitates locally-restricted procedures and implications to deal with this imbalanced learning situation.

Accordingly, this result raises the significance of the context of learning on the knowledge development of both genders in general and the local context in particular.

A further contribution is that most of the previous studies generally indicated that the grammatical collocations were more challenging than the lexical ones. In contrast, in the current study, grammatical collocations of simple syntactic structures were generally less challenging than the lexical ones. Additionally, the study results indicated that the difficulty of the syntactic structure could be overcome by concentrating on the other features of collocations such as frequency. For example, learners can acquire collocations of difficult syntactic structures by increasing their exposure to those items. This contribution has its implications and importance mainly in pedagogical contexts. Teachers could make use of the other features of collocations in their teaching to overcome a challenge created by a specific collocational feature.

Finally, the last contribution was eliciting a model summary of the collocational knowledge and development from reviewing the existing literature. The model can be employed as a basis for developing and establishing a better theoretical notion for L2 learners' collocational knowledge and development than can be fruitful in theory and practice.

7.3 Pedagogical implications

The results can have its significance and pedagogical implications in the local context in terms of curricula textbooks settings and educational programmes. Accordingly, they can help in the sophistication of the process of teaching and learning English as a foreign language, particularly, in the local context. Accordingly, the importance of the study results and their pedagogical implications can be demonstrated in various aspects. First, locally, i.e., Kurdistan Region, it is, to the best of my knowledge, the first study that deals with collocations and its influence on the learners' knowledge and development in this breadth and depth. Additionally, it is the first study of the English course textbooks of 11 years, Year 1 to Year 11, which explores learners' knowledge of this aspect of English language, and examines the factors which may affect students' collocational knowledge and development. Second, the majority of the previous studies which were conducted in the local setting were in the form of questionnaires in which participants expressed their attitudes towards the

curriculum textbooks. This is the first study in which participants took systematic knowledge exploration tests based on their textbooks. Third, it draws teachers' attention, who seem to be unaware or have neglected collocations, to this aspect of language. It was also noticed that participants were explicitly unaware of this phenomenon as a term, though they were practically familiar with it in the form of matching tests. Accordingly, this implies that teachers are recommended to draw students' attention to this aspect of language and its significance in developing their language proficiency.

The results also indicated students' low productive knowledge of collocations. This necessitates that teachers should pay more attention to developing their collocational knowledge and use of collocations. This can be achieved through increasing productive assignments such as writing short essays, encouraging conversations inside classrooms, acting short plays, and making discussions from time to time about topics that interest students with providing clues and useful collocations on the boards to encourage and motivate learners to use the language.

Additionally, since the influence of the gender factor was evidenced in the current study, females' underachievement compared to males', and the belief that this shortage in achievement may not be genetic but circumstantial, this can have social, political, economic, and pedagogical implications. Socially, females' self-esteem could be raised through raising people's awareness of equity principle between males and females and females' important role and contribution in the different aspects of daily life, specifically in villages where females are under higher social, cultural, and traditional pressures than those live in cities and town. Additionally, raising parents' awareness of the necessity of a fair treatment of their male and female children and giving them equal attention and support specifically during their early years of life. This may increase females' self-confidence and future expectations, accordingly, their motivation to learn. Politically, issuing official regulations and instructions that guarantee females' equity of chance with males in terms of education, work, legal, and political status. Economically, having equal opportunities with males to work and earn money can motivate females to have higher expectations for themselves, accordingly, higher desire to learn and promote their academic knowledge in the future. Educationally, including female-related subjects in the

curricula inputs that shed light on females' rights and their role in society. This emphasis on female students could be more effective and required specifically at the early stages of learning, specially Year 1 to Year 9, where they may need more encouragement to overcome the potential negative influence of the contextual factors on their self-esteem, motivation, education, and knowledge development.

Furthermore, due to the evidenced influence of L1 on the participants' collocational knowledge and development, teachers are advised to raise learners' awareness to the L1-L2 differences whenever they experience collocations which have no L1 equivalents. Additionally, the results indicated that some patterns or types were more challenging than others such as lexical patterns vs grammatical patterns or that the adj+n pattern was more challenging than the others. Accordingly, more emphasis is required on such patterns and types of collocations.

In sum, this study is hoped to be of benefit to the local Ministry of Education in setting future English textbooks. Its usefulness could be in various ways. First, the results imply raising learners' awareness of collocations. Accordingly, educators in the ministry need to highlight the significance of collocations in designing future new English textbooks for students. This can be through the explicit content of collocations in the textbooks and providing teachers with explicit guidelines for teaching collocations. The study implies also a shortage in the input. Accordingly, increasing the inputs with specific emphasis on collocations can be fruitful. It is also hoped to be of help to teachers of English language in identifying their students' strength and weakness areas, setting tests, and adapting their teaching plans, accordingly, increasing their collocational knowledge and language proficiency.

7.4 Limitations of the study

The present research was not without issues and limitations. Among the issues was participant attrition. Some participants left the study while others joined, as explained in the methodology chapter. Another issue was the difficulty of supervising all the participants alone without the school teachers' help to provide a healthy, appropriate environment for the participants at the time of conducting the tests due to the large number of participants and unavailability of a large hall that could have all participants in one place under my direct

supervision. A further issue was unavailability of sufficient numbers of items with the required features in the textbooks which resulted in using some items which were not of strongly fitting the criteria set for the constructs, consequently, being deleted at the reliability stage. As a result, having the exact number of items for each feature became difficult and the balance between items according to the required features changed a little.

Another limitation was that items of the receptive test 1 were different from that of the test 2, and items of the productive test 1 were different from those of the test 2. To avoid the influence of repetitions on the participants, testing them on the same items receptively or productively was initially avoided. Consequently, the difference in the test items did not allow following knowledge development of the same items during the school year within the same knowledge dimension, i.e., receptively or productively. Rather, the comparisons were made between the knowledge development of two different sets of items. Moreover, this might be one of the reasons behind being participants' knowledge of receptive 2 less than their knowledge of receptive 1. Consequently, the results did not show a development in the receptive knowledge but a decline in knowledge of some aspects of collocations which was not the expected result in such situation, i.e., knowledge development over the school year.

Another limitation was that examining participants' receptive and productive knowledge at the same time required eliciting and including a large number of items in the tests which put pressure on the tests setting and participants at the time of taking the tests. Dealing with one aspect of knowledge could have reduced the pressure on the whole study specifically in terms of number of items which meet the applied features.

Furthermore, the COLLMATCH3 and gap-filling tests which were employed for this study contained equal numbers of items in terms of their frequency, the four syntactic patterns, semantic transparency, and congruency with L1. However, the balance among these aspects could not be completely maintained due to subjecting these items to reliability criteria. Preserving balance among the different aspects of the tested items could have provided more consistency in the results. Furthermore, due to the practical difficulty of achieving this aim, some of the items might have not been quite strong in relation to the features on which they were tested specifically with transparency and congruency features.

Consequently, this might have resulted in part of the inconsistency of the test results on these aspects.

Another limitation of the study was removing some items from the collocational knowledge tests which seemed to have negative influences on the reliability of the tests. Although the rationale behind removing the items was to increase the reliability of the tests, the removal might have had negatively affected the study findings. The removal might have resulted in losing some significant data that could have had significant pedagogical implications such as removing the most and the least recurrent items in the learners' curriculum textbooks on which learners' scored the highest marks on the earlier and the lowest on the latter.

Finally, the potential incompatibility between the two compared sets of items, receptively and productively, in terms of frequency of the items in the textbooks and learners' knowledge of those items can be regarded as another limitation of the study. As guaranteeing compatibility between both sets of items before administering the tests, specifically in terms of their frequency and knowledge, was not practically feasible, the nonsignificant knowledge development in the learners' receptive knowledge, the slow development in their productive and general knowledge of collocations over the school year could be attributed not only to the influence of poor instructional quality, insufficient input, and EFL context, but also to the knowledge measurement tools I used in the study.

7.5 Directions for further research

The study adopted a hybrid approach definition of collocations which is a rather new approach that combines the frequency-based and phraseological approaches in one design. Due to the complex nature of collocations and learners' slow knowledge development at this level of learning, I encourage conducting future research adopting the same definition and design but for longer periods that range between three to six years. On the one hand, the hybrid approach can involve the key characteristics of collocations such as frequency, syntactic relations, semantics, and combinability of the component items, accordingly, yielding more comprehensive view of collocations. On the other hand, as learners' collocational knowledge develops slowly over time, longitudinal studies of longer time spans may allow more chance to identify change and observe nature and development of learners' collocational knowledge. This type of research may provide deeper, more comprehensive,

and more informative insights into L2 learners' collocational knowledge and development such as exploring the influence of features of collocations and learning context on learners' collocational knowledge and development. The present longitudinal study lasted for only one school year. Extending future studies for three years or more in the same local context may provide better explorations and depiction of the knowledge development patterns.

Additionally, the study employed appropriateness judgement test and gap-filling tests. However, incorporating other types of tests such as essay writing and translation tests in future studies, in addition to the tests utilised in this study, may assist in shedding light on some other aspects of learners' knowledge and development.

As participants' performance on the receptive tests was not consistent as it was on the productive tests, extending the data collection period to more than one year may provide more consistent and stable results. In terms of participants' level of proficiency, more studies are required at the high school level as most studies have been conducted on learners of higher proficiency levels, mostly university students.

In addition, although the current study did not aim at evaluating the school textbooks, the results revealed that such study designs with some modifications can be used in future studies as useful tools for assessing learners' performance and curricula textbooks contents evaluations. Accordingly, results of such studies can be of significant pedagogical implications in teaching and curriculum settings.

Furthermore, the study results of the individual test analyses revealed that males achieved higher than females at this proficiency level. These results give chance for future research to verify this point and to investigate the potentiality of existing a real difference between males' and females' collocational knowledge and development, and exploring the reasons behind the difference in more detail. This can help in identifying and dealing with any social, economic, political, cultural, educational, or even genetic factors which can lie behind this difference in knowledge development.

Furthermore, from the pilot tests conducted on native speakers, it was noticed that even native speakers of English made some mistakes in filling the gap-

filling tests. The mistakes were not in grammar but in mismatching. Future research is required to explore native collocational knowledge development at the same level, Year 11, to verify whether the mistakes were due to lacking clarity in the contextual sentences or lack of collocational knowledge of native speakers at this stage of language knowledge.

Most of the studies have investigated mainly academic collocations. Exploring collocational knowledge and development in natural daily life and non-academic contexts will be worthwhile which can give chance for more investigations, accordingly, knowledge about this linguistic phenomenon. A further area of investigation is that the study explored the grammatical collocations of only simple structures which encourages exploring the influence of grammatical collocations of more complex structures as well.

Finally, testing learners' collocational knowledge and development is still in need of more sophisticated evaluating and assessing tools specifically in studies conducted from dynamic perspectives which lack effective tools in this respect. Accordingly, designing new tools or developing the current ones to further explore learners' collocational knowledge from static or dynamic perspectives are recommended.

The above suggested studies are some ideas for future research, but they are not exhaustive as many other ideas can be built on the present study results. There are still many other aspects of learners' collocational knowledge which need to be investigated at this level of language proficiency or lower, i.e., secondary school level.

Appendices

Appendix (A): Ethical Approval Certificate and Consent Form

MSc, PhD, EdD & DEdPsych theses.



Graduate School of Education

Certificate of ethical research approval

PhD thesis

To activate this certificate you need to first sign it yourself, and then have it signed by your supervisor and finally by the Chair of the School's Ethics Committee.

For further information on ethical educational research access the guidelines on the BERA web site: <http://www.bera.ac.uk/publications> and view the School's Policy online.

READ THIS FORM CAREFULLY AND THEN COMPLETE IT ON YOUR COMPUTER (the form will expand to contain the text you enter). **DO NOT COMPLETE BY HAND**

Your name: Jamal Muhammad Ramadhan

Your student no: 620033032

Return address for this certificate: Top floor flat 92 Sidwell Sreet, Exeter, Devon, EX4 6PH

Degree/Programme of Study: 4 year PhD programme

Project Supervisor(s): Dr. Philip Durrant and Dr. Esmael Abdollahzadeh

Your email address: jmr218@exeter.ac.uk or jamalbotani2@yahoo.com

Tel: UK/ 07907275137 or Kurdistan/ 009647508269595

I hereby certify that I will abide by the details given overleaf and that I undertake in my thesis to respect the dignity and privacy of those participating in this research.

I confirm that if my research should change radically, I will complete a further form.


Signed:...Jamal Muhammad Ramadhan **date:** 25/02/2015

Chair of the School's Ethics Committee
updated: March 2013

Certificate of ethical research approval

TITLE OF YOUR PROJECT:

A Longitudinal Study of the Development of Kurdish Students' English Collocational Knowledge

1. Brief description of your research project:

The study examines the development of collocational knowledge of Kurdish learners who are learning English as a foreign language. The study attempts to identify how features of collocations affect or relate to learners' knowledge of collocation and how these features are connected with each other and it attempts to explain the pattern of learners' knowledge development from the Dynamic Systems Theory perspective.

2. Give details of the participants in this research (giving ages of any children and/or young people involved):

The study is conducted on about 200 high school Kurdish students, Year 11, whose age ranges between 17-18. Participants are of both sexes, male and female; 100 male and 100 female chosen arbitrarily to give a chance to any student likes to participate or withdraw from the study. Since the researcher will not be teaching the participants at the time of conducting the study, a researcher-participant relationship will be established during the data collection period.

Give details (with special reference to any children or those with special needs) regarding the ethical issues of:

3. informed consent: Where children in schools are involved this includes both headteachers and parents). **Copy(ies) of your consent form(s) you will be using must accompany this document.** a blank consent form can be downloaded from the GSE student access on-line documents: **Each consent form MUST be personalised with your contact details.**

All participants will be informed of the study aims, the tests taken, what is expected from them to do and the use of their data and the benefits anticipated from the study results to learning English in general and students in particular. Informed consent forms will be given to them to sign with their utmost freedom to voluntarily participate in the study. Additionally, the form will be written in Kurdish, and the English translation of the form is provided with this application. In case of student with special needs, specific arrangements will be followed and adjusted according needs and degree of impairment. These procedures will be explained in more detail in part 7.

Concerning securing permission from the sample schools to test the student, I took permission from the local authority, General Directorate of Education-Duhok, who were very cooperative and granted me permission to collect data from the assigned schools, Bijareh High School and Midya High School. After being authorized by the General Directorate of Education, I went to these schools and met both head teachers and showed them the permission letter. On their parts, they welcomed my study and demonstrated their readiness to cooperate with in collecting data from their students next academic year.

4. anonymity and confidentiality

Chair of the School's Ethics Committee
updated: March 2013

It will be assured to the participants that their names, data and personal information will be kept safe, confidential and anonymous and since the study data are quantitative test results, the data will be presented through summary statistics, and no reference needs to be made to individuals. Additionally, it will be certain to them that these data will be used for research purposes only and they will not be used by another party without their prior approval.

5. Give details of the methods to be used for data collection and analysis and how you would ensure they do not cause any harm, detriment or unreasonable stress:

Data collection will be carried out by doing pen-and-paper tests in a comfortable and safe place without practising any physical or psychological pressure on the part of teachers, headmasters or any other parties. The tests will also take study loads into consideration through conducting these tests at times where they have not schools exams or assessment and they will be consulted about the most convenient settings for their tests. The tests will take place at the school examination halls in which the participants study. The tests are planned to take place on three sittings; the first is October, 2015, the second in February, 2016 and the third in May, 2016. Participants who decide to withdraw from the study their names and data will be destroyed if they have taken part in earlier tests.

6. Give details of any other ethical issues which may arise from this project - e.g. secure storage of videos/recorded interviews/photos/completed questionnaires, or

Before scoring the test papers, all the data will be stored in a safe closet by the researcher at home. After scoring, the data and test results will be coded and save on the researcher's computer which is in turn protected by a password known only by the researcher.

7. Special arrangements made for participants with special needs etc.

If any participants with special needs are identified, necessary adjustments will be made to the tests to enable them to take part fairly.

8. Give details of any exceptional factors, which may raise ethical issues (e.g. potential political or ideological conflicts which may pose danger or harm to participants):

Participants of the study are all from the same ethnic group, Kurdish, religion and no political or ideological conflicts are expected to pose danger or harm to the participants. Additionally, the study focus, linguistics, helps in keeping the study away from ideological and political conflicts. The research will take place in Duhok/Kurdistan-Iraq. The city is a safe place far from the conflicts and participants are expected to face any threats from any kind. I, as a student, will be staying with some relatives in Duhok and all kinds of contacts with me or them will be available such mobile phones and internet services. Accordingly, I will keep updated on FCO travel advice. To ensure keeping updated, I will keep in weekly contact with my both supervisors to be aware of the latest news regarding the different aspects of the study and the security status of the region and place of the study.

This form should now be printed out, signed by you on the first page and sent to your supervisor to sign. Your supervisor will forward this document to the School's Research Support Office for the Chair of the School's Ethics Committee to countersign. A unique approval reference will be added and this certificate will be returned to you to be included at the back of your dissertation/thesis.

N.B. You should not start the fieldwork part of the project until you have the signature of your supervisor

This project has been approved for the period: 23/09/2013

until: 22/09/2018

Chair of the School's Ethics Committee
updated: March 2013

P.L.D. *

By (above mentioned supervisor's signature): ... date:.....

N.B. To Supervisor: Please ensure that ethical issues are addressed annually in your report and if any changes in the research occur a further form is completed.

GSE unique approval reference:..... D/14/15/42.....

J.K. Cole

Signed:..... date:..05.05.2015.....

