

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Humanities

Languages, Cultures and Linguistics

**Assessment of Collocation Progression in Second Language Learners' Writing at
Different Levels of Proficiency: A Corpus- Based Study**

by

AlShahad Adnan AlDereihim

Thesis for the Degree of Doctor of Philosophy

June 2023

Research Thesis: Declaration of Authorship

Print name: AlShahad Adnan AlDereihim

Title of thesis: Assessment of Collocation Progression in Second Language Learners' Writing at different Levels of Proficiency: A corpus- Based Study.

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- None of this work has been published before submission

Signature:

Date: 18/06/2023

Dedication

To my Loving mother Prof.Siham AlKadhib

Acknowledgements

I would like to thank all those who contributed to accomplishing this study. I give my thanks to my supervisory team Dr. Ying Zheng and Prof. Roumyana Slabakova. A special thanks to my main supervisory team Dr. Ying Zheng for her insightful comments, constant encouragement, patience, and kind support.

I would like to also express my gratitude to the students who participated in this study, and to the University staff and tutors for their cooperation and for giving me access to their classrooms.

I owe my deepest gratitude to my good friends Ghufuran AlOthaim, Batool Bakhsh, Dr. Ashwaq AlThowibi, Dr. Samah Alshehri, Dr. Hind AlEnezi for their support, encouragement, and advice.

Special thanks to the Saudi Cultural Bureau, Princess Norah University, and my sponsor Prince Sattam Bin Abdulaziz University who generously funded this thesis through a scholarship.

Finally, this thesis would have never been possible without the warm-hearted love, support, and encouragement of my parents, my father Mr. Adnan AlDerehim, and my mother Prof. Siham AlKadhib. I am also thankful to my uncle Adel AlDerehim, to all my brothers, sisters-in-law, nieces, and nephews for providing me with love and support throughout this journey.

University of Southampton

Abstract

Faculty of Humanities

Languages, Cultures and Linguistics

Thesis for the Degree of Doctor of Philosophy

Assessment of Collocation Progression in Second Language Learners' Writing at Different Levels of

Proficiency: A Corpus- Based Study

by

AlShahad Adnan AlDereihim

Research on second language learners' collocation knowledge has looked into many areas such as investigating factors that affect collocation learning e.g. (Nguyen and Webb, 2016), which found that there was a close relationship between the participants' knowledge of collocations and single-word items. Also types known by learners as in (Laufer and Waldman, 2011), or looked into group of learners use of collocations at a certain proficiency level e.g. (Sivanova-Chanturia, 2015).

This study intends to contribute to the literature by investigating English language learners of different level of proficiency use of collocation and looks into their language development over time, for this purpose the study focused on finding out the types of collocations learners use, the degree of fixedness of extracted collocations, and the progress students have made across three different study years by using corpus analysis tools. The study took place at a Saudi Arabian University. A total of 807 samples were collected from 269 students for this study at three different times over the course of an academic year. All samples were transcribed and uploaded to an online corpus (Sketch Engine). Overall, verb-noun collocations are the most commonly used collocations in students' writing, accounting for 50% of total collocations. This is followed by adjectival-noun collocations, which account for 41.2% of all collocations. The number of collocations between nouns and noun phrases was the lowest, accounting for only 8% of all collocations. When examining the production of collocations according to the type of collocation and the year of study, it is evident that year one, year two, and year three produced more verb-noun collocations than any other type of collocations (54.3%, 52.7%, and 44.7%, respectively). It is then followed by adjective-noun collocations (40%, 38.2%, and 44.7%), respectively. Noun-noun collocations came last and scored only (5.7%, 9.1%, and 10.5) respectively. According to the results of the analysis of the four levels of fixedness among phases and years, in phase one year 1 most students noted that the adjective + noun collocations were low (n=11, 47.8%), followed by high (n=6, 26.1%) and medium (n=6, 26.1%). In terms of verb+noun collocations, the students

tend to have a medium level (n=12, 52.2%). The average number of adjective + noun collocations found in year 2 was medium (15, 46.7%), low (n=10, 33.3%), and high (n=5, 16.7%). In terms of verb+noun collocations, the students showed a medium level (n=13, 56.6%). Also, most students in phase one of 3 years showed a medium level (n=9, 40.9%), then a low level (n=8, 36.4%) in adjective + noun collocations. With respect to verb+noun collocations, students tended to have a medium level (n=8, 47.1%), followed by a high level (n=6, 35.3%). In phase two, less than half of the students had a high level (n=3, 42.9%) in verb+noun in year 1. While in year 2, at the same phase, the levels of students were equally distributed between low, medium, and high (n=3, 33.3%) in noun + noun collocation. , and students tended to have a medium level (n=11, 61.1%) in verb + noun collocations. In 3 years. During phase 2, the student showed medium levels (n=12, 46.2%) in adjective + noun collocations. For phase three, students tended to have either a high level (n=4, 50%) or a medium level (n=4,50%) in verb + noun collocations. In year 1, it was found that verb + noun collocations were the most commonly used, and students tended to have a medium level of fixedness (n=11, 61.1%) in verb + noun collocations in year 2. Also, in phase three, year 3, it was found that students tended to have a medium level (n=7, 70%) in verb + noun collocation. The outcomes will attempt to fill a gap in the knowledge about the use of corpora to identify measures that would be useful to educators, and examiners to rate second language learners.

Second language learners can produce acceptable collocations regardless of their proficiency level. Even though collocations are not explicitly taught in English classes, university students are able to produce acceptable collocations at three levels of fixedness.

List of Abbreviations

BNC	British National Corpus
COCA	The Corpus of Contemporary American English
ESL	English as a second language
EFL	English as a foreign language
POS	Part of Speech
CIA	Contrastive Interlanguage Analysis

Table of Contents

Research Thesis: Declaration of Authorship	i
Dedication	ii
Acknowledgements.....	iii
Abstract	iv
List of Abbreviations	vi
Table of Contents	vii
Table of Tables	xi
Table of Figures	xv
Chapter 1 Introduction	1
1.1 Background of the Study	1
1.2 Statement of Problem	2
1.3 Purpose of the Study	2
1.4 Research Questions	3
1.5 Organization of this Thesis	3
Chapter 2 Collocations	5
2.1 Definitions and Approaches of Identifying Collocations.....	5
2.2 The Phraseological Approach	6
2.3 The Textual Approach	8
2.4 A Complementary Approach to Identifying Collocation in this Approach.....	14
2.5 Collocation Types	15
2.6 The Nature of Collocation Knowledge	17
2.7 Importance of Collocation Knowledge in Second Language Learning	19
2.8 Collocation Fixedness	20
2.9 Second Language Learners Issues with Collocations	26
2.10 Collocation Deficiency: Overuse, Underuse, and Misuse	29
2.11 Summary.....	38
Chapter 3 Learners' Proficiency.....	39
3.1 Collocation Knowledge across Proficiency Levels.....	39
3.2 L1 Interference	42