Chair of the School's Ethics Committee

Chair of the School's Ethics Committee
updated: March 2013



GRADUATE SCHOOL OF EDUCATION

Title of Research Project: **A Longitudinal Study of the Development of Kurdish Students' English Collocational Knowledge**

Combining the word *fast* with *cars* to produce *fast cars* but not **quick cars* because it is wrong or inappropriate in English use is called collocation. This study aims at examining how high school Kurdish students combine two words correctly and how their knowledge of collocations develops over time.

If you agree to participate in this study, you will be asked to take three waves of tests over the academic year. Each test will take about 60 minutes and will take place at your school on October 15th, 2015, February 15th, 2016 and May 15th, 2016.

You do not have to take part in this study and the results of your tests will not affect your class grades or be revealed to you teachers.

The results of this study are hoped to be useful for teachers and students in learning English collocations to improve students' language ability in Kurdistan.

CONSENT FORM

I have been fully informed about the aims and purposes of the project. I understand that there is no compulsion for me to participate in this research project and, if I do choose to participate, I may at any stage withdraw my participation and may also request that my data be destroyed. Additionally, I have been informed that I have the right to refuse permission for the publication of any information about me and that any information which I give will be used solely for the purposes of this research project, which may include publications or academic conference or seminar presentations. In addition, if applicable, the information, which I give, may be shared between any of the other researcher(s) participating in this project in an anonymised form and all information I give will be treated as confidential and that the researcher(s) will make every effort to preserve my anonymity.

.....
(Printed name of participant)

.....
(Signature of participant)

.....
(Date)

One copy of this form will be kept by the participant; a second copy will be kept by the researcher(s)
Contact phone number of researcher(s): UK/004407907275137 or Kurdistan/009647508269595
If you have any concerns about the project that you would like to discuss, please contact:

The researcher: Jamal Muhammad Ramadhan
Email: jamalbotani2@yahoo.com
UK/004407907275137 or Kurdistan/009647508269595
Or
The project supervisor: Dr Philip Durrant
Email: P.L.Durrant@exeter.ac.uk

* when research takes place in a school, the right to withdraw from the research does NOT usually mean that pupils or students may withdraw from lessons in which the research takes place



GRADUATE SCHOOL OF EDUCATION

تلفون: کوردستان/۰۰۹۶۴۷۵۰۸۲۶۹۰۹۵ - پان بریتانیا/۰۰۴۴۰۹۷۰۷۲۷۵۱۳۷

پان

سرپرستی فکولتی: د. فیلیپ دورانت

نیمیل: P.L.Durrant@exeter.ac.uk

* دیمی فکولتی دهنه نعلامدان ل فوتایختن مانی فکولتی ل فکولتی ر امانا هندی نادت فوتلی تمبندار یوی نهچنه و انیت

خو

Data Protection Act: The University of Exeter is a data collector and is registered with the Office of the Data Protection Commissioner as required to do under the Data Protection Act 1998. The information you provide will be used for research purposes and will be processed in accordance with the University's registration and current data protection legislation. Data will be confidential to the researcher(s) and will not be disclosed to any unauthorised third parties without further agreement by the participant. Reports based on the data will be in anonymised form.

إقليم كوردستان - العراق
مجلس الوزراء
وزارة التربية
المديرية العامة لتربية محافظة دهوك
مديرية تربية قضاء دهوك / الشرقية
قسم التخطيط



هه ریمی کوردستان - عیراق
نه نجومه نی و مزیران
و دزاره نی په رورده
ریشه بهریا گشتی یا په رورده دهوک
ریشه بهریا په رورده قهزا دهوک / رۆژ شهلات
پشکا نه خشه دانان

No:

Date: ٢٠١٥ / ٣ / ١٠

ژماره: ٢٥١٩

ریشه و ت: ٢٣٥ / ١

ب / نو نامادها (٣) ئادار

ب / ئاسنگاری

نفسارا ریشه بهریا گشتی یا په رورده دهوک / ر. نه خشه دانانا په رورده / هوبا نه خشه دانان یا ژماره
(٤١٢٢) ل ٢٠١٥ / ٣ / ١٠، تکایه هاریکاری و ئاسنگاری بو بهریز (جمال محمد رمضان) قوتایی دکتورایی
بهینه کرن ژبو ب. دهسفه ئینانا پیزانینا ژ بو بهرنامی خو بی دکتورایی.

دگهل ریزگرتنی

هه فینج:

نفسارا ریشه بهریا گشتی یا په رورده دهوک / ر. نه خشه دانانا په رورده / هوبا نه خشه دانان

ل سمیر عبدالعزیز حگی

ریشه بهر

٢٠١٥ / ٣ / ١٠

وینهک بو :

- ریشه بهریا گشتی یا په رورده دهوک / ر. نه خشه دانانا په رورده / هوبا نه خشه دانان

- بهکا سهر په رشتیاری و دلنیایی و جوری / روزهلات

- پشکا نه خشه دانان

- باددکا زفروک

Appendix (B): Test items list

No	Collocations	Frequency in COCA	MI	Frequency in textbooks	Collocates frequency in textbooks	Nodes frequency in textbooks	Congruent/Incongruent	Transparent/Opaque	Pattern	Type: Lexical/ Grammatical	Collocation item's appearance in the textbooks		First collocation appearance in the textbooks/ year
											Collocate/ Year	Node/ Year	
1	comb (her) hair	382	5.8	1	2	63	C	T	vn	L	comb 2	hair 2	Y 2
2	Kiss (her) cheek	477	7.0	3	6	3	C	T	vn	L	kiss 8	cheek 8	Y 8
3	ask a question	6820	3.1	19	458	366	C	T	vn	L	ask 1	question 3	Y 3
4	climb a tree	199	3.4	5	44	55	C	T	vn	L	climb 8	tree 5	Y 8
5	develop (his) skills	1348	4.1	2	27	29	C	T	vn	L	develop 9	skill 7	Y 11
6	add sugar	984	3.9	2	53	19	C	T	vn	L	add 3	sugar 6	Y 6
7	drink tea	404	4.3	24	105	34	C	T	vn	L	drink 3	tea 3	Y 6
8	answer a question	8235	4.9	74	457	366	C	T	vn	L	answer 1	question 3	Y 4
9	brush (his) teeth	475	6.0	2	3	9	IC	T	vn	L	brush 5	teeth 2	Y 5
10	pay a debt	918	3.8	3	27	7	IC	T	vn	L	pay 9	debt	Y 10
11	launch a satellite	123	4.5	6	7	21	IC	T	vn	L	launch 10	satellite 8	Y 10
12	fly a plane	376	3.8	6	67	29	IC	T	vn	L	fly 4	plane 3	Y 8
13	heat oil	3172	5.2	2	12	31	IC	T	vn	L	heat 8	oil 9	Y 10
14	lend money	379	3.67	1	7	135	IC	T	vn	L	lend 8	money 5	Y 8
15	play the guitar	479	3.7	27	404	65	IC	T	vn	L	play 1	guitar 3	Y 3
16	play a game	4165	3.0	83	404	192	IC	T	vn	L	play 1	game1	Y 3
17	bend (her) knees	222	5.3	1	7	4	C	O	vn	L	bend 7	knee 7	Y 7
18	attract tourists	114	5.1	2	22	70	C	O	vn	L	attract 8	tourist 5	Y 8
19	ride a bicycle	995	6.1	7	53	35	C	O	vn	L	ride 3	bicycle 4	Y 4
20	cross the river	406	3.1	7	28	69	C	O	vn	L	cross 5	river 5	Y 8
21	pass an exam	170	4.2	1	26	9	C	O	vn	L	pass 8	exam 6	Y 9
22	climb mountains	280	3.9	4	44	99	C	O	vn	L	climb 8	mountain 5	Y 8
23	cook dinner	315	3.2	6	57	58	C	O	vn	L	cook 4	dinner 4	Y 7
24	follow the path	579	3.2	1	78	15	C	O	vn	L	follow 3	path 5	Y 11
25	follow the instruction	463	4.2	1	78	12	IC	O	vn	L	follow 3	instruction 4	Y 4

	s												
26	pay attention	6876	5.0	1	27	2	IC	O	vn	L	pay 9	attention 9	Y 9
27	lose weight	2466	4.4	6	62	9	IC	O	vn	L	lose 9	weight 9	Y 9
28	make a mistake	2834	3.1	11	504	51	IC	O	vn	L	make 3	mistake 6	Y 6
29	shake hands	1198	3.7	1	4	60	IC	O	vn	L	shake 8	hand 1	Y 8
30	borrow money	749	4.5	2	14	135	IC	O	vn	L	borrow 6	money 5	Y 8
31	catch a bus	63	3.7	6	30	48	IC	O	vn	L	catch 7	bus 3	Y 8
32	make a decision	7358	3.7	5	504	80	IC	O	vn	L	make 3	decision 11	Y 11
33	electric motors	465	6.8	1	5	1	C	T	adjn	L	electric 8	motor 11	Y 11
34	active volcanoes	125	6.5	1	427	10	C	T	adjn	L	active 9	volcano 9	Y 9
35	global warming	5209	9.6	19	35	63	C	T	adjn	L	global 9	warming 7	Y 9
36	daily routines	566	5.8	5	8	15	C	T	adjn	L	daily 5	routine 5	Y 5
37	prehistoric monsters	10	5.6	1	3	16	C	T	adjn	L	prehistoric 9	monster 9	Y 9
38	radioactive materials	162	6.0	2	2	7	C	T	adjn	L	radioactive 9	material 6	Y 9
39	my best friend	6450	4.8	7	117	226	C	T	adjn	L	best 7	friend 3	Y 7
40	correct answers	587	3.9	9	177	457	C	T	adjn	L	correct 3	answer 1	Y 7
41	modal verbs	9	10.5	4	5	81	IC	T	adjn	L	modal 11	verb 5	Y 11
42	blonde hair	779	7.0	2	2	63	IC	T	adjn	L	blonde 9	hair 2	Y 9
43	big bangs	1602	6.1	5	171	6	IC	T	adjn	L	big 3	bang 5	Y 5
44	fizzy drinks	12	7.3	7	7	105	IC	T	adj+n	L	fizzy 6	drink 3	Y 6
45	a loud noise	407	6.0	1	13	18	IC	T	adjn	L	loud 7	noise 8	Y 11
46	precious metals	110	5.1	2	7	20	IC	T	adjn	L	precious 9	metal 6	Y 9
47	Native Americans	5955	5.2	5	6	92	IC	T	adjn	L	native 8	American 7	Y 8
48	the main entrance	509	5.5	1	17	8	IC	T	adjn	L	main 8	entrance 9	Y 9
49	the industrial revolution	861	6.9	4	25	4	C	O	adjn	L	industrial 11	revolution 11	Y 11
50	the wild west	928	4.4	2	21	40	C	O	adjn	L	wild 6	west 7	Y 10
51	bad luck	1515	5.0	10	85	45	C	O	adjn	L	bad 5	luck 3	Y 9
52	the average temperature	353	3.7	5	9	31	C	O	adjn	L	average 11	temperature 11	Y 11
53	an open mouth	2590	4.1	1	113	14	C	O	adjn	L	open 5	mouth 4	Y 11

54	weak syllables	22	5.6	2	5	4	C	O	adjn	L	weak 7	syllable 8	Y 11
55	mobile phones	603	5.7	13	15	104	C	O	adjn	L	mobile 3	phone 3	Y 3
56	the same time	359 52	3.2	8	83	396	C	O	adjn	L	same 4	time 3	Y 8
57	alien spaceships	34	7.2	1	9	11	IC	O	adjn	L	alien 7	spaceship 8	Y 8
58	developed countries	205	3.4	1	27	156	IC	O	adjn	L	developed 9	country 5	Y 10
59	happy birthday	151 9	6.6	10	110	85	IC	O	adj+n	L	happy 3	birthday 5	Y 5
60	dark hair	315 1	5.0	7	43	63	IC	O	adjn	L	dark 7	hair 2	Y 7
61	terrible mistakes	408	5.8	1	43	51	IC	O	adjn	L	terrible 7	mistake 6	Y 10
62	a gentle voice	287	4.2	1	8	12	IC	O	adjn	L	gentle 11	voice 7	Y 11
63	good news	108 17	3.4	4	243	50	IC	O	adjn	L	good 1	news 5	Y 9
64	heavy rain	682	5.0	4	21	63	IC	O	adjn	L	heavy 7	rain 3	Y 9
65	throw away	181 3	4.4	4	59	25	C	T	P v	G	away 3	throw 6	Y 6
66	dive into	705	3.7	1	131	4	C	T	P v	G	into 5	dive 9	Y 9
67	sit down	948 7	4.8	29	87	78	C	T	P v	G	down 1	sit 1	Y 1
68	talk about	626 77	4.3	162	892	328	C	T	P v	G	about 3	talk 3	Y 4
69	pull out	558 6	4.3	2	176	8	C	T	P v	G	out 5	pull 8	Y 8
70	pass through	284 7	4.1	3	53	26	C	T	P v	G	through 8	pass 8	Y 9
71	go ahead	144 93	6.0	7	17	793	C	T	P v	G	ahead 9	go 3	Y 9
72	worry about	149 25	5.5	5	892	37	C	T	P v	G	about 3	worry 5	Y 6
73	pull off	301 6	4.1	1	43	8	IC	T	P v	G	off 6	pull 8	Y 10
74	complain about	226 1	3.9	1	892	1	IC	T	P v	G	about 3	complain	Y 11
75	pick up	213 95	4.8	22	279	35	IC	T	P v	G	up 1	pick 4	Y 4
76	look at	106 031	3.2	228	24	666	IC	T	P v	G	at 1	look 1	Y 1
77	divide into	112 5	4.4	1	131	4	IC	T	P v	G	into 5	divide 9	Y 11
78	argue against	431	3.4	3	2	10	IC	T	P v	G	against 10	argue 9	Y 11
79	ought to	242 25	3.4	16	360 9	18	IC	T	P v	G	to 1	ought 11	Y 11
80	go back	260 71	3.6	10	105	793	IC	T	P v	G	back 4	go 3	Y 5
81	calm down	242 4	4.4	2	87	9	C	O	P v	G	down 1	calm 9	Y 9
82	slow down	481 0	4.4	3	87	36	C	O	P v	G	down 1	slow 8	Y 10
83	wake up	837 3	4.7	8	279	8	C	O	P v	G	up 1	wake 8	Y 8
84	stand up	713 4	4.7	24	279	93	C	O	P v	G	up 1	stand 1	Y 1

85	move away	188 1	3.3	1	59	67	C	O	P v	G	away 3	move 4	Y 11
86	take away	638 1	3.3	3	59	210	C	O	P v	G	away 3	take 2	Y 10
87	go on	206 41	2.8	7	972	793	C	O	P v	G	on 3	go 3	Y 7
88	write down	212 3	3.7	5	87	751	C	O	P v	G	down 1	write 2	Y 8
89	cut off	939 4	4.0	4	43	63	IC	O	P v	G	off 6	cut 6	Y 11
90	shut up	551 2	3.6	2	279	3	IC	O	P v	G	up 1	shut 9	Y 9
91	look forward	387 4	3.7	7	9	666	IC	O	P v	G	forward 6	look 1	Y 6
92	clean up	521 0	3.0	4	279	46	IC	O	P v	G	up 1	clean 3	Y 8
93	bump into	466	4.4	1	131	1	IC	O	P v	G	into 5	bump 9	Y 9
94	check out	829 6	4.3	1	176	144	IC	O	P v	G	out 5	check 5	Y 9
95	set up	247 37	4.2	4	279	16	IC	O	P v	G	up 1	set 5	Y 11
96	get up	122 71	2.3	98	279	609	IC	O	P v	G	up 1	get 4	Y 4
97	at least	132 125	4.5	3	24	4	C	T	prep n	G	at 1	least 8	Y 8
98	till tomorrow	134	4.2	1	8	54	C	T	prep n	G	till 10	tomorrow 5	Y 11
99	at (three) o'clock	296 9	3.3	30	24	30	C	T	prep n	G	at 1	o'clock 3	Y 3
100	for example	916 98	3.5	33	853	82	C	T	prep n	G	for 3	example 5	Y 5
101	under control	546 8	4.7	1	40	15	C	T	prep n	G	under 3	control 8	Y 8
102	at crossroads	799	4.3	1	24	9	C	T	prep + n	G	at 1	crossroads 5	Y 5
103	on television	123 15	3.0	33	972	176	C	T	prep n	G	on 3	television 4	Y 5
104	out of control	407 3	5.8	7	176	15	C	T	prep n	G	out 5	control 8	Y 10
105	on duty	187 2	3.1	1	972	1	IC	T	prep n	G	on 3	duty 9	Y 9
106	on foot	242 7	2.5	3	972	24	IC	T	prep n	G	on 3	foot 3	Y 7
107	on a picnic	35	2.2	5	972	54	IC	T	prep n	G	on 3	picnic 5	Y 5
108	on the floor	155 23	4.2	4	972	5	IC	T	prep n	G	on 3	floor 7	Y 7
109	on the screen	602 9	3.8	2	972	6	IC	T	prep n	G	on 3	screen 5	Y 5
110	on the train	166 5	2.1	9	972	67	IC	T	prep n	G	of 3	train 3	Y 10
111	at the end	330 70	4.2	18	24	72	IC	T	prep n	G	at 1	end 3	Y 4
112	on the right	596 9	2.7	16	972	187	IC	T	prep n	G	on 3	right 3	Y 5
113	in fact	788 04	4.7	1	54	37	C	O	prep n	G	in 3	fact 5	Y 5
114	in light (of)	467 7	4.6	1	54	39	C	O	prep n	G	in 3	light 5	Y 11
115	in return	511	2.1	1	54	26	C	O	prep	G	in 3	return 8	Y 9