Table of Contents

3.3	<i>L2 Exposure</i>	49
3.4	<i>A Usage-Based Approach to Second Language Acquisition</i>	54
3.5	<i>Learner Corpora in Second Language Learning and Assessment</i>	56
3.6	<i>Designing a Learner Corpus</i>	60
3.7	<i>Frequency in Learner Corpora</i>	61
3.8	<i>The Focus of the Current Thesis</i>	62
3.9	<i>Summary</i>	62
Chapter 4 Methodology		63
4.1	<i>Participants</i>	63
4.2	<i>Research Instruments</i>	63
4.3	<i>Research Design</i>	64
4.4	<i>Data Analysis Method</i>	66
4.5	<i>Identifying Collocations in the Corpora</i>	69
4.6	<i>Distribution of Collocations in Learner Corpora</i>	69
4.7	<i>Degree of Fixedness for the Identified Collocations</i>	70
	<i>Building Corpus using Sketch Engine</i>	71
	<i>Methodological Limitation</i>	74
	<i>Summary</i>	75
Chapter 5 Findings.....		76
5.1	<i>Extracting Learners' Collocations</i>	76
	5.1.1 <i>Verb- Noun Collocations</i>	78
	5.1.2 <i>Adjective- Noun Collocations</i>	80
	5.1.3 <i>Noun-noun Collocations</i>	82
5.2	<i>Degree of Fixedness for the Identified Collocations</i>	85
	5.2.1 <i>The First Set of Data Collection</i>	85
a.	<i>Collocations Extracted from Year 1</i>	85
b.	<i>Collocations Extracted from Year 2</i>	88
c.	<i>Collocation Extracted from Year 3</i>	92
	5.2.2 <i>The Second Set of Data Collection</i>	95
a.	<i>Collocation Extracted from Year 1</i>	95

b.	Collocation Extracted from Year 2	97
c.	Collocation Extracted from in Year 3.....	100
	5.2.3 The Third Set of Data Collection.....	105
a.	Collocation Extracted from Year 1	105
b.	Collocation Extracted from Year 2	106
c.	Collocation Extracted from Year 3	108
	5.3 Collocation progression	110
	5.4 Difference in levels of fixedness between phases and years	113
	Chapter 6 Discussion.....	117
	6.1 Proficiency level and Collocations Knowledge	117
	6.2 Collocation’s fixedness	122
	6.3 First Language Transfer.....	125
	Chapter 7 Conclusion	130
	7.1 Implications of the study	130
	7.2 Limitations of the Study	131
	7.3 Suggestions for future research.....	131
	Appendix A	133
	Appendix B	138
	Bibliography	140

Table of Tables

Table 1 Grammatical Collocations (GC) (Benson et al., 1997)	15
Table 2 Lexical Collocations (LC) (Benson et al., 1997)	16
Table 3 What is involved in knowing a word? (Nation, 2001: 27)	17
Table 4 Composition of the spoken BNC (Lancaster.ac.uk, 2018)	61
Table 5: Crosstabulation for year of study and type of collocation.....	76
Table 6 Antilog10 summary statistics of relative frequency and ANOVA between the three years	78
Table 7 Summary statistics and ANOVA between the three years for relative frequency of verb + noun Collocation.....	79
Table 8 Summary statistics and ANOVA between the three years for Log Dice score of verb + noun Collocation.....	80
Table 9 Summary statistics and ANOVA between the three years for Log Dice score of adjective + noun collocation.	81
Table 10 Spearman's correlation between relative frequency of adjective + noun collocation and Years 1,2, and 3.....	82
Table 11 Summary statistics and ANOVA between the three years for relative frequency of noun + noun Collocation.....	83
Table 12 Summary statistics and ANOVA between the three years for Log Dice score of noun + noun collocation.	84
<i>Table 13 Degree of Fixedness for Verb + Noun collocations identified from Year1 the first phase of data collection</i>	<i>85</i>
Table 14 Degree of Fixedness for Adj + Noun collocations identified from Year1 the first phase of data collection	87
Table 15 Degree of Fixedness for Noun + Noun collocations identified from Year1 the first phase of data collection.....	88

Table of Tables

Table 16	Degree of Fixedness for Verb + Noun collocations identified from Year2 the first phase of data collection.....	88
Table 17	Degree of Fixedness for Adj + Noun collocations identified from Year2 the first phase of data collection	89
Table 18	Degree of Fixedness for Noun + Noun collocations identified from Year2 the first phase of data collection.....	91
Table 19	Degree of Fixedness for Verb + Noun collocations identified from Year3 the first phase of data collection.....	92
Table 20	Degree of Fixedness for Adj + Noun collocations identified from Year3 the first phase of data collection.....	93
Table 21	Degree of Fixedness for Noun + Noun collocations identified from Year3 the first phase of data collection	95
Table 22	Degree of Fixedness for Verb + Noun collocations identified from Year1 the second phase of data collection	95
Table 23	Degree of Fixedness for Adj + Noun collocations identified from Year1 the second phase of data collection.....	96
Table 24	Degree of Fixedness for Noun + Noun collocations identified from Year1 the second phase of data collection	97
Table 25	Degree of Fixedness for Verb + Noun collocations identified from Year2 the second phase of data collection	97
Table 26	Degree of Fixedness for Adj + Noun collocations identified from Year2 the second phase of data collection	98
Table 27	Degree of Fixedness for Noun + Noun collocations identified from Year2 the second phase of data collection	99
Table 28	Degree of Fixedness for Verb + Noun collocations identified from Year3 the second phase of data collection	100
Table 29	Degree of Fixedness for Adj + Noun collocations identified from Year3 the second phase of data collection.....	102

Table 30	Degree of Fixedness for Noun + Noun collocations identified from Year3 the second phase of data collection.....	104
<i>Table 31</i>	<i>Degree of Fixedness for Verb + Noun collocations identified from Year1 the third phase of data collection</i>	<i>105</i>
<i>Table 32</i>	<i>Degree of Fixedness for Verb + Noun collocations identified from Year1 the third phase of data collection</i>	<i>106</i>
Table 33	Degree of Fixedness for Noun + Noun collocations identified from Year1 the third phase of data collection	106
<i>Table 34</i>	<i>Degree of Fixedness for Verb + Noun collocations identified from Year2 the third phase of data collection</i>	<i>106</i>
<i>Table 35</i>	<i>Degree of Fixedness for adj + Noun collocations identified from Year2 the third phase of data collection</i>	<i>108</i>
Table 36	Degree of Fixedness for Noun + Noun collocations identified from Year2 the third phase of data collection.....	108
<i>Table 37</i>	<i>Degree of Fixedness for Verb + Noun collocations identified from Year3 the third phase of data collection</i>	<i>108</i>
<i>Table 38</i>	<i>Degree of Fixedness for Adj + Noun collocations identified from Year3 the third phase of data collection</i>	<i>109</i>
Table 39	Degree of Fixedness for Noun + Noun collocations identified from Year3 the third phase of data collection.....	110
Table 40	Distribution for levels of fixedness (advance, high, low and medium) for each phase and collocations in year 1	111
Table 41	Distribution for levels of fixedness (advance, high, low and medium) for each phase and collocations in year 2	112
Table 42	Distribution for levels of fixedness (advance, high, low and medium) for each phase and collocations in year 3	113
Table 43	Mean Difference in levels of fixedness between phases for each year using one-way ANOVA	114

Table of Tables

Table 44 Mean Difference in levels of fixedness between the three years using one-way ANOVA	115
Table 45 Test of Normality	138
Table 46 Test of Normality	138
Table 47 Test of Normality	139
Table 48 Spearman's correlation	139

Table of Figures

Figure 1	Phraseological categories (Howarth, 1998)	6
Figure 2	Breakdown of (Halliday, 2005) example	9
Figure 3	Translation Task (Laufer and Girsai, 2008).....	34
Figure 4	Fill-in- the-gap task (Fernández and Schmitt, 2015)	35
Figure 5	Collocation recognition test (Pellicer-Sánchez, 2017)	35
Figure 6	Concordance with POS tags displayed (Sketchengine.co.uk, 2018)	73
Figure 7	Word Sketch Interface	73
Figure 8	Combinations that has the word "job"	74
Figure 9	Example of Combinations Strength using Word Sketch	74
Figure 10	Distribution of relative frequency between the three years	77
Figure 11	Distribution of Log10 relative frequency between the three years	78
Figure 12	Relative frequency of verb+ noun collocation for year one, two and three.	79
Figure 13	Mean log dice score of verb+ noun collocation for year one, two and three.	80
Figure 14	Mean log dice score of adjective+ noun collocation for year one, two and three.	82
Figure 15	Relative frequency of noun+ noun collocation for year one, two and three.	83
Figure 16	Mean log dice score of noun+ noun collocation for year one, two and three.	84
Figure 17	Distribution of Log Dice between phase and years.	115
Figure 18	Distribution of Log Dice between the three years	116