5		6							n				
11 6	in horror	942	2.7	2	54	21	C	O	prep n	G	in 3	horror 11	Y 11
11 7	in a row	423 3	4.3	1	54	2	C	O	prep n	G	in 3	row 6	Y 6
11 8	in charge	105 75	4.1	6	54	6	C	O	prep n	G	in 3	charge 8	Y 8
11 9	for a while	138 11	4.0	6	853	56	C	O	prep n	G	for 3	while 8	yea r 11
12 0	for a walk	114 2	4.3	8	853	221	C	O	prep n	G	for 3	walk 3	Y 7
12 1	in a circle	277 1	3.1	2	54	19	IC	O	prep n	G	in 3	circle 3	Y 7
12 2	across countries	926 8	7.6	2	36	156	IC	O	prep n	G	across 8	country 5	Y 10
12 3	for ages	124 4	3.0	4	853	33	IC	O	prep n	G	for 3	age 7	Y 9
12 4	around the world	149 58	6.0	11	49	247	IC	O	prep n	G	around 8	world 4	Y 4
12 5	on sale	184 4	3.9	1	972	3	IC	O	prep n	G	on 3	sale 9	Y 9
12 6	by accident	152 9	4.5	1	190	16	IC	O	prep n	G	by 7	accident 6	Y 9
12 7	over the years	205 27	4.0	1	111	329	IC	O	prep n	G	over 4	year 5	Y 10
12 8	of course	109 767	4.7	47	190 6	54	IC	O	prep n	G	of 3	course 5	Y 5

Appendix (C): Classifying collocations

First group				
v+n	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	comb (her) hair	brush (his/her) teeth	bend (her/his) knees	follow the instructions
	Kiss (her) cheek	pay a debt	attract tourists	pay attention
	ask a question	launch a satellite	ride a bicycle	lose weight
	climb a tree	fly a plane	cross the river	make a mistake
adj+n	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	electric motors	modal verbs	(the) industrial revolution	alien spaceships
	active volcanoes	blonde hair	the wild west	developed countries
	global warming	big bangs	bad luck	happy birthday
	daily routines	fizzy drinks	the average temperature	dark hair
v+prep	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	throw away	pull off	calm down	cut off
	dive into	complain about	slow down	shut up
	sit down	pick up	wake up	look forward
	talk about	look at	stand up	clean up
prep+n	transparent		opaque	

	congruent	non-congruent	congruent	non-congruent
	at least	on duty	in fact	in a circle
	till tomorrow	on foot	in light of	across the country
	at (three) o'clock	on a picnic	in return	for ages
	for example	on (the) floor	in horror	around (the) world
Second group				
v+n	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	develop skills	heat oil	pass an exam	shake hands
	add sugar	lend money	climb mountains	borrow money
	drink tea	play the guitar	cook dinner	catch a bus
	answer a question	play a game	follow (the) path	make a decision
adj+n	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	prehistoric people	a loud noise	an open mouth	terrible mistakes
	radioactive materials	precious metals	weak syllables	a gentle voice
	(my) best friend	Native Americans	mobile phones	good news
	correct answers	the main entrance	the same time	heavy rain
v+prep	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	pull out	divide into	move away	bump into
	pass through	argue against	take away	check out

	go ahead	ought to	go on	set up
	worry about	go back	write down	get up
prep+n	transparent		opaque	
	congruent	non-congruent	congruent	non-congruent
	under control	on the screen	in a row	on sale
	at crossroads	on a train	in charge	by accident
	on television	at the end	for a while	over years
	out of control	on the right	for a walk	of course

Appendix (D): The receptive tests



Participant's name:

School:

Code:

Gender: F / M

COLLMATCH3: Test 1

نمف بهشه ژتاقیرنی ژ ٦٤ جووت باندیان پیکهاتیبه (٦٤-١). نمرکی ته نمره کو بریار دهی کا نمف ووشنن لیکدایی ل نینگلیزیی دا بقی شیوهی دهینه بکارنننن یان نه. همکه تو دبینی کو ددرستن هیمایی راست لیده بو (بملی) یان هیمایی شانس لیده بو (نه). بو نمونه
 65. answer the telephone 66. reply the telephone
 ل نمونا سمری جووته ژماره ٦٥ دهینه بکارنننن ل٦٦ ناهینه بکارنننن ل نینگلیزیی دا.
 هیقیه بهرسفا همی جووتکا بدن. سوپاس بو بهشدار یا هموین بهریز ----

- 1 comb her hair 2 kiss his cheek 3 enquire a question 4 climb a tree
- 5 brush his teeth 6 give a debt 7 launch a satellite 8 fly a plane
- 9 bow his knees 10 pull tourists 11 ride a bicycle 12 cross the river
- 13 follow the instructions 14 pay attention 15 lose weight 16 do a mistake
- 17 electric motors 18 active volcanoes 19 global warming 20 daily conventions
- 21 modal verbs 22 blonde hair 23 big bang 24 fizzy drinks
- 25 the industrial uprising 26 the wild west 27 bad luck 28 the average temperature
- 29 foreign spaceships 30 developed countries 31 glad birthday 32 dark hair
- 33 throw away 34 dive within 35 sit down 36 talk about
- 37 pull off 38 complain around 39 pick up 40 look at
- 41 calm low 42 slow down 43 wake high 44 stand up
- 45 cut off 46 shut up 47 look onward 48 clean up
- 49 on least 50 till tomorrow 51 in 3 o'clock 52 for example
- 53 on duty 54 on foot 55 in a picnic 56 on the floor
- 57 in fact 58 on light of 59 in return 60 in horror
- 61 in a circle 62 past the country 63 to ages 64 around the world

Thank you

Participant's name:

School:

Code:

Gender: F / M

COLLMATCH3: Test 2

٦٤ جوت بٺديان بٺكهاٺبٺبه (٦٤-١). ٺٺركٺ ٺه ٺهوه كو بٺريار دهٺ كا ٺهٺف ووشٺن لٺكدا بٺي ل ٺٺنگلٺز بٺي دا
 بٺي شٺوهٺي دهٺنه بكار ٺٺنان يان نه. ههكه ٺو دبٺبٺي كو ددرستن هٺما بٺي راست لٺده بو (بٺلٺ) يان هٺما بٺي شاش لٺده بو (نه). بو نمونه
 65. answer the telephone 66. reply the telephone
 ل نمونا سهري جوٺه ژماره ٦٥ دهٺنه بكار ٺٺنان لٺ ٦٦ ناهٺنه بكار ٺٺنان ل ٺٺنگلٺز بٺي دا.
 هٺقٺبه برسفا ههٺمي جوٺنكا بدهن. سوٺاس بو بهشداريا ههريٺن بهريز ----

- 1 develop skills 2 add sugar 3 drink tea 4 reply a question
- 5 heat oil 6 lend money 7 play the guitar 8 play a game
- 9 cross an exam 10 climb a mountain 11 cook dinner 12 trail a path
- 13 shake hands 14 borrow money 15 catch a bus 16 create a decision
- 17 prehistoric people 18 radioactive materials 19 finest friends 20 correct answers
- 21 a loud noise 22 precious metals 23 Native Americans 24 chief entrance
- 25 open mouth 26 weak syllables 27 movable phones 28 same time
- 29 terrible mistake 30 gentle voice 31 good news 32 thick rain
- 33 pull out 34 pass through 35 go away 36 worry around
- 37 divide into 38 argue versus 39 ought to 40 go back
- 41 move away 42 take away 43 go on 44 write down
- 45 bump into 46 check out 47 set up 48 get up
- 49 below control 50 at crossroads 51 on television 52 out of control
- 53 on the screen 54 on the train 55 at the end 56 in the right
- 57 in a row 58 in charge 59 for a while 60 for a walk
- 61 on sale 62 with accident 63 over years 64 of course

Thank You

Appendix (E): The productive tests



Participant's name:

School:

Code:

Gender: F / M

Gap-filling: Test 1

ئەف بەشە ياژ ٦٤ بوشاييان يڭ دەوت (١--٦٤). ئەركىن تە فەن بوشاييان بووتۇن گونجايى تۆزى كەى. وەك ناسانكار بەك پىتائىكىن ژ ووتۇن پىنقى بو فەن بوشاييا ھاتىە دان وەكى ئەف نەمونا ل خارى:

65. Azad is good at m..... friends.

65 Azad is good at making friends.

ھىقىە ھەمى بوشاييا تۆزى كەن. سوپاس بو بەشداريا ھەمىن بەوز.

1. Could you please tell me what skills I should d..... to get this job?
2. Some people do not a..... sugar to tea.
3. People d..... tea in very small cups.
4. I could not a..... any questions of the yesterday test. They were very difficult.
5. Lucy, get a pan and h..... a little oil in it to fry an egg.
6. Could you L..... me some money for an ice cream?
7. I can't p.....the guitar very well.
8. Let's p..... a game, shall we?
9. You should p..... your exams if you want to go to the university.
10. Can you c.....this mountain? It is very high.
11. Julie c..... dinner for us yesterday.
12. If you want to succeed, f.....my path.
13. When we meet someone's parents, we s..... hands.
14. I have not got any money. Could I b..... some money from you?
15. You can c..... a bus to the city centre. You don't have to walk the whole way.
16. Miss Havisham and her friend m..... the decision to get married.
17. In the ancient times, P..... people put rollers under heavy objects to move them from a place to another.
18. Marie Curie discovered some r..... materials such as Radium and Polonium.
19. I was excited about my future, but I was sad to leave my b..... friends.
20. Please read the questions and choose the c..... answers.
21. Suddenly, Pip heard L..... noises of guns that seemed to come from the sea.
22. In the past, coins were made of p..... metals such as gold and silver.
23. America was the land of five million n..... Americans, who fought hard to keep the Europeans out.
24. You can get buses to the city centre at the bus stop outside the m..... entrance of the village.
25. Mr Jagers saw me looking at Molly with wide eyes and an o..... mouth.
26. Words in English consist of w..... and strong syllables.
27. Students are not allowed to use m..... phones in the class.
28. Some people believe that if two people look in a mirror at the s..... time, they will have an argument.

29. Julie made a **t**..... mistake; she threw water on the burning oil.
30. Magwitch spoke to Pip in a **g**..... voice and said, "Thank you, my dear boy."
31. I have got some **g**..... news! I have just had a phone call from UK.
32. It can be dangerous to fly in **h**..... rain as the hang-glider can get wet.
33. Marco Polo pulled **o**..... packet after packet of the precious stones they had brought from Cathay.
34. Rays of light passed **t**..... a tiny hole in my dark room.
35. You go **a**..... I will stay here.
36. We all worry **a**..... pollution.
37. The 'Holland Code' is a useful tool that divides all people **i**..... just six groups.
38. Each speaker can argue **a**..... or for other speakers' ideas.
39. We ought **t**..... find an old farm building for our project.
40. Joe had to go **b**..... to the village, because he thought that London was not for him.
41. I'd hate to move **a**..... from Kurdistan. I like it so much.
42. Linda has taken **a**..... all love from my heart.
43. I can't go **o**..... in this job. It is too hard.
44. Students should write **d**.....new words after each class.
45. While I was walking, I bumped **i**..... a man unintentionally.
46. Some of us are checking **o**..... the music scene here in Duhok.
47. Emma has set **u**..... a new school website.
48. I usually get **u**..... at 7 o'clock.
49. By the evening, the great fire in the building was completely under **c**.....
50. My friend had an accident **a**.....the crossroads near my house yesterday.
51. There is a good music programme **o**..... television now.
52. Wildfires have been burning **o**.....of control in Australia for days.
53. A story can be made much more interesting **o**..... the screen by using special effects.
54. Mr Fix followed Mr Fogg and Passepartout to Calcutta **o**..... the train.
55. **A**..... the end of the training, we said good-bye to each other.
56. When you go to the city centre, our house is **o**..... the right.
57. You must do this exercise four times **i**..... a row
58. The director is **i**..... charge of the documentary film.
59. Your brother stood watching us **f**..... a while and then walked slowly towards home.
60. I usually go **f**..... a walk in the city centre at 4 p.m.
61. In this year, the first iphone 6s went **o**..... sale to the public.
62. I met my friend in front of a hair salon **b**.....accident.
63. Hang-gliding is a popular sport for **o**..... years.
64. Could you say hello to him for me? Yes, **o**..... course.

Thank You

Participant's name:

School:

Code:

Gender: F / M

Gap-filling: Test 2

ئەف بەشە يا ژ ٦٤ بوشاييان ٽيک دھت (٦٤_١). ئەرکئ تە ڤان بوشاييان بووئئن گونجايي تزي کەي. وەک ئاسانکار يەک پيتائکئ
ژ ووئئن ٽئفئ بو ڤان بوشاييا هاتيه دان وەکئ ئەف نمونال خارئ:

65. Azad is good at **m**..... friends.

65 Azad is good at **making** friends.

هيفيه هەمی بوشاييا تزي کەن. سوپاس بو بەشداريا هەوئن بەئز.

Now, fill in the gaps in these sentences with the most appropriate nodes or collocates guided by the priming letters given for each blank.

1. I **c**..... my hair with a plastic comb every day.
2. When you meet someone's parents, you **k**..... them twice on the cheek.
3. Excuse me, can I **a**..... a question?
4. Cats can **c**..... trees.
5. I **b**..... my teeth every day.
6. Joe had **p**..... all my debts for me before he left.
7. The Russians **L**..... a satellite called Sputnik.
8. Pilots can **f**..... planes high.
9. I am watching Silvia. She is **b**..... her knees now.
10. Every year, Zaweta and similar beauty spots in Kurdistan **a**..... more tourists.
11. Let's **r**..... our bicycles.
12. How did they **c**..... the river to the other side?
13. You can make one of these pictures. **F**..... the instructions.
14. No one **p**..... much attention to Bell's telephone until the Emperor of Brazil picked it up.
15. You do not have to **L**..... weight. You are not fat.
16. Where did Ed **m**..... a mistake?
17. Some new vehicles are hybrids that have both an **e**..... motor and a petrol engine.
18. Mount Ruapehu (m), is an **a**..... volcano.
19. The scientists have been warning us about **g**..... warming for ages.
20. Can you, please, talk about your **d**..... routines?
21. We can use **m**..... verbs to express possibility and certainty.
22. Maddy has got **b**..... hair just like the colour of coffee.
23. Tonight there's going to be a great **b**..... bang at the Music Now stage!
24. Can I have a fizzy **d**..... please?
25. In the early nineteenth century, the Industrial **R**..... started spreading across Europe.
26. Mr Fogg took a train to the **w**..... West.
27. In the Spanish culture, if you put your hat on a bed, it will bring **b**..... luck.
28. **A**..... temperatures have been rising steadily in the past few years.

29. The name "Flying saucers" is the popular name for a..... spaceships.
30. Japan has now become one of the most d..... countries on Earth.
31. On your birthday, everybody says, "H..... Birthday".
32. Dana has got short d..... hair.
33. If you send a cassette to a record company, they throw it a.....
34. A film-maker dived i..... the sea with his camera.
35. Why don't you sit d..... and have a rest? You look tired.
36. Can you talk a..... your future plans?
37. Three angry priests began attacking Passepartout and pulling o..... his shoes.
38. Joe was gentle man and he never complained a..... Mrs Joe.
39. Uncle Pumblechook smiled and picked u..... his knife and fork to eat the meat pie.
40. I was afraid to look a..... her eyes, so I looked around me.
41. Emma asked Mike to calm d.....
42. I had a car accident because I did not slow d..... fast enough.
43. The football coach said that we had to wake u..... fast.
44. Our teacher told us to stand u..... when we answer a question.
45. Everyone who saw the film believed that the Queen's head really had been cut o.....
46. Dave told Maddy to shut u.....
47. If you travel to Britain, what will you look f..... to most?
48. The practical lesson showed us how to clean u..... litter.
49. Hang-gliders need a steady wind of a..... least 12 kilometers an hour.
50. Our school doesn't start t..... tomorrow.
51. The train leaves a..... six o'clock in the morning.
52. There have been frightening monsters in films, f..... example.
53. There is always a nurse o..... duty in the health center on the first floor.
54. Kawa is wearing white trainers o..... his feet.
55. In Kurdistan, I like going o..... picnics.
56. When I was leaving I saw some money o..... the floor.
57. Have you ever been abroad before? i..... fact, I have never been outside Kurdistan.
58. Pip walked in and looked round i..... the light of the evening moon.
59. A rich man gave a poor family some land to grow their crops. i..... return, the poor family worked for him.
60. When Miss Havisham's dress was on fire, she stood up and cried out i..... horror.
61. Women usually dance i..... a circle.
62. The modern form of football was quickly accepted by clubs a..... the country.
63. I have not seen them f..... ages.
64. Mr Fogg said, "I have won! I have been a..... the world in eighty days."

Thank You

Appendix (F): Characteristics of the test items

receptive 1/productive 2				
v+n	transparent		opaque	
v+n	congruent	non-congruent	congruent	non-congruent
v+n	comb (her) hair	brush (his/her) teeth	bend (her/his) knees	follow the instructions
v+n	Kiss (her) cheek	pay a debt	attract tourists	pay attention
v+n	ask a question	launch a satellite	ride a bicycle	lose weight
v+n	climb a tree	fly a plane	cross the river	make a mistake
adj+n	transparent		opaque	
adj+n	congruent	non-congruent	congruent	non-congruent
adj+n	electric motors	modal verbs	(the) industrial revolution	alien spaceships
adj+n	active volcanoes	blonde hair	the wild west	developed countries
adj+n	global warming	big bangs	bad luck	happy birthday
adj+n	daily routines	fizzy drinks	the average temperature	dark hair
phrasal verb	transparent		opaque	
phrasal verb	congruent	non-congruent	congruent	non-congruent
phrasal verb	throw away	pull off	calm down	cut off
phrasal verb	dive into	complain about	slow down	shut up
phrasal verb	sit down	pick up	wake up	look forward
phrasal verb	talk about	look at	stand up	clean up
prep+n	transparent		opaque	
prep+n	congruent	non-congruent	congruent	non-congruent
prep+n	at least	on duty	in fact	in a circle
prep+n	till tomorrow	on foot	in light of	across the country
prep+n	at (three) o'clock	on a picnic	in return	for ages
prep+n	for example	on (the) floor	in horror	around (the) world
receptive 2/productive 1				
v+n	transparent		opaque	
v+n	congruent	non-congruent	congruent	non-congruent

v+n	develop skills	heat oil	pass an exams	shake hands
v+n	add sugar	lend money	climb mountains	borrow money
v+n	drink tea	play the guitar	cook dinner	catch a bus
v+n	answer a question	play a game	follow (the) path	make a decision
adj+n	transparent		opaque	
adj+n	congruent	non-congruent	congruent	non-congruent
adj+n	prehistoric monsters	a loud noise	an open mouth	terrible mistakes
adj+n	radioactive materials	precious metals	weak syllables	a gentle voice
adj+n	(my) best friend	Native Americans	mobile phones	good news
adj+n	correct answers	the main entrance	the same time	heavy rain
phrasal verb	transparent		opaque	
phrasal verb	congruent	non-congruent	congruent	non-congruent
phrasal verb	pull out	divide into	move away	bump into
phrasal verb	pass through	argue against	take away	check out
phrasal verb	go ahead	ought to	go on	set up
phrasal verb	worry about	go back	write down	get up
prep+n	transparent		opaque	
prep+n	congruent	non-congruent	congruent	non-congruent
prep+n	under control	on the screen	in a row	on sale
prep+n	at crossroads	on a train	in charge	by accident
prep+n	on television	at the end	for a while	over years
prep+n	out of control	on the right	for a walk	of course

Appendix (G): Item analysis-receptive 1

Item no.	Items	N	Total scores on the item	Facility value	Std. Deviation
1	comb hair	252	192	0.76	0.43
2	Kiss cheek	252	152	0.60	0.49
3	ask a question	252	202	0.80	0.40
4	climb a tree	252	224	0.89	0.32
5	brush his teeth	252	202	0.80	0.40
6	pay a debt	252	147	0.58	0.49
7	launch a satellite	252	112	0.44	0.50
8	fly a plane	252	164	0.65	0.48
9	bend his knees	252	156	0.62	0.49
10	attract tourists	252	170	0.67	0.47
11	ride a bicycle	252	228	0.90	0.29
12	cross the river	252	230	0.91	0.28
13	follow the instructions	252	197	0.78	0.41
14	pay attention	252	138	0.55	0.50
15	lose weight	252	223	0.88	0.32
16	make a mistake	252	97	0.38	0.49
17	electric motors	252	201	0.80	0.40
18	active volcanoes	252	136	0.54	0.50
19	global warming	252	161	0.64	0.48
20	daily routines	252	116	0.46	0.50
21	modal verbs	252	127	0.50	0.50
22	blond hair	252	199	0.79	0.41
23	big bang	252	226	0.90	0.31
24	fizzy drinks	252	186	0.74	0.44
25	the industrial revolution	252	150	0.60	0.49
26	the wild west	252	179	0.71	0.46
27	bad luck	252	237	0.94	0.24
28	the average temperature	252	147	0.58	0.49
29	alien spaceships	252	144	0.57	0.50
30	developed countries	252	202	0.80	0.40
31	happy birthday	252	188	0.75	0.44
32	dark hair	252	236	0.94	0.24
33	throw away	252	211	0.84	0.37
34	dive into	252	193	0.77	0.42

35	sit down	252	243	0.96	0.19
36	talk about	252	241	0.96	0.21
37	pull off	252	148	0.59	0.49
38	complain about	252	149	0.59	0.49
39	pick up	252	236	0.94	0.24
40	look at	252	243	0.96	0.19
41	calm down	252	207	0.82	0.38
42	slow down	252	222	0.88	0.32
43	wake up	252	204	0.81	0.39
44	stand up	252	240	0.95	0.21
45	cut off	252	167	0.66	0.47
46	shut up	252	221	0.88	0.33
47	Look forward	252	183	0.73	0.45
48	clean up	252	204	0.81	0.39
49	at least	252	137	0.54	0.50
50	till tomorrow	252	192	0.76	0.43
51	at three o'clock	252	183	0.73	0.45
52	for example	252	243	0.96	0.19
53	on duty	252	125	0.50	0.50
54	on foot	252	127	0.50	0.50
55	on a picnic	252	64	0.25	0.44
56	on the floor	252	224	0.89	0.32
57	in fact	252	206	0.82	0.39
58	in light of	252	160	0.63	0.48
59	in return	252	103	0.41	0.49
60	in horror	252	176	0.70	0.46
61	in a circle	252	203	0.81	0.40
62	across the country	252	125	0.50	0.50
63	for ages	252	150	0.60	0.49
64	around the world	252	240	0.95	0.21

Appendix (H): Item analysis-receptive 2

Item no.	Items	Number of participants	Total scores on the item	Facility value	Std. Deviation
1	develop skills	252	130	0.52	0.50
2	add sugar	252	187	0.74	0.44
3	drink tea	252	232	0.92	0.27
4	answer a question	252	152	0.60	0.49
5	heat oil	252	94	0.37	0.49
6	lend money	252	131	0.52	0.50
7	play the guitar	252	237	0.94	0.24
8	play a game	252	233	0.92	0.27
9	pass an exam	252	181	0.72	0.45
10	climb mountains	252	232	0.92	0.27
11	cook dinner	252	237	0.94	0.24
12	follow the path	252	184	0.73	0.45
13	shake hands	252	192	0.76	0.43
14	borrow money	252	147	0.58	0.49
15	catch a bus	252	206	0.82	0.39
16	make a decision	252	134	0.53	0.50
17	prehistoric people	252	145	0.58	0.50
18	radioactive materials	252	120	0.48	0.50
19	best friend	252	166	0.66	0.48
20	correct answers	252	236	0.94	0.24
21	loud noise	252	166	0.66	0.48
22	precious metals	252	154	0.61	0.49
23	native Americans	252	153	0.61	0.49
24	main entrance	252	187	0.74	0.44
25	open mouth	252	221	0.88	0.33
26	weak syllables	252	95	0.38	0.49
27	mobile phones	252	175	0.69	0.46
28	same time	252	227	0.90	0.30
29	terrible mistake	252	209	0.83	0.38
30	gentle voice	252	178	0.71	0.46
31	good news	252	237	0.94	0.24
32	heavy rain	252	174	0.69	0.46
33	pull out	252	141	0.56	0.50
34	pass through	252	167	0.66	0.47
35	go away	252	237	0.94	0.24
36	worry about	252	206	0.82	0.39
37	divide into	252	154	0.61	0.49

38	argue against	252	184	0.73	0.45
39	ought to	252	211	0.84	0.37
40	go back	252	232	0.92	0.27
41	move away	252	168	0.67	0.47
42	take away	252	209	0.83	0.38
43	go on	252	237	0.94	0.24
44	write down	252	209	0.83	0.38
45	bump into	252	91	0.36	0.48
46	check out	252	212	0.84	0.37
47	set up	252	215	0.85	0.36
48	get up	252	232	0.92	0.27
49	under control	252	166	0.66	0.48
50	at the crossroads	252	145	0.58	0.50
51	on television	252	223	0.88	0.32
52	out of control	252	233	0.92	0.27
53	on the screen	252	192	0.76	0.43
54	o the train	252	179	0.71	0.46
55	at the end	252	224	0.89	0.32
56	on the right	252	106	0.42	0.50
57	in a row	252	142	0.56	0.50
58	in charge	252	193	0.77	0.42
59	for a while	252	211	0.84	0.37
60	for a walk	252	175	0.69	0.46
61	on sale	252	180	0.71	0.45
62	by accident	252	126	0.50	0.50
63	over years	252	175	0.69	0.46
64	of course	252	184	0.73	0.45

Appendix (I): Item analysis- productive 1

Item no.	Items	Number of participants	Total scores on the item	Facility value	Std. Deviation
1	develop his skills	252	6	0.02	.15
2	add sugar	252	121	0.48	.50
3	drink tea	252	218	0.87	.34
4	answer a question	252	159	0.63	.48
5	heat oil	252	37	0.15	.36
6	lend money	252	137	0.54	.50
7	play the guitar	252	232	0.92	.27
8	play a game	252	236	0.94	.24
9	pass an exam	252	125	0.50	.50
10	climb mountains	252	219	0.87	.34
11	cook dinner	252	221	0.88	.33
12	follow the path	252	131	0.52	.50
13	shake hands	252	120	0.48	.50
14	borrow money	252	141	0.56	.50
15	catch a bus	252	189	0.75	.43
16	make a decision	252	169	0.67	.47
17	prehistoric people	252	6	0.02	.15
18	radioactive materials	252	37	0.15	.36
19	best friend	252	214	0.85	.36
20	correct answers	252	218	0.87	.34
21	loud noise	252	156	0.62	.49
22	precious metals	252	27	0.11	.31
23	native Americans	252	31	0.12	.33
24	main entrance	252	51	0.20	.40
25	open mouth	252	129	0.51	.50
26	weak syllables	252	42	0.17	.37
27	mobile phones	252	169	0.67	.47
28	same time	252	224	0.89	.32
29	terrible mistake	252	179	0.71	.46
30	gentle voice	252	11	0.04	.21
31	good news	252	185	0.73	.44
32	heavy rain	252	78	0.31	.46
33	pull out	252	63	0.25	.43
34	pass through	252	40	0.16	.37
35	go away	252	18	0.07	.26
36	worry about	252	172	0.68	.47
37	divide into	252	153	0.61	.49

38	argue against	252	5	0.02	.14
39	ought to	252	194	0.77	.42
40	go back	252	100	0.40	.49
41	move away	252	136	0.54	.50
42	take away	252	41	0.16	.37
43	go on	252	83	0.33	.47
44	write down	252	38	0.15	.36
45	bump into	252	115	0.46	.50
46	check out	252	62	0.25	.43
47	set up	252	110	0.44	.50
48	get up	252	210	0.83	.37
49	under control	252	108	0.43	.50
50	at the crossroads	252	120	0.48	.50
51	on television	252	217	0.86	.35
52	out of control	252	120	0.48	.50
53	on the screen	252	167	0.66	.47
54	o the train	252	207	0.82	.38
55	at the end	252	174	0.69	.46
56	on the right	252	214	0.85	.36
57	in a row	252	158	0.63	.49
58	in charge	252	164	0.65	.48
59	for a while	252	151	0.60	.49
60	for a walk	252	184	0.73	.45
61	on sale	252	152	0.60	.49
62	by accident	252	141	0.56	.50
63	over years	252	32	0.13	.33
64	of course	252	243	0.96	.19

Appendix (J): Item analysis-productive 2

Item no.	Items	Number of participants	Total scores on the item	Facility value	Std. Deviation
1	comb hair	252	102	.40	0.49
2	Kiss cheek	252	112	.44	0.50
3	ask a question	252	237	.94	0.24
4	climb a tree	252	184	.73	0.45
5	brush his teeth	252	196	.78	0.42
6	pay a debt	252	159	.63	0.48
7	launch a satellite	252	47	.19	0.39
8	fly a plane	252	203	.81	0.40
9	bend his knees	252	28	.11	0.32
10	attract tourists	252	29	.12	0.32
11	ride a bicycle	252	164	.65	0.48
12	cross the river	252	163	.65	0.48
13	follow the instructions	252	36	.14	0.35
14	pay attention	252	82	.33	0.47
15	lose weight	252	170	.67	0.47
16	make a mistake	252	222	.88	0.32
17	electric motors	252	82	.33	0.47
18	active volcanoes	252	33	.13	0.34
19	global warming	252	155	.62	0.49
20	daily routines	252	136	.54	0.50
21	modal verbs	252	94	.37	0.49
22	blond hair	252	120	.48	0.50
23	big bang	252	124	.49	0.50
24	fizzy drinks	252	174	.69	0.46
25	the industrial revolution	252	51	.20	0.40
26	the wild west	252	64	.25	0.44
27	bad luck	252	192	.76	0.43
28	the average temperature	252	30	.12	0.32
29	alien spaceships	252	12	.05	0.21
30	developed countries	252	105	.42	0.49
31	happy birthday	252	246	.98	0.15
32	dark hair	252	205	.81	0.39
33	throw away	252	132	.52	0.50
34	dive into	252	131	.52	0.50
35	sit down	252	218	.87	0.34
36	talk about	252	201	.80	0.40
37	pull off	252	52	.21	0.41

38	complain about	252	129	.51	0.50
39	pick up	252	209	.83	0.38
40	look at	252	194	.77	0.42
41	calm down	252	174	.69	0.46
42	slow down	252	184	.73	0.45
43	wake up	252	195	.77	0.42
44	stand up	252	210	.83	0.37
45	cut off	252	135	.54	0.50
46	shut up	252	187	.74	0.44
47	Look forward	252	158	.63	0.49
48	clean up	252	133	.53	0.50
49	at least	252	100	.40	0.49
50	till tomorrow	252	91	.36	0.48
51	at three o'clock	252	188	.75	0.44
52	for example	252	194	.77	0.42
53	on duty	252	164	.65	0.48
54	on foot	252	189	.75	0.43
55	on a picnic	252	133	.53	0.50
56	on the floor	252	203	.81	0.40
57	in fact	252	163	.65	0.48
58	in light of	252	167	.66	0.47
59	in return	252	123	.49	0.50
60	in horror	252	187	.74	0.44
61	in a circle	252	194	.77	0.42
62	across the country	252	25	.10	0.30
63	for ages	252	209	.83	0.38
64	around the world	252	155	.62	0.49

Appendix (K): Item-total statistics- receptive test 1

No		Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	comb hair	.14	0.73
2	Kiss cheek	.16	0.73
3	ask a question	.11	0.73
4	climb a tree	.28	0.72
5	brush his teeth	.02	0.73
6	pay a debt	.05	0.73
7	launch a satellite	.28	0.72
8	fly a plane	-.06	0.74
9	bend his knees	-.001	0.73
10	attract tourists	.34	0.72
11	ride a bicycle	.35	0.72
12	cross the river	.23	0.72
13	follow the instructions	.25	0.72
14	pay attention	.16	0.73
15	lose weight	.16	0.73
16	make a mistake	.007	0.73
17	electric motors	.18	0.72
18	active volcanoes	.19	0.72
19	global warming	.21	0.72
20	daily routines	.07	0.73
21	modal verbs	.05	0.73
22	blond hair	.13	0.73
23	big bang	.20	0.72
24	fizzy drinks	.03	0.73
25	the industrial revolution	.13	0.73
26	the wild west	.37	0.72
27	bad luck	.16	0.73
28	the average temperature	.16	0.73
29	alien spaceships	.09	0.73
30	developed countries	.23	0.72
31	happy birthday	.20	0.72
32	dark hair	.21	0.72
33	throw away	.24	0.72
34	dive into	.07	0.73
35	sit down	.24	0.73
36	talk about	.29	0.72

37	pull off	.29	0.72
38	complain about	.28	0.72
39	pick up	.34	0.72
40	look at	.04	0.73
41	calm down	.38	0.72
42	slow down	.22	0.72
43	wake up	.29	0.72
44	stand up	.35	0.72
45	cut off	.24	0.72
46	shut up	.40	0.72
47	Look forward	.36	0.72
48	clean up	.32	0.72
49	at least	.22	0.72
50	till tomorrow	.24	0.72
51	at three o'clock	.09	0.73
52	for example	.29	0.72
53	on duty	.18	0.72
54	on foot	.12	0.73
55	on a picnic	-.03	0.73
56	on the floor	.24	0.72
57	in fact	.25	0.72
58	in light of	.29	0.72
59	in return	.07	0.73
60	in horror	.03	0.73
61	in a circle	.20	0.72
62	across the country	-.008	0.73
63	for ages	.09	0.73
64	around the world	.24	0.72

Appendix (L): Receptive test 1 item-total statistics after deleting four items

Item number	Collocations	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	comb hair	0.14	0.75
2	Kiss cheek	0.16	0.75
3	ask a question	0.11	0.75
4	climb a tree	0.29	0.74
5	brush his teeth	0.02	0.75
6	pay a debt	0.03	0.75
7	launch a satellite	0.29	0.74
10	attract tourists	0.34	0.74
11	ride a bicycle	0.35	0.74
12	cross the river	0.22	0.75
13	follow the instructions	0.24	0.75
14	pay attention	0.16	0.75
15	lose weight	0.16	0.75
16	make a mistake	-0.01	0.76
17	electric motors	0.20	0.75
18	active volcanoes	0.20	0.75
19	global warming	0.23	0.75
20	daily routines	0.06	0.75
21	modal verbs	0.07	0.75
22	blond hair	0.15	0.75
23	big bang	0.20	0.75
24	fizzy drinks	0.04	0.75
25	the industrial revolution	0.12	0.75
26	the wild west	0.37	0.74
27	bad luck	0.17	0.75
28	the average temperature	0.18	0.75
29	alien spaceships	0.06	0.75
30	developed countries	0.23	0.75

31	happy birthday	0.19	0.75
32	dark hair	0.21	0.75
33	throw away	0.24	0.75
34	dive into	0.07	0.75
35	sit down	0.23	0.75
36	talk about	0.31	0.75
37	pull off	0.30	0.74
38	complain about	0.27	0.74
39	pick up	0.35	0.74
40	look at	0.05	0.75
41	calm down	0.37	0.74
42	slow down	0.24	0.75
43	wake up	0.27	0.74
44	stand up	0.36	0.74
45	cut off	0.26	0.74
46	shut up	0.42	0.74
47	Look forward	0.36	0.74
48	clean up	0.32	0.74
49	at least	0.22	0.75
50	till tomorrow	0.26	0.74
51	at three o'clock	0.09	0.75
52	for example	0.30	0.75
53	on duty	0.19	0.75
54	on foot	0.14	0.75
56	on the floor	0.26	0.75
57	in fact	0.26	0.74
58	in light of	0.28	0.74
59	in return	0.07	0.75
60	in horror	0.04	0.75
61	in a circle	0.22	0.75
63	for ages	0.09	0.75
64	around the world	0.23	0.75

Appendix (M): Receptive test 2 item-total statistics

No.	Items	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	develop skills	0.21	0.76
2	add sugar	0.26	0.76
3	drink tea	0.12	0.76
4	answer a question	0.09	0.76
5	heat oil	0.14	0.76
6	lend money	0.18	0.76
7	play the guitar	0.16	0.76
8	play a game	0.29	0.76
9	pass an exam	0.23	0.76
10	climb mountains	0.28	0.76
11	cook dinner	0.10	0.76
12	follow the path	0.08	0.76
13	shake hands	0.35	0.75
14	borrow money	0.33	0.75
15	catch a bus	0.27	0.76
16	make a decision	0.23	0.76
17	prehistoric people	0.11	0.76
18	radioactive materials	0.26	0.76
19	best friend	0.09	0.76
20	correct answers	0.15	0.76
21	loud noise	0.20	0.76
22	precious metals	0.17	0.76
23	native Americans	0.24	0.76
24	main entrance	0.19	0.76
25	open mouth	0.13	0.76
26	weak syllables	0.20	0.76
27	mobile phones	0.14	0.76
28	same time	0.24	0.76
29	terrible mistake	0.26	0.76
30	gentle voice	0.21	0.76
31	good news	0.28	0.76
32	heavy rain	0.08	0.76
33	pull out	0.27	0.76
34	pass through	0.17	0.76
35	go away	0.22	0.76
36	worry about	0.21	0.76
37	divide into	0.28	0.76
38	argue against	-0.01	0.77
39	ought to	0.36	0.75
40	go back	0.23	0.76

41	move away	0.11	0.76
42	take away	0.01	0.76
43	go on	0.26	0.76
44	write down	0.34	0.75
45	bump into	0.21	0.76
46	check out	0.31	0.76
47	set up	0.25	0.76
48	get up	0.31	0.76
49	under control	0.23	0.76
50	at the crossroads	0.25	0.76
51	on television	0.33	0.76
52	out of control	0.31	0.76
53	on the screen	0.24	0.76
54	o the train	0.06	0.76
55	at the end	0.30	0.76
56	on the right	0.15	0.76
57	in a row	0.29	0.75
58	in charge	0.27	0.76
59	for a while	0.25	0.76
60	for a walk	0.09	0.76
61	on sale	0.25	0.76
62	by accident	0.06	0.76
63	over years	0.14	0.76
64	of course	-0.19	0.77

Appendix (N): Item total statistics-productive test1

Item number		Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
1	develop his skills	0.20	0.93
2	add sugar	0.47	0.93
3	drink tea	0.26	0.93
4	answer a question	0.29	0.93
5	heat oil	0.39	0.93
6	lend money	0.52	0.93
7	play the guitar	0.28	0.93
8	play a game	0.33	0.93
9	pass an exam	0.36	0.93
10	climb mountains	0.45	0.93
11	cook dinner	0.33	0.93
12	follow the path	0.38	0.93
13	shake hands	0.56	0.93
14	borrow money	0.59	0.93
15	catch a bus	0.51	0.93
16	make a decision	0.48	0.93
17	prehistoric people	0.08	0.93
18	radioactive materials	0.34	0.93
19	best friend	0.32	0.93
20	correct answers	0.45	0.93
21	loud noise	0.57	0.93
22	precious metals	0.36	0.93
23	native Americans	0.40	0.93
24	main entrance	0.48	0.93
25	open mouth	0.60	0.93
26	weak syllables	0.45	0.93
27	mobile phones	0.41	0.93
28	same time	0.41	0.93
29	terrible mistake	0.52	0.93
30	gentle voice	0.24	0.93
31	good news	0.42	0.93
32	heavy rain	0.38	0.93
33	pull out	0.44	0.93
34	pass through	0.41	0.93
35	go away	0.07	0.93
36	worry about	0.42	0.93
37	divide into	0.56	0.93
38	argue against	0.13	0.93
39	ought to	0.45	0.93
40	go back	0.28	0.93

41	move away	0.51	0.93
42	take away	0.46	0.93
43	go on	0.39	0.93
44	write down	0.35	0.93
45	bump into	0.37	0.93
46	check out	0.35	0.93
47	set up	0.23	0.93
48	get up	0.28	0.93
49	under control	0.58	0.93
50	at the crossroads	0.44	0.93
51	on television	0.31	0.93
52	out of control	0.51	0.93
53	on the screen	0.46	0.93
54	o the train	0.32	0.93
55	at the end	0.51	0.93
56	on the right	0.42	0.93
57	in a row	0.48	0.93
58	in charge	0.50	0.93
59	for a while	0.49	0.93
60	for a walk	0.58	0.93
61	on sale	0.42	0.93
62	by accident	0.41	0.93
63	over years	0.23	0.93
64	of course	0.26	0.93

Appendix (O): Productive test 2 item total statistics

	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
comb hair	0.43	0.95
Kiss cheek	0.53	0.95
ask a question	0.23	0.95
climb a tree	0.39	0.95
brush his teeth	0.37	0.95
pay a debt	0.51	0.95
launch a satellite	0.37	0.95
fly a plane	0.45	0.95
bend his knees	0.36	0.95
attract tourists	0.38	0.95
ride a bicycle	0.49	0.95
cross the river	0.55	0.95
follow the instructions	0.28	0.95
pay attention	0.46	0.95
lose weight	0.54	0.95
make a mistake	0.42	0.95
electric motors	0.53	0.95
active volcanoes	0.36	0.95
global warming	0.56	0.95
daily routines	0.56	0.95
modal verbs	0.32	0.95
blond hair	0.38	0.95
big bang	0.33	0.95
fizzy drinks	0.60	0.95
the industrial revolution	0.42	0.95
the wild west	0.41	0.95
bad luck	0.42	0.95
the average temperature	0.27	0.95
alien spaceships	0.25	0.95
developed countries	0.49	0.95
happy birthday	0.32	0.95
dark hair	0.31	0.95
throw away	0.50	0.95
dive into	0.51	0.95
sit down	0.43	0.95
talk about	0.44	0.95

pull off	0.26	0.95
complain about	0.59	0.95
pick up	0.49	0.95
look at	0.56	0.95
calm down	0.57	0.95
slow down	0.63	0.95
wake up	0.60	0.95
stand up	0.65	0.95
cut off	0.51	0.95
shut up	0.66	0.95
Look forward	0.39	0.95
clean up	0.51	0.95
at least	0.52	0.95
till tomorrow	0.58	0.95
at three o'clock	0.51	0.95
for example	0.60	0.95
on duty	0.55	0.95
on foot	0.62	0.95
on a picnic	0.31	0.95
on the floor	0.51	0.95
in fact	0.64	0.95
in light of	0.53	0.95
in return	0.55	0.95
in horror	0.60	0.95
in a circle	0.53	0.95
across the country	0.17	0.95
for ages	0.48	0.95
around the world	0.36	0.95

References

- Aghbar, A. (1990). Fixed expressions in written texts: Implications for assessing writing sophistication. *The Meeting of the English Association of Pennsylvania State System Universities*, Pennsylvania, PA. Retrieve from <http://eric.ed.gov/?id=ED329125>. (ED329125)
- Agrawal, S., Sanyal, R., & Sanyal, S. (2014). Statistics and linguistic rules in multiword extraction: A comparative analysis. *International Journal of Reasoning-Based Intelligent Systems*, 6(1-2), 59-70. doi:10.1504/IJRIS.2014.063954
- Ahmadian, M., & Darabi, A. (2012). A study of the relationship between EFL learners' knowledge of near synonyms and their collocational behaviour. *Journal of Educational and Social Research*, 2(3), 179-188. doi:10.5901/jesr.2012.v2n3p179
- Aisenstadt, E. (1979). Collocability restrictions in dictionaries. In R. Hartmann (Ed.), *Dictionaries and their users: Papers from the 1978 BAAL. Seminar on lexicography* (pp. 71-74). Exeter, UK: University of Exeter.
- Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL. Review of Applied Linguistics*, 53, 53-61. doi:10.1075/itl.53.04ais
- Alali, F., & Schmitt, N. (2012). Teaching formulaic sequences: The same as or different from teaching single words? *TESOL Journal*, 3(2), 153-180. doi:10.1002/tesj.13
- Akbarian, I. (2010). The relationship between vocabulary size and depth for ESP/EAP learners. *System*, 38(3), 391-401. doi:10.1016/j.system.2010.06.013
- Alharthi, T. (2014). The dynamic and incremental features of vocabulary acquisition. *International Journal of English Linguistics*, 4(6), 70-77. doi:10.5539/ijel.v4n6p70
- Alias, M. (2005). Assessment of learning outcomes: Validity and reliability of classroom tests. *World Transactions on Engineering and Technology Education*, 4(2), 235-238. Retrieved from <http://eprints.uthm.edu.my/6811/>
- Alsakran, R. (2011). *The productive and receptive knowledge of collocations by advanced Arabic-speaking ESL/EFL learners*. (Master's thesis, Colorado State University). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.818.9152&rep=rep1&type=pdf>
- Alsulayyi, N. (2015). The use of grammatical collocations by advanced Saudi EFL learners in the UK and KSA. *International Journal of English Linguistics*, 5(1), 32-43. doi:10.5539/ijel.v5n1p32
- Altenberg, B., & Granger, S. (2001). The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics*, 22(2), 173-195. doi:10.1093/applin/22.2.173
- Anderson, R., & Freebody, P. (1979). *Vocabulary knowledge*. Technical Report (No. 136). University of Illinois: Centre for the Study of Reading. Retrieved from <http://files.eric.ed.gov/fulltext/ED177480.pdf>

- Arnaud, P., & Savignon, S. (1997). Rare words, complex lexical units and the advanced learner. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp.157-173). Cambridge, UK: Cambridge University Press.
- Ashiyani, Z., & Salehi, H. (2017). A comparison of male and female learners' English collocation learning through using WhatsApp. *International Journal of Research Studies in Educational Technology*, 6(1), 25-42.
doi:10.5861/ijrset.2016.1615
- Ashouri, S. Arjmandi, M., & Rahimi, R. (2014). The impact of corpus-based collocation instruction on Iranian EFL learners' collocation learning. *Universal Journal of Educational Research*, 2(6), 470-479.
doi:10.13189/ujer.2014.020604
- Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal*, 86(4), 525-45. doi:10.1111/1540-4781.00159
- Atkinson, D., Churchill, E., Nishino T., & Okada, H. (2007). Alignment and interaction in a sociocognitive approach to second language acquisition. *The Modern Language Journal*, 91(2), 169-88.
doi:10.1111/j.15404781.2007.00539.x
- Badawi, F. (2008). *Investigating EFL prospective teachers' ability to translate culture-bound expressions*. Retrieved from <http://eric.ed.gov/?id=ED503396> (ED503396)
- Bahardoust, M., & Moeini, M. (2012). Lexical and grammatical collocations in writing production of EFL learners. *The Journal of Applied Linguistics*, 5(1), 61-86. Retrieved from http://www.iauournals.ir/article_520093_6fcd1743e12899e7db8b346a44c4a0cd.pdf
- Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101-114. doi:10.1016/0346-251X(93)90010-E
- Baker, M. (1992). *In other words: A coursebook on translation*. London, UK: Routledge.
- Baker, M. (2001). *The atoms of language*. New York, NY: Basic Books.
- Bandpay, A. (2012). A survey on the impact of L1 grammar negative transfer on L2 writing skill in Iranian EFL students in Rudsar City. *The Iranian EFL Journal*, 8(6), 204-226. Retrieved from <http://profdoc.um.ac.ir/articles/a/1032123.pdf#page=227>
- Bardel, C., Lindqvist, C., & Laufer, B. (Eds.). (2013). *Research on L2 learners' collocational competence and development: A progress report, L2 vocabulary acquisition, knowledge and use* (pp. 5-10). EUROSLA Monographs Series 2. Retrieved from <http://www.eurosla.org/monographs/EM02/Henriksen.pdf>
- Barfield, A., & Gyllstad, H. (Eds.). (2009). *Researching collocations in another language: Multiple interpretations*. Basingstoke, UK: Palgrave Macmillan.
- Barlow, M., & Kemmer, S. (Eds.). (2000). *Usage-based models of language*. Stanford, CA: CSLI.
- Bartsch, S. (2004). *Structural and functional properties of collocations in English*. Tübingen, Germany: Gunter Narr Verlag.

- Bartsch, S., & Evert, S. (2014). Towards a Firthian notion of collocation. In A. Abel, & L. Lemnitzer (Eds.), *Network Strategies, Access Structures and Automatic Extraction of Lexicographical Information. 2nd Work Report of the Academic Network Internet Lexicography, OPAL–Online publizierte Arbeiten zur Linguistik. Institut für Deutsche Sprache, Mannheim* (pp. 48-61). Retrieved from www.stefan-evert.de/PUB/BartschEvert2014.pdf
- Becker, L. (2000). *Effect size (ES)*. Retrieved from <http://www.uv.es/~friasnav/EffectSizeBecker.pdf>
- Bell, H. (2009). The messy little details: A longitudinal case study of the emerging lexicon. In T. Fitzpatrick, & A. Barfield (Eds.), *Lexical processing in second language learners: Papers and perspectives in honor of Paul Meara* (pp. 111-127). Bristol, UK: Multilingual Matters. Retrieved from <https://www.amazon.de/Lexical-Processing-Second-Language-Learners/dp/1847691528>
- Belle, G., Fisher, L., Heagerty, P., & Lumley, T. (2004). *Biostatistics: A methodology for the health sciences* (2nd ed.). New Jersey, NJ: John Wiley and Sons Inc. Retrieved from <http://www.cos.ufrj.br/~bioestatistica/livros/Introduction%20to%20Biostatistics.pdf>
- Benson, M., Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of English: A guide to word combinations*. Amsterdam, Netherland: John Benjamins Publishing Company. doi:10.1075/z.bbi1(1st)
- Bergström, K. (2008). *Vocabulary and receptive knowledge of English collocations among Swedish upper secondary school students*. (Bachelor's degree thesis, Stockholm University). Retrieved from <http://su.diva-portal.org/smash/record.jsf?pid=diva2:219552>
- Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, 26, 28-41. doi:10.1016/j.jslw.2014.09.004
- Biber, D. (1993). Co-occurrence patterns among collocations: A tool for corpus-based lexical knowledge acquisition. *Computational Linguistics*, 19(3), 531-538. Retrieved from <http://www.aclweb.org/anthology/J93-3004?CFID=887564313&CFTOKEN=81644679>
- Blum, S., & Levenston, E. (1980). Lexical simplification in second-language acquisition. *Studies in Second Language Acquisition*, 2(2), 43-63. doi:10.1017/S0272263100000073
- Bolinger, D. (1979). Meaning and memory. In G. Haydu (Ed.), *Experience forms: Their cultural and individual place and function* (pp. 95-111). New York, NY: Mouton Publishers.
- Bonk, W. (2000). *Testing ESL learners' knowledge of collocations* (Report No. FL 801 384). Retrieved from <http://files.eric.ed.gov/fulltext/ED442309.pdf> (ED442309)
- Boers, F., Lindstromberg, S., & Eyckmans, J. (2014). Some explanations for the slow acquisition of L2 collocations. *Vigo International Journal of Applied Linguistics*, 11, 41-62. Retrieved from <http://vialjournal.webs.uvigo.es/pdf/Vial-2014-Article2.pdf>

- Bouma, G. (2009). Normalized (pointwise) mutual information in collocation extraction. *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology* (Vol. 156). Germany: Potsdam. Retrieved from <https://svn.spraakdata.gu.se/repos/gerlof/pub/www/Docs/npmi-pfd.pdf>
- Brezina, V., Mcenery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 202(2015), 139-173. doi:10.1075/ijcl.20.2.01bre
- Bueraheng, N., & Laohawiriyanon, C. (2014). Does learners' degree of exposure to English language influence their collocational knowledge? *International Journal of English and Literature*, 4(3), 1-10. Retrieved from http://www.academia.edu/7700468/DOES_LEARNERS_DEGREE_OF_EXPOSURE_TO_ENGLISH_LANGUAGE_INFLUENCE_THEIR_COLLOCATIONAL_KNOWLEDGE
- Burr, J., & Nesselrode, J. (1990). Change measurement. In A. von Eye (Ed.), *Statistical methods in longitudinal research* (pp. 3-34). San Diego, CA: Academic Press, Inc.
- Bybee, J. (2008). Usage-based grammar and second-language acquisition. In N. Ellis, & P. Robinson (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216-236). New York, NY: Taylor and Francis.
- Bybee, J., & Hopper, P. (2001). *Frequency and the emergence of language structure*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Caspi, T. (2010). *A dynamic perspective on second language development*. (Doctoral dissertation, University of Groningen). Retrieved from <http://irs.ub.rug.nl/ppn/329338412>
- Caspi, T., & Lowie, W. (2013). The dynamics of L2 vocabulary development: A case study of receptive and productive knowledge. *Revista Brasileira de Linguística Aplicada*, 13(2), 437-462. Retrieved from http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-63982013000200005&lng=en&tlng=en
- Cambridge Advanced Learner's Dictionary* (2015). Cambridge, UK: Cambridge University Press. Retrieved from <http://dictionary.cambridge.org/dictionary/english/>
- Catalán, R. (2002). The concept of lexical competence in second language learning and teaching studies. *Atlantis*, XXIV(2), 149-162. Retrieved from http://www.atlantisjournal.org/old/Papers/24_1/jimenez.pdf
- Cellucci, C., Albano, A., & Rapp, P. (2005). Statistical validation of mutual information calculations: Comparisons of alternative numerical algorithms. *Physical Review E*, 71(6), 1-14. doi:10.1103/PhysRevE.71.066208
- Channell, J. (1981). Applying semantic theory to vocabulary teaching. *English Language Teaching Journal*, 35(2), 115-122. doi:10.1093/elt/XXXV.2.115
- Chen, H. (2010). Contrastive learner corpus analysis of epistemic modality and interlanguage pragmatic competence. *Arizona Working Papers in SLA & Teaching*, 17, 2-51. Retrieved from <https://pdfs.semanticscholar.org/aeaa/fde2a1672ff5105ebc82ad465930cb6df6e2.pdf>

- Choi, S. (2016). Processing and learning of enhanced English collocations: An eye movement study. *Language Teaching Research*. doi:10.1177/1362168816653271
- Chomsky, N. (1976). *Reflections on Language*. London, UK: Fontana.
- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36(1), 1-22. Retrieved from <http://www.mitpressjournals.org/doi/pdf/10.1162/0024389052993655>
- Christiansen, M., & Kirby, S. (2003). Language evolution: Consensus and controversies. *Trends in Cognitive Sciences*, 7(7), 300-307. doi:10.1016/S1364-6613(03)00136-0
- Churchill, E. (2007). A dynamic systems account of learning a word: From ecology to form relations. *Applied Linguistics*, 29(3), 339-358. doi:10.1093/applin/amm019
- Clear, J. (1993). From Firth principles: Computational tools for the study of collocation. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair* (pp. 271-292). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. *Paper presented at the Annual Conference of the British Educational Research Association*, University of Exeter, England.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London, UK: Routledge.
- Collins English Dictionary (2017). Harper Collins Publishers. Retrieved from <https://www.collinsdictionary.com/dictionary/english/lemmatize>
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89. doi:10.1093/applin/amm022
- Cover, T., & Thomas, J. (1991). *Elements of information theory*. New York, NY: John Wiley & Sons, Inc.
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223-235. doi:10.1093/applin/II.3.223
- Cowie, A. (1994). Phraseology. In R. Asher (Ed.), *The encyclopedia of language and linguistics* (pp. 3168-3171). Oxford, UK: Pergamon Press.
- Cowie, A. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. Oxford, UK: Clarendon Press.
- Coxhead, A. (2000). A new academic wordlist. *TESOL Quarterly*, 34(2), 213-238. doi:10.2307/3587951
- Crain, S., & Lillo-Martin, D. (1999). *An introduction to linguistic theory and language acquisition*. Oxford, UK: Blackwell.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11(4), 367-383. doi:10.1017/S0272263100008391

- Cruse, D. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.
- De Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. Abingdon, UK: Routledge.
- De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21. doi:10.1017/S1366728906002732
- De Bot, K., Lowie, W., Thorne, S., & Verspoor, M. (2013). Dynamic Systems Theory as a comprehensive theory of second language development. In M. Mayo, M. Gutierrez-Mangado, & M. Adrián (Eds.), *Contemporary approaches to second language acquisition* (pp. 199-220). Amsterdam, Netherlands: John Benjamins Publishing Company. doi:10.1075/aals.9.13ch10
- De Bot, K., Paribakht, T., & Wesche, M. (1997). Toward a lexical processing model for the study of second language vocabulary acquisition: Evidence from ESL reading. *Studies in Second Language Acquisition*, 19(3), 309-29. Retrieved from http://journals.cambridge.org/abstract_S0272263197003021
- Dinçkan, Y. (2010). Culture-bound collocations in bestsellers: A study of their translations from English into Turkish. *Translators' Journal*, 55(3), 456-473. doi:10.7202/045065ar
- Dörnyei, Z., Durow, V., & Zahran, K. (2004). Individual differences and their effects on formulaic sequence acquisition. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 87-106). Amsterdam, Netherland: John Benjamins Publishing Company.
- Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty, & M. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Oxford, UK: Blackwell Publishing.
- Doughty, C., & Long, M. (Eds.). (2005). *The Handbook of Second Language Acquisition*. *Language Acquisition*. Blackwell Publishing. Blackwell reference Online. doi: 10.1111/b.9781405132817.2005.00002.x
- Duan, M., & Qin, X. (2012). Collocation in English teaching and learning. *Theory and Practice in Language Studies*, 2(9), 1890-1894. doi:10.4304/tpls.2.9.1890-1894
- Dupuy, H. (1974). The rationale, development and standardization of a basic word vocabulary test. *Vital and Health Series*, 2(60), 1-85. Washington, D.C.: U.S. Government Printing Office. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/25102343>
- Durrant, P. (2008). *High-frequency collocations and second language learning*. (Doctoral thesis, University of Nottingham). Retrieved from <http://etheses.nottingham.ac.uk/622/>
- Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations. A meta-analysis. *International Journal of Corpus Linguistics*, 19(4), 443-477. doi:10.1075/ijcl.19.4.01dur
- Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6(2), 125-155. doi:10.1515/CLLT.2010.006

- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177. doi:10.1515/iral.2009.007
- Durrant, P., & Schmitt, N. (2010). Adults learners' retention of collocations from exposure. *Second Language Research*, 26(2), 163-188. doi:10.1177/0267658309349431
- Dziemianko, A. (2010). Paper or electronic? The role of dictionary form in language reception, production and the retention of meaning and collocations. *International Journal of Lexicography*, 23(3), 257-273. doi:10.1093/ijl/ecp040
- Dziemianko, A. (2011). Does dictionary form really matter? In K. Akasu, & U. Satoru (Eds.), *ASIALEX 2011 Proceedings. Lexicography: Theoretical and Practical Perspectives* (pp. 92-101). Kyoto, Japan: Asian Association for Lexicography. Retrieved from http://www.academia.edu/3194420/Does_dictionary_form_really_matter
- Ebrahimi-Bazzaz, F., Abd Samad, A., Ismail, I., & Noordin, N. (2014). Verb-noun collocation proficiency and academic years. *International Journal of Applied Linguistics & English Literature*, 3(1), 152-162. doi:10.7575/aiac.ijalel.v.3n.1p.152
- Ebrahimi, S., & Toosi, F. (2013). A comparative analysis of lexical collocations in Molavi's Mathnavi and its translation. *International Journal of English and Education*, 2(2), 62-73. Retrieved from http://ijee.org/yahoo_site_admin/assets/docs/6.89113220.pdf
- Ellis, N. (1998). Emergentism, connectionism, and language learning. *Language Learning*, 48(4), 631-664. doi:10.1111/0023-8333.00063
- Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24(2), 143-188. doi:10.1017/S0272263102002024
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.
- Ellis, R., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375-396. doi:10.1002/j.1545-7249.2008.tb00137.x
- Erman, B. (2007). Cognitive processes as evidence of the idiom principle. *International Journal of Corpus Linguistics*, 12(1), 25-53. doi:10.1075/ijcl.12.1.04erm
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Interdisciplinary Journal for the Study of Discourse*, 20(1), 29-62. doi:10.1515/text.1.2000.20.1.29
- Evert, S. (2005). *The Statistics of word cooccurrences: Word pairs and collocations*. (Doctoral thesis, University of Stuttgart). doi:10.1073/pnas.141413598
- Fan, M. (2009). An exploratory study of collocational use by ESL students: A task based approach. *System*, 37(1), 110-123. doi:10.1016/j.system.2008.06.004

- Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Review of Applied Linguistics in Language Teaching*, 33(4), 315-331. doi:10.1515/iral.1995.33.4.315
- Farrall, S. (2006). What is qualitative longitudinal research? Papers in social research methods. *Papers in Social Research Methods, Qualitative Series*, 11, 1-25. Retrieved from <http://www.lse.ac.uk/methodology/pdf/QualPapers/Stephen-Farrall-Qual%20Longitudinal%20Res.pdf>
- Farrokh, P. (2012). Raising Awareness of Collocation in ESL/EFL Classrooms. *Journal of Studies in Education*, 2(3), 55-74. doi:10.5296/jse.v2i3.1615
- Firth, J. (1957). *Papers in linguistics*. London, UK: Oxford University Press.
- Fitzpatrick, T. (2012). Tracking the changes: Vocabulary acquisition in the study abroad context. *The Language Learning Journal*, 40(1), 81-98. doi:10.1080/09571736.2012.658227
- Foster, P., & Skehan, P. (1999). The influence of source of planning on task-based performance. *Studies in Second Language Acquisition*, 3(3), 215-247. doi:10.1177/136216889900300303
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21(3), 354-375. doi:10.1093/applin/21.3.354
- Gaballa, H., & Al-Khayri, M. (2014). Testing collocational knowledge of Taif University English seniors. *Journal of Humanities and Social Science*, 19(11), 63-90. Retrieved from <http://www.iosrjournals.org/iosr-jhss/papers/Vol19-issue11/Version-3/K191136390.pdf>
- Gabrielatos, C., & Language, E. (2005). Corpora and Language Teaching: Just a fling or wedding bells? *Teaching English as a Second or Foreign Language-Electronic Journal*, 8(4), 1-39. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1068106.pdf>
- Gaines, R. (1969). Linear and nonlinear models of the human controller. *International Journal of Man-Machine Studies*, 1(4): 333-360. doi:10.1016/S0020-7373(69)80001-5
- Gallego, M., & Llach, M. (2009). Exploring the increase of receptive vocabulary knowledge in the foreign language: A longitudinal study. *International Journal of English Studies*, 9(1), 113-133. Retrieved from <http://revistas.um.es/ijes/article/view/90681>
- Ganji, M. (2012). On the effect of gender and years of instruction on Iranian EFL learners' collocational competence. *English Language Teaching*, 5(2), 123-133. doi:10.5539/elt.v5n2p123
- Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language Pedagogy. *ELT Journal*, 55(3), 238-246. doi:10.1093/elt/55.3.238
- Gitsaki, C. (1996). *The development of ESL collocational knowledge*. (Doctoral dissertation, University of Queensland). Retrieved from <http://espace.library.uq.edu.au/view/UQ:205374>
- Gitsaki, C. (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco, CA: International Scholars Publications.

- Gledhill, C. (2000). *Collocations in science writing*. Tübingen, Germany: Gunter Narr Verlag. Retrieved from https://www.academia.edu/1997042/Collocations_in_Science_Writing
- Goldstein, H. (1968). Longitudinal studies and the measurement of change. *The Statistician*, 18(2), 93-117. Retrieved from <http://www.bristol.ac.uk/cmm/team/hg/full-publications/1968/longitudinal-studies-measurement-of-change.pdf>
- González Fernández, B. & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics*, 166(1), 94-126. doi:10.1075/itl.166.1.03fer
- Goulden, R., Nation, I., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4): 341-363. doi:10.1093/applin/11.4.341
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145-160). Oxford, UK: Clarendon Press.
- Granger, S., & Paquot, M. (2008). Disentangling the phraseological web. In S. Granger, & F. Meunier (Eds.), *Phraseology in foreign learning and teaching* (pp. 27-49). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Greenbaum, S. (1974). Some verb-intensifier collocations in American and British English. *American Speech*, 49(1/2), 79-89. doi:10.2307/3087920
- Groot, F. (2012). Intra-individual learner variation in English L2 writing: A case study of complexity, and accuracy development in intensive English language programs. *FLLT Journal*, 1(1), 49-69. Retrieved from http://ejournal.litu.tu.ac.th/pdf/FLLT_Journal_Volume_1.pdf
- Gyllstad, H. (2007). *Testing English collocations: Developing receptive tests for use with advanced Swedish learners*. (Doctoral dissertation, Lund University). Retrieved from <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=599011&fileId=2172422>
- Gyllstad, H. (2013). Looking at L2 vocabulary knowledge dimensions from an assessment perspective: Challenges and potential solutions. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use* (pp.11-28). EUROSLA Monographs Series 2. Retrieved from www.eurosla.org/monographs/EM02/EM02tot.pdf
- Gyllstad, H., & Wolter, B. (2015). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, 66(2), 296-323. doi:10.1111/lang.12143
- Halliday, M. (1966). Lexis as a linguistic level. In C. Bazell, J. Catford, M. Halliday, & R. Robins (Eds.), *In memory of J.R. Firth* (pp. 148-162). London, UK: Longmans.
- Hama, H. (2010). *Major sources of collocational errors made by EFL learners at Koya University*. (Master's thesis, Bilkent University). Retrieved from <http://www.thesis.bilkent.edu.tr/0003969.pdf>

- Hamilton, H. (2012). The efficacy of dictionary use while reading for learning new words. *American Annals of the Deaf*, 157(4), 358-372. Retrieved from <http://www.files.eric.ed.gov/fulltext/ED535143.pdf>
- Hase, S. (2014). Learning is non-linear. Why not curriculum? Retrieved from <http://teachthought.com/learning/learning-non-linear-curriculum/>
- Hauser, M., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298(5598), 1569-1579. Retrieved from www.chomsky.info/articles/20021122.pdf
- Hausmann, F.-J. (1984). Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortwendungen. *Praxis des neusprachlichen Unterrichts*, 31(1), 395-406.
- Hausmann, F. (1989). Le dictionnaire de collocations. In F. Hausmann, O. Reichmann, H. Wiegand, & L. Zgusta (Eds.), *Wörterbücher : Ein internationales handbuch zur lexicographie. Dictionaries, dictionnaires* (pp. 1010-1019). Berlin, Germany: De Gruyter.
- Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, 21(2), 303-317. Retrieved from <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/757C3EE5BBA3DDD5110C77259E4E6E40/S0272263199002089a.pdf/div-class-title-three-dimensions-of-vocabulary-development-div.pdf>
- Henriksen, B. (2013). Research on L2 learners' collocational competence and development: A progress report. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use* (pp. 29-56). EUROSLA Monographs Series 2. Retrieved from <http://www.eurosla.org/monographs/EM02/EM02tot.pdf>
- Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (Ed.), *Teaching Collocation: Further developments in the Lexical Approach* (pp.47-69). Hove, UK: Language Teaching Publications.
- Hoey, M. (1991). *Patterns of lexis in text*. Oxford, UK: Oxford University Press.
- Hoey, M. (2003). Lexical priming and the properties of text. In A. Partington, J. Morley, & L. Haarman (Eds.) *Corpora and Discourse* (pp. 1-15). Bern, Switzerland: Peter Lang. Retrieved from <http://www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm>
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Abingdon, UK: Routledge.
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tübingen, Germany: Niemeyer.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44. doi:10.1093/applin/19.1.24
- Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161-186). Oxford, UK: Clarendon Press.

- Hu, H. (2013). The effects of word frequency and contextual types on vocabulary acquisition from extensive reading: A case study. *Journal of Language Teaching and Research*, 4(3), 487-495. doi:10.4304/jltr.4.3.487-495
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.
- Hunston, S. (2003). Lexis, word form and complementation pattern: A corpus study. *Functions of Language*, 10(1), 31-60. doi:10.1075/fol.10.1.03hun
- Hulstijn, J. (1997). Mnemonic methods in foreign language vocabulary learning: Theoretical considerations and pedagogical implications. In J. Coady, & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 203-224). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139524643.015
- Jabir, J. (2011). The notion of collocation revisited with special reference to the lexical item /nafsa/ (self) in the Holy Quran. *The Arabic Gulf Magazine*, 39(1-2), 13-63. Retrieved from <http://www.iasj.net/iasj?func=fulltext&ald=61659>
- Jaff, D. (2013). *Collocation in English language*. (University paper, Koya University). Retrieved from https://www.academia.edu/3609169/Collocation_In_English_Language
- Jespersen, O. (1922). *Language; its nature, development and origin*. London, UK: G. Allen & Unwin, Ltd.
- Jule, A. (2008). *A beginner's guide to language and gender* (Vol. 1). Clevedon, UK: Multilingual Matters Ltd.
- Kameli, S., Mostapha, G., & Baki, R. (2012). The influence of formal language learning environment on vocabulary learning strategies. *Journal of Language Teaching and Research*, 3(1), 23-29. doi:10.4304/jltr.3.1.23-29
- Karan, M., Snajder, J., & Basic, B. (2012). Evaluation of classification algorithms and features for collocation extraction in Croatian. *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (pp. 657-662). Istanbul, Turkey: Ministry of Science, Education and Sports, Republic of Croatia. Retrieved from http://lrec.elra.info/proceedings/lrec2012/pdf/796_Paper.pdf
- Kellerman, E. (1978). Giving learners a break: Native language intuitions as a source of predictions about transferability. *Working Papers on Bilingualism*, 15, 59-92. Retrieved from <http://eric.ed.gov/?id=ED153469> (ED153469)
- Kennedy, G. (2003). Amplifier collocations in the British National Corpus: Implications for English language. *TESOL Quarterly*, 37(3), 467-487. doi:10.2307/3588400
- Keshavarz, H., & Salimi, H. (2007). Collocational competence and cloze test performance: A study of Iranian EFL learners. *International Journal of Applied Linguistics*, 17(1), 81-92. doi:10.1111/j.1473-4192.2007.00134.x
- Kilgarriff, A., & Kosem, I. (2013). Corpus tools for lexicographers. In S. Granger, & M. Paquot (Eds.), *Electronic lexicography* (pp.32-56). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199654864.003.0003
- Kim, D. (2008). *A study on the use of lexical collocations of Korean heritage learners: Identifying the sources of errors*. (Master's thesis, University of

- Southern California). Retrieved from <http://search.proquest.com/docview/304488320?accountid=10792>
- Kirsner, K. (1994). Second language vocabulary learning: The role of implicit processes. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 283-311). London, UK: Academic Press.
- Kjellmer, G. (1982). Some problems relating to the study of collocations in the Brown Corpus. In S. Johansson (Ed.), *Computer corpora in English language research* (pp. 25-33). Bergen, Norway: Norwegian Computing Centre for the Humanities.
- Kjellmer, G. (1984). Some thoughts on collocational distinctiveness. In J. Aarts, & W. Meijs (Eds.), *Corpus linguistics: Recent developments in the use of computer corpora in English language research* (pp.163-171). Amsterdam, Netherland: Rodopi.
- Kjellmer, G. (1991). A mint of phrase. In K. Aijmer, & B. Altenberg (Eds), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 111-127). London, UK: Longman.
- Knowles, G., & Don, Z. (2004). The notion of a “lemma”: Headwords, roots and lexical sets. *International Journal of Corpus Linguistics*, 9(1), 69-81. doi:10.1075/ijcl.9.1.04kno
- Kolesnikova, O., & Gelbukh, A. (2012). Semantic relations between collocations: A Spanish case study. *Revista Signos*, 45(78), 44-59. doi:10.4067/S0718-09342012000100003
- Koya, T. (2005). *The acquisition of basic collocations by the Japanese learners of English*. (Doctoral dissertation, The Graduate School of Waseda University). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8598943>
- Kramsch, C. (Ed.). (2002). *Language acquisition and language socialization: Ecological perspectives*. London, UK: Continuum.
- Kramsch, C. (2012). Why is everyone so excited about complexity theory in applied linguistics? *Mélanges CRAPEL*, 33, 9-24. Retrieved from <http://www.atilf.fr/IMG/pdf/02.pdf>
- Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The Modern Language Journal*, 73(4), 440-464. Retrieved from <http://www.jstor.org/stable/326879>
- Kristen, C., Römmer, A., Müller, W., & Kalter, F. (2005). Longitudinal studies for education reports: European and North American examples. *Education Reform, 10*. Berlin: German Federal Ministry of Education and Research. Retrieved from http://web.archive.org/web/20130515160142/http://www.bmbf.de/pub/bildungsreform_band_10_eng.pdf
- Kurosaki, S. (2012). *An analysis of the knowledge and use of English collocations by French and Japanese learners*. (Doctoral dissertation, University of London). Retrieved from https://repository.royalholloway.ac.uk/file/d65f4c4b-8bac-d738-3735-83b5e5e031a7/8/Shino_K._Thesis_20121015.pdf
- Langacker, R. (2000). A dynamic usage-based model. In M. Barlow, & S. Kemmer (Eds.), *Usage-based models of language* (pp.1-63). Stanford, CA: CSLI.

- Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 14-65. Retrieved from www.uibk.ac.at/anglistik/staff/freeman/course.../diane_chaos_paper.pdf
- Larsen-Freeman, D. (2002). Language acquisition and language use from a chaos/complexity theory perspective. In C. Kramsch (Ed.), *Language acquisition and language socialization: Ecological perspectives* (pp. 33-46). London, UK: Continuum.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619. doi:10.1093/applin/aml029
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford, UK: Oxford University Press.
- Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In M. McCarthy, & N. Schmitt (Eds.), *Vocabulary description, acquisition and pedagogy* (pp.140-155). Cambridge, UK: Cambridge University Press.
- Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics*, 19(2), 255-271. doi:10.1093/applin/19.2.255
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436. doi:10.1111/j.0023-8333.2004.00260.x
- Laufer, B., & Nation, I. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322. doi:10.1093/applin/16.3.307
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672. doi: <https://doi.org/10.1111/j.1467-9922.2010.00621.x>
- Laval, V., & Bernicot, J. (2002). Tu es dans la lune: Understanding idioms in French-speaking children and adults. *Pragmatics*, 12(4), 399-413. doi:10.1075/prag.12.4.01lav
- Leedham, M. (2011). *A corpus-driven study of features of Chinese students' undergraduate writing in UK universities*. (Doctoral dissertation, The Open University). Retrieved from http://oro.open.ac.uk/29228/1297/Leedham_Thesis.pdf
- Leeuwen, M. (2005). Questions for the Dynamicist: The use of Dynamical Systems Theory in the philosophy of cognition. *Minds and Machines*, 15(3-4), 271-333. doi:10.1007/s11023-004-8339-2
- Léon J. (2007). Meaning by collocation. The Firthian filiation of Corpus Linguistics. In D. Kibbee (Ed.), *Proceedings of ICHoLS X, 10th International Conference on the History of Language Sciences* (pp. 404-415). Amsterdam, Netherland: John Benjamins Publishing Company.
- Lerner, R., & Kaufman, M. (1985). The concept of development in contextualism. *Developmental Review*, 5(4), 309-333. doi:10.1016/0273-2297(85)90016-4

- Lesniewska, J. (2006). Collocations and second language use. *Studia Lingüística*, 123, 95-105. Retrieved from <http://www.filg.uj.edu.pl/documents/41616/4333138/12306-Lesniewska.pdf>
- Levelt, W. (1993). The architecture of normal spoken language use. In G. Blanken, J. Dittmann, H. Grimm, J. Marshall, & C. Wallesch (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Berlin, Germany: Walter de Gruyter. Retrieved from <http://pubman.mpdl.mpg.de/pubman/faces/viewItemOverviewPage.jsp?itemId=escidoc:64342>
- Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Hove, UK: Language Teaching Publications.
- Lewis, M. (1997). *Implementing the lexical approach*. Hove, England: Language Teaching Publications.
- Lewis, M. (2000). *Teaching collocation: Further development in the lexical approach*. Boston, MA: Thomson Heinle.
- Lewis, M., & Hill, J. (Eds.). (1997). *A dictionary of selected collocations*. Hove, UK: Language Teaching Publications.
- Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing*, 18(2), 85-102. doi:10.1016/j.jslw.2009.02.001
- Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 22-46). London, UK: Continuum.
- Li, W., & Lu, Q. (2011). A hybrid extraction model for Chinese noun/verb synonymous bi-gram collocations. *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation* (pp. 430-439). Hong Kong, China: The Hong Kong Polytechnic University. Retrieved from <http://aclweb.org/anthology/Y11-1045>
- Lindstromberg, S., & Eyckmans, J. (2014). How big is the positive effect of assonance on the recall of L2 collocations? *International Journal of Applied Linguistics*, 165(1), 19-45. doi:10.1075/itl.165.1.02lin
- Liu, L. (2002). *A corpus-based lexical semantic investigation of verb-noun miscollocations in Taiwan learners' English*. Tamkang University, Taiwan: Graduate Institute of Western Languages and Literature.
- Lowie, W. (2012). Dynamic systems theory approaches to second language acquisition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1806-1813). Groningen, Netherlands: Blackwell Publishing Ltd. doi:10.1002/9781405198431.wbeal0346
- Lowie, W., Caspi, T., Van Geert, P., & Steenbeek, H. (2011). Modeling development and change. In M., Verspoor, K., de Bot, & W Lowie (Eds.), *A dynamic approach to second language development: methods and techniques* (pp. 99-122). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Lowie, W., Verspoor, M., & de Bot, K. (2009). A dynamic view of second language development across the lifespan. In K. de Bot, & R. Schrauf (Eds.),

- Language development over the lifespan* (pp. 125-146). New York, NY: Routledge.
- Macis, M., & Schmitt, N. (2017). The figurative and polysemous nature of collocations and their place in ELT. *ELT Journal*, 71(1), 50-59. doi:10.1093/elt/ccw044
- Mahmoud, A. (2005). Collocational errors made by Arab learners of English. In P. Robertson, & J. Jung (Eds.), *The Asian EFL Journal Quarterly: Professional teaching articles collection* (pp.117-126). Retrieved from <http://asian-efl-journal.com/1912/teaching-articles/2005/07/collocation-errors-made-by-arab-learners-of-english/#squelch-taas-tab-content-0-3p>
- Mahvelati, E., & Mukundan, J. (2012). The effects of input flood and consciousness-raising approach on collocation knowledge development of language learners. *International Journal of Applied Linguistics & English Literature*, 1(6), 182-192. doi:10.7575/ijalel.v.1n.6p.182
- Manning, C., & Schütze, H. (1999). *Foundations of statistical natural language processing*. Cambridge, UK: MIT Press.
- Marin, M., & Peltzer-Karpp, A. (2009). Towards a dynamic systems approach to the development of language and music. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*(pp. 284-292). Jyväskylä, Finland: University of Jyväskylä. Retrieved from https://jyx.jyu.fi/dspace/bitstream/handle/123456789/20893/urn_nbn_fi_jyu-2009411277.pdf?sequence=1
- Martinez, R., & Murphy, V. (2011). Effect of frequency and idiomaticity in second language reading comprehension. *TESOL Quarterly*, 45(2), 267-290. doi:10.5054/tq.2011.247708
- Marton, W. (1977). Foreign vocabulary learning as problem No. 1 of language teaching at the advanced level. *Interlanguage Studies Bulletin*, 2(1), 33-57. <http://www.jstor.org/stable/pdf/43135158.pdf>
- Mazaji, B. (2015). Effects of digital games on vocabulary acquisition of Iranian low proficiency male and female EFL learners. *International Journal of Foreign Language Teaching and Research*, 3(10), 83-92. Retrieved from http://jfl.iaun.ac.ir/article_12543_3316cb9f6963ef71bd20daec9f0455a4.pdf
- McKeown, K., & Radev, D. (2000). Collocations. In R. Dale, H. Moisl, & H. Somers (Eds.), *A handbook of natural language processing* (pp. 507-523). New York, NY: Marcel Dekker. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=74B89DC1F30CAB9DC26DC93317CAEB20?doi=10.1.1.16.2814&rep=rep1&type=pdf>
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. London, UK: Routledge.
- Meara, P. (1996). The vocabulary knowledge framework. *Lognostics: Tools for vocabulary researchers*. Retrieved from <http://www.lognostics.co.uk/vlibrary/meara1996c.pdf>
- Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research*, 18(4), 393-407. doi:10.1191/0267658302sr211xx

- Mehrpour, S., Razmjoo, S., & Kian, P. (2011). The relationship between depth and breadth of vocabulary knowledge and reading comprehension among Iranian EFL Learners. *Journal of English Language Teaching and Learning*, 53(222), 97-127. Retrieved from <http://www.ensani.ir/storage/Files/20120427094740-8079-5.pdf>
- Melka, F. (1997). Receptive versus productive aspects of vocabulary. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 84-102). Cambridge, UK: Cambridge University Press.
- Menard, S. (2002). *Longitudinal research* (2nd ed.). Thousand Oaks, CA: Sage.
- Miller, R. (2000). *Researching life family histories*. London, UK: Sage.
- Miller, P. (2011). *Theories of developmental psychology* (5th ed.). New York, NY: Worth Publishers. Retrieved from <http://dl.umsu.ac.ir/bitstream/Hannan/3074/1/9781429216340.pdf>
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters.
- Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research* (pp. 211-232). EUROSLA Monographs Series 1. Retrieved from <http://eurosla.org/monographs/EM01/211-232Milton.pdf>
- Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 Vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis* (pp. 57 -78). EUROSLA Monographs Series 2. Retrieved from <http://www.eurosla.org/monographs/EM02/Milton.pdf>
- Mohajeri, M., & Ketabi, S. (2013). The knowledge and use of collocations and their relation with English speaking proficiency among upper- intermediate to advanced Iranian EFL learners. *ELT Voice-India*, 3(5), 2230-9136. Retrieved from http://eltvoices.in/Volume3/Issue_5/EVI_35_1.pdf
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford, UK: Clarendon Press.
- Mounya, A. (2010). *Teaching lexical collocations to raise proficiency in foreign language writing*. (Master's thesis, University of Mentouri). Retrieved from <http://bu.umc.edu.dz/theses/anglais/ABD1089.pdf>
- Mueller, C. (2011). English learners' knowledge of prepositions: Collocational knowledge or knowledge based on meaning? *System*, 39(4), 480-490. doi:10.1016/j.system.2011.10.012
- Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning*, 48(3), 323-363. Retrieved from <https://cdn.auckland.ac.nz/assets/.../about/.../rote-or-rule-0023-8333.pdf>
- Nagy, W., & Herman, P. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M. McKeown, & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 19-35). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Nagy, W., & Scott, J. (1990). Word schemas: Expectations about the form and meaning of new words. *Cognition and Instruction*, 7(2), 105-127. doi:10.1207/s1532690xci0702_2
- Nation, I. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle Cengage Learning
- Nation, I. (2013). *Learning vocabulary in another language* (2nd ed.). New York, NY: Cambridge University Press.
- Nation, I., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). Cambridge, UK: Cambridge University Press.
- Nattinger, J. (1988). Some current trends in vocabulary teaching. In R. Carter, & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 62-82). London, UK: Longman.
- Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford, UK: 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. doi:10.1093/applin/24.2.223
- Nesselhauf, N. (2004). What are collocations? In D. Allerton, N. Nesselhauf, & P. Skandera (Eds.), *Phraseological units: Basic concepts and their application* (pp. 1–21). Basel, Switzerland: Schwabe.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Neuman, Y., Assaf, D., Cohen, Y., Last, M., Argamon, S., Howard, N., & Frieder, O. (2013). Metaphor identification in large texts corpora. *PLoS ONE*, 8(4), 1-9. doi:10.1371/journal.pone.0062343
- Nguyen, H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298-320. doi:10.1177/1362168816639619
- Nizonkiza, D. (2015). Measuring receptive collocational competence across proficiency levels. *Stellenbosch Papers in Linguistics*, 44(1983), 125-146. doi:10.5774/44-0-186
- Nizonkiza, D., Van Dyk, T., & Louw, H. (2013). First-year university students' productive knowledge of collocations. *Stellenbosch Papers in Linguistics Plus*, 42(1), 165-181. doi:10.5842/42-0-143
- O'Grady, W. (1997). *Syntactic development*. Chicago, IL: University of Chicago Press.
- Ortega, L., & Byrnes, H. (Eds.). (2008). *The longitudinal study of advanced L2 capacities*. London, UK: Routledge.
- Otero, P. (2008). The meaning of syntactic dependencies. *Linguistik Online*, 35(3), 33-53. Retrieved from www.linguistik-online.com/35_08/gamallo.pdf
- Oxford Dictionary of English. (2015). MacBook Pro.

- Pace-Sigge, M. (2013). The concept of lexical priming in the context of language use. *ICAME Journal*, 37, 149-173. Retrieved from http://clu.uni.no/icame/ij37/Pages_149-174.pdf
- Palmer, H. (1966). *Second interim report on English collocations, submitted to the Tenth Annual Conference of English Teachers, under the auspices of the Institute for Research in English Teaching, Dept. of Education, Tokyo, October 1933*. Tokyo, Japan: Kaitakusha.
- Papachristos, G. (2012). Case study and system dynamics research: Complementarities, pluralism and evolutionary theory development. *Proceeding of the 30th International Conference of the System Dynamics Society* (pp. 2720-2748). *St. Gallen, Switzerland*. Retrieved from www.systemdynamics.org/conferences/2012/proceed/papers/P1125.pdf
- Pawley, A. (2007). Developments in the study of formulaic language since 1970: A personal view. In P. Skandera (Ed.), *Phraseology and culture in English* (pp. 3-34). Berlin, Germany: Mouton de Gruyter.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191-227). London, UK: Longman.
- Pearce, D. (2002). A comparative evaluation of collocation extraction techniques. *Proceeding of the 3rd International Conference on Language Resources and Evaluation (LREC)*(pp.1530-1536), Las Palmas, Canary Islands, Spain. Retrieved from <http://www.lrec-conf.org/proceedings/lrec2002/pdf/169.pdf>
- Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, 18(1),1-28. Retrieved from <http://files.eric.ed.gov/fulltext/EJ759833.pdf>
- Pignot-Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge. *Language Studies Working Papers*, 4, 37-45. Retrieved from http://wap.rdg.ac.uk/web/files/english-language-and-literature/elal_LSWP_Vol_4_Pignot_Shahov.pdf
- Plumert, J., & Spencer, J. (2007). *The emerging spatial mind*. Oxford, UK: Oxford University Press.
- Poulsen, S. (2005). *Collocations as a language resource: A functional and cognitive study in English phraseology*. (Doctoral thesis, University of Southern Denmark). Retrieved from http://static.sdu.dk/mediafiles/Files/Om_SDU/Fakulteterne/Humaniora/Phd/afhandlinger/2005/Afhandling%2048_s_poulsen%20pdf.pdf
- Qian, D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, 52(3), 513-536. doi:10.1111/1467-9922.00193
- Rajulton, F. (2001). The fundamentals of longitudinal research: An overview. *Canadian Studies in Population*, 28(2), 169-185. Retrieved from <https://ejournals.library.ualberta.ca/index.php/csp/article/viewFile/15869/12674>
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

- Read, J. (2007). Second language vocabulary assessment: Current practices and new directions. *International Journal of English Studies*, 7(2), 105-125. Retrieved from <http://files.eric.ed.gov/fulltext/EJ1072194.pdf>
- Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 128-143). London, UK: Routledge.
- Tabrizi, N., & Moghadam, F. (2016). The effect of collocation instruction via Short Message Service (SMS) on the acquisition of collocations by Iranian intermediate EFL learners: With a focus on gender. *International Journal of Humanities and Cultural Studies*, 3(1), 1339-1359. Retrieved from <http://www.ijhcs.com/index.php/ijhcs/article/view/1425/1298>
- Robinson, P., & Ellis, N. (2008). *Handbook of cognitive linguistics and second language acquisition*. New York, NY: Routledge. Retrieved from https://www.academia.edu/6633691/HANDBOOK_OF_COGNITIVE_LINGUISTICS_AND_SECOND_LANGUAGE_ACQUISITION
- Roberts, W. (1986). Nonlinear models of development: An example from the socialization of competence. *Child Development*, 57(5), 1166-1178. <http://faculty.tru.ca/wlroberts/nonlinear.pdf>
- Sadeghi, Z. (2010). The importance of collocation in vocabulary teaching and learning. *Translation Journal*, 14(2), 1-24. Retrieved from <http://translationjournal.net/journal/52collocation.htm>
- Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319-353). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139524780.013
- Schilk, M. (2011). *Structural nativization in Indian English lexicogrammar*. Philadelphia, PA: John Benjamins Publishing Company.
- Schmid, J. (2003). Collocation: Hard to pin down, but bloody useful. *Journal of English and American Studies*, 51(3), 235-58. Retrieved from http://www.anglistik.uni-muenchen.de/personen/professoren/schmid/schmid_publ/collocation.pdf
- Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt, & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (1998a). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning*, 48(2), 281-317. Retrieved from [http://www.norbertschmitt.co.uk/uploads/schmitt-n-\(1998\)-tracking-the-incremental-acquisition-of-second-language-vocabulary-a-longitudinal-study-language-learning-48-2-281-317.pdf](http://www.norbertschmitt.co.uk/uploads/schmitt-n-(1998)-tracking-the-incremental-acquisition-of-second-language-vocabulary-a-longitudinal-study-language-learning-48-2-281-317.pdf)
- Schmitt, N. (1998c). Quantifying word association responses: What counts as native-like? *System*, 26(3), 389-401. doi:10.1016/S0346-251X(98)00019-0
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, UK: Cambridge University Press.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan.

- Schmitt, N., & Carter, R. (2000). Lexical phrases in language learning, *The Language Teacher*, 24(8), 6-10. Retrieved from http://jalt-publications.org/tlt/issues/2000-08_24.8
- Schmitt, N., Dörnyei, Z., Adolphs, S., & Durow, V. (2004). Knowledge and acquisition of formulaic sequences: A longitudinal study. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 55-86). Amsterdam, Netherland: John Benjamins Publishing Company.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition*, 19(1), 17-36. doi:10.1017/S0272263197001022
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Level Test. *Language Testing*, 18(1), 55-88. doi:10.1177/026553220101800103
- Schwartz, H., Eichstaedt, J., Kern, M., Dziurzynski, L., Ramones, S., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, Martin E., & Ungar, L. (2013). Personality, gender, and age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE*, 8(9), 1-16. doi:10.1371/journal.pone.0073791
- Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26(2), 173-199. doi:10.1017/S0272263104262027
- Seretan, V. (2011). *Syntax-based collocation extraction: Text, speech and language technology series* (Vol. 44). Geneva, Switzerland: Springer Science and Business Media. doi:10.1007/978-94-007-0134-2_4
- Seretan V., Nerima, L., & Wehrli, E. (2003). Extraction of multi-word collocations using syntactic bigram composition. *Proceedings of International Conference on Recent Advances in NLP* (pp. 131-138). Borovets, Bulgaria. Retrieved from www.latl.unige.ch/personal/vseretan/publ/ranlp2003_vs_ln_ew.pdf
- Seretan V., Nerima, L., & Wehrli, E. (2004). A tool for multi-word collocation extraction and visualization in multilingual corpora. *Proceedings of the Eleventh EURALEX International Congress* (pp. 755-766). Lorient, France. Retrieved from http://www.latl.unige.ch/personal/vseretan/publ/EURALEX2004_VS_LN_EW.pdf
- Shehata, A. K. (2008). *L1 Influence on the reception and production of collocations by advanced ESL/EFL Arabic learners of English*. (Master's thesis, Ohio University). Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=ohiou1218237449&disposition=inline
- Shen, Z. (2008). The roles of depth and breadth of vocabulary knowledge in EFL reading performance. *Asian Social Science*, 4(12), 135-137. doi:10.5539/ass.v4n12p135
- Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62(4), 339-348. Retrieved from <http://eltj.oxfordjournals.org/content/62/4/339.full.pdf>. doi:10.1093/elt/ccm091

- Shooshtari, Z., & Karami, N. (2013). Lexical collocation instruction and its impact on Iranian non-academic EFL learners' speaking ability. *Journal of Language Teaching and Research*, 4(4), 767-776. doi:10.4304/jltr.4.4.767-776
- Shokouhi, H., & Mirsalari, G. (2010). Collocational knowledge versus general linguistic knowledge among Iranian EFL learners. *Teaching English as a Second or Foreign Language*, 14(4), 1-24. Retrieved from <http://tesl-ej.org/pdf/ej52/a7.pdf>
- Simpson-Vlach, R., & Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. doi:10.1093/applin/amp058
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
- Sinclair, J. (1996). The search for units of meaning. In J. Sinclair (2004, reprinted), *Trust the text: Language, corpus and discourse* (pp. 24-48). London, UK: Routledge.
- Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter, & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp.140-158). Harlow, UK: Longman.
- Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London, UK: Routledge.
- Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, 53, 148-160. doi:10.1016/j.system.2015.07.003
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. (2011). Seeing a phrase 'time and again' matters: The role of phrasal frequency in the processing of multi-word sequences. *Journal of Experimental Psychology: Language, Memory, and Cognition*, 37(3), 776-784. doi:10.1037/a0022531
- Siyanova-chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, 36(5), 549-569. doi:10.1093/applin/amt054
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review*, 64(3), 429-458. doi:10.3138/cmlr.64.3.429
- Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1),143-177. Retrieved from <http://www.anthology.aclweb.org/J/J93/J93-1007.pdf>
- Smith, L. (2009). Dynamic systems, sensorimotor processes, and the origins of stability and flexibility. In J. Spencer (Ed.), *Toward a unified theory of development Connectionism and Dynamic System Theory re-consider* (pp. 68-83). Oxford Scholarship Online: Oxford University Press. doi:10.1093/acprof:oso/9780195300598.003.0004
- Smith, P., & Torrey, B. (1996). The future of the behavioral and social sciences. *Science*, 271(5249), 611-612. doi:10.1126/science.271.5249.611
- Spencer, J., Clearfield, M., Corbetta, D., Ulrich, B., Buchanan, P., & Schöner, G. (2006). Moving toward a grand theory of development: In memory of Esther Thelen. *Child Development*, 77(6), 1521-1538. doi:10.1111/j.1467-8624.2006.00955.x

- Spencer, J., Perone, S., & Buss, A. (2011). Twenty years and going strong: A dynamic systems revolution in motor and cognitive development. *Child Development Perspectives*, 5(4), 260-266. doi:10.1111/j.1750-8606.2011.00194.x
- Spivey, M. (2007). *The continuity of mind*. Oxford, UK: Oxford University Press.
- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(4), 532-553. doi:10.1093/applin/amq001
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, 2(1), 23-55. Retrieved from <http://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf>
- Stubbs, M. (1996). *Text and corpus analysis: Corpus-assisted studies of language and culture*. Oxford, UK: Blackwell.
- Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford, UK: Blackwell Publishing.
- Szudarski, P., & Carter, R. (2014). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics*, 26(2), 245-265. doi:10.1111/ijal.12092
- Tabrizi, A., & Moghadam, F. (2016). The effect of collocation instruction via Short Message Service (SMS) on the acquisition of collocations by Iranian Intermediate EFL learners: With a focus on gender. *International Journal of Humanities and Cultural Studies*, 3(1), 1339-1359. Retrieved from <http://www.ijhcs.com/index.php/ijhcs/index>
- Tahmasebi, G., Ghaedrahmat, M., & Haqverdi, H. (2013). The relationship between language proficiency and Iranian EFL learners' knowledge of vocabulary depth versus vocabulary breadth. *Latin American Journal of Content and Language Integrated Learning*, 6(2), 96-111. doi:10.5294/laclil.2013.6.2.5 eISSN 2322-9721
- Takač, V., & Lukač, M. (2013). How word choice matters: An analysis of adjective-noun collocations in a corpus of learner essays. *Linguistics*, 14(2-3): 385-402. Retrieved from https://www.academia.edu/4006007/How_word_choice_matters_An_analysis_of_adjective-noun_collocations_in_a_corpus_of_learner_essays
- Takahashi, C. (2012). Impact of dictionary use skills instruction on second language writing. *Working Papers in TESOL & Applied Linguistics*, 12(2), 81-103. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.838.4342&rep=rep1&type=pdf>
- Taplin, S. (2005). *Methodological design issues in longitudinal studies of children and young people in out-of-home care: Literature review*. (Research Report). New Department of Community Services. Retrieved from http://www.community.nsw.gov.au/__data/assets/file/0004/358474/methodological_issues_oohc.pdf
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. doi:10.5116/ijme.4dfb.8dfd

- Thelen, E. (1985). Developmental origins of motor coordination: Leg movements in human infants. *Developmental Psychobiology*, 18(1), 1-22. doi:10.1002/dev.420180102
- Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic Dialogues*, 15(2), 255-283. Retrieved from <http://psysc613.wikispaces.com/file/view/Dynamic+Systems+Theory+and+the+Complexity+of+Change.pdf>
- Thelen, E., & Smith, L. (1994). *A Dynamic Systems Approach to the development of cognition and action*. London, UK: MIT Press.
- Thomas, M., McClelland, J., Richardson, F., Schapiro, A., & Baughman, F. (2009). Dynamic and Connectionist approaches to development: Toward a future of mutually beneficial coevolution. In J. Spencer (Ed.), *Toward a unified theory of development Connectionism and Dynamic System Theory re-consider* (pp. 338-350). Oxford, UK: Oxford University Press.
- Todiraşcu, A., Tufiş, D., Heid, U., Gledhill, C., Ştefanescu, D., & Rousselot, F. (2008). A hybrid approach to extracting and classifying verb + noun constructions. *Proceedings of the Sixth International Language Resources and Evaluation (LREC)*(pp.1956-1963). Retrieved from http://lrec-conf.org/proceedings/lrec2008/pdf/500_paper.pdf
- Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M. (2004). What kind of evidence could refute the UG hypothesis?: Commentary on Wunderlich. *Studies in Language*, 28(3), 642-645. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.384.2596&rep=rep1&type=pdf>
- Tomasello, M., & Abbot-Smith, K. (2002). A tale of two theories: Response to Fisher. *Cognition*, 83(2), 207-214. doi:10.1016/S0010-0277(01)00172-X
- Torabian, H., Maros, M., & Subakir, M. (2014). Lexical collocational knowledge of Iranian undergraduate learners: Implications for receptive & productive performance. *Procedia-Social and Behavioral Sciences*, 158, 343-350. doi:10.1016/j.sbspro.2014.12.098
- Trantescu, A. (2010). Semantic transparency and opacity in fixed expressions. *Philological Sciences Series. linguistics*, XXXII (1-2), 375-397. Retrieved from http://www.cis01.central.ucv.ro/litere/activ_st/articole.../trantescu_ana-maria.pdf
- Trofimovich, P. (2011). Language experience in L2 phonological learning: Effects of psycholinguistic and sociolinguistic variables. *International Review of Applied Linguistics in Language Teaching*, 49(2), 135-156. doi:10.1515/iral.2011.007
- Tseng, W., & Schmitt, N. (2008). Toward a model of motivated vocabulary learning: A Structural Equation Modeling Approach. *Language Learning*, 58(2), 357-400. doi:10.1111/j.1467-9922.2008.00444.x
- Tutin, A. (2005). Annotating lexical functions in corpora: Showing collocations in context. In Y. Apresjan, & L. Iomdin (Eds.), *Proceedings of the Second*

- International Conference on Meaning-Text Theory* (pp. 497-502). Moscow, Russia. Retrieved from <http://www.ugrenoble3.fr/tutin/Publis/MTT.pdf>
- Van Geert, P. (2008). The Dynamic Systems approach in the study of L1 and L2 acquisition: An introduction. *Modern Language Journal*, 92(2), 179-199. doi:10.1111/j.1540-4781.2008.00713.x
- Van Geert, P., & Steenbeck, H. (2005). Explaining after by before: Basic aspects of a dynamic systems approach to the study of development. *Developmental Review*, 25(3-4), 408-442. doi:10.1016/j.dr.2005.10.003
- Vechtomova, O., & Wang, Y. (2006). A study of the effect of term proximity on query expansion. *Journal of Information Science*, 32(4), 324-333. doi:10.1177/0165551506065787
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217-234. Retrieved from <https://www.cambridge.org/core/services/aop-cambridge-core/content/view/6A7CD827022283582903ED690EF4EE06/S0142716401002041a.pdf/div-class-title-breadth-and-depth-of-vocabulary-in-relation-to-l1-l2-acquisition-and-frequency-of-input-div.pdf>
- Verspoor, M., De Bot, K., & Lowie, W. (Eds.). (2011). *A dynamic approach to second language development: Methods and techniques* (Vol. 29). Amsterdam, Netherlands: John Benjamins Publishing Company.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231. doi:10.1111/j.1540-4781.2008.00715.x
- Verspoor, M., Schmid, M., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing*, 21(3), 239-263. doi:10.1016/j.jslw.2012.03.007
- Waring, R. (2002). Scales of vocabulary knowledge in second language vocabulary assessment. *Kiyo, The occasional papers of Notre Dame Seishin University*, 46(1), 35-41. Retrieved from <http://www.robwaring.org/papers/various/scales.htm>
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79-95. doi:10.1017/S0272263108080042
- Wei, Y. (1999). Teaching collocations for productive vocabulary development. *The Annual Meeting of the Teachers of English to Speakers of Other Languages*. New York, NY. Retrieved from <http://eric.ed.gov/?id=ED457690>. (ED457690)
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32(4), 430-449. doi:10.1093/applin/amr011
- Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, 35(3), 451-482. doi:10.1017/S0272263113000107
- Wood, D. (2004). An empirical investigation into the facilitating role of automatized lexical phrases in second language fluency development. *Journal*

- of *Language and Learning*, 2(1), 27-50. Retrieved from http://webpace.buckingham.ac.uk/kbernhardt/journal/jllearn/2_2/wood2_2.html
- Wood, D. (2008). Mandarin Chinese speakers in a study abroad context: Does acquisition of formulaic sequences facilitate fluent speech in English? *The East Asian Learner*, 3(2), 43-62.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.
- Wu, W. (1996). Lexical collocations: One way to make passive vocabulary active. *Proceedings of the 11th Conference on English Teaching and Learning in the Republic of China* (pp. 461-480). Taipei, Taiwan: Crane. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.546.2776&rep=rep1&type=pdf>
- Wu, S., Witten, I., & Franken, M. (2010). Utilizing lexical data from a Web-derived corpus to expand productive collocation knowledge. *ReCALL*, 22(1), 83-102. doi:10.1017/S0958344009990218
- Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics*, 27(1), 103-129. doi:10.1093/applin/ami045
- Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly*, 44(4), 647-668. doi:10.5054/tq.2010.235998
- Yazdandoost, Z., AmalSaleh, E., & Kafipour, R. (2014). The relationship among collocation knowledge and listening, speaking, reading and writing proficiency of Iranian EFL learners. *Language, Individual & Society, Journal of International Scientific Publications*, 8, 408-419. Retrieved from <https://www.scientific-publications.net/get/1000006/1408972464185312.pdf>
- Yuksel, I. (2013). ELT majors' cross sectional evaluation of academic lexical competence and performance. *Journal of Language Teaching and Research*, 4(2), 244-252. doi:10.4304/jltr.4.2.244-252
- Yule, G. (2003). *The study of language* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Zaabalawi, R. S., & Gould, A. M. (2017). English collocations: A novel approach to teaching the language's last bastion. *Ampersand*, 4, 21-29. doi:10.1016/j.amper.2017.03.002
- Zhang, X. (1993). *English collocations and their effect on the writing of native and non-native college freshmen*. (Doctoral dissertation, Indiana University of Pennsylvania).
- Zhao, G. (2010). An understanding of language development models: Pidginization from the perspective of Chaos Theory. *English Language Teaching*, 3(2), 88-93. doi: <http://dx.doi.org/10.5539/elt.v3n2p88>
- Zheng, Y. (2014). The fluctuating development of cross-linguistic semantic awareness: A longitudinal multiple-case study. *Language Awareness*, 23(4), 369-388. doi:10.1080/09658416.2013.864298
- Zhong, H. (2011). Learning a word: From receptive to productive vocabulary use. *The Asian Conference on Language Learning Official Proceedings* (pp.

116-126), Osaka, Japan. Retrieved from http://iafor.org/offprints/acll-split-offprints/ACLL2011_0102.pdf

Zhong, H. (2016). The relationship between receptive and productive vocabulary knowledge: A perspective from vocabulary use in sentence writing. *The Language Learning Journal*, 1736(4), 1-17.
doi:10.1080/09571736.2015.1127403

Software:

- NVivo 10
- Statistical Package for the Social Sciences (SPSS 22).
- The Constituent Likelihood Automatic Word-tagging System (CLAWS 5). Retrieved from <http://ucrel.lancs.ac.uk/claws/trial.html>.