Chapter 1 Introduction

1.1 Background of the Study

The interest toward large computerized databases of written or spoken language (corpora) has increased in recent years. Corpora could vary in their size and their content as they consist of different type of texts. For example, there are historical corpora that consist of old texts. There are general language corpora that consist of language from various sources and do not belong to a single text type. There are specific corpora dedicated to one particular use of languages like academic journals corpora or newspaper corpora, or a specific genre of languages like medical corpora, and art corpora (Akeel, 2016).

In language research, early examples of corpora use to go back to the 1940s. For instance, Fries and Traver in 1940 and Bongers in 1947 used corpus in classroom research (McEnery and Wilson, 2003). From the 1970-1980, computers became more accessible to scholars, and they started using corpora to address different linguistic matters (Weigle and Goodwin, 2016). Corpora have helped in producing materials such as wordlists – later known as dictionaries- and textbooks by providing real and natural examples of language rather than depending on one's intuition to build up patterns (McEnery and Wilson, 2003).

In language pedagogy, attention to corpora has grown since the mid-1980s (McEnery and Xiao, 2011). They have been used to reference publishing, design syllabus, develop materials, teachers development, and in language testing. What makes language corpora unique is the fact that they are databases (Barker, 2014) that contains a great number of users' language making the analysis more objective and thus adds to it empirical strength (McEnery and Wilson, 2003).

The corpus-based approach means using the pre-existing linguistic description as evidence in language studies (Tognini-Bonelli, 2001). Corpus-based approach provides an essential input on the frequencies of specific structures in language. It provides information about specialized collocations and patterns that are not available in textbooks or grammar books and can only be found in corpora (Boulton, 2012). Many researchers have recognized this approach in second language research, for instance, to examine a number of occurrences of one or more linguistic features.

Chapter 1

1.2 Statement of Problem

Collocations are identified as frequently co-occurring lexical combinations, and they are considered as one of the important and most frequent types of prefabricated formulaic language (Nesselhauf, 2005). Research has shown that English language learners struggle with collocation even at an advanced level. For example, Siyanova and Schmitt (2008) looked at advanced learners of English found that they used adjective-noun collocations in similar quantities to native speakers, but the quality of their collocations differed, with 25% being considered atypical.. One of the justifications behind collocation misuse is that language learners tend to create their own collocations using learned rules to create new expressions that are not found in the target language (Wray, 2005).

Despite the importance of collocations in learners' language development, there is not enough research on investigating learners' collocations use at different levels of proficiency (Staples *et al.*, 2013). Learner corpora contain previous intuited facts about language structure, functions and uses of language with evidence of expert and no expert users. Language researchers and testers can benefit from it to identify learners' errors by their language background or proficiency level, and this will narrow the focus while editing test items and task and other test preparation materials. For example, learner errors can reveal collocations misuse, which test writers can employ as distracter items for multiple-choice questions rather than writing poor distracters that are not found in learner output (Barker, 2014).

There is an increased interest in using corpora to inform and enhance language teaching and assessment (Weigle and Goodwin, 2016). However, most of the research that has been done only quantitatively measured the performance of learners at one point of time and this does not answer questions related to language development over time (Belz, 2007). Many researchers have urged to increase investigations on employing corpora in language testing and assessment especially to explore individual differences between different levels of language proficiency (Park, 2014). This study aims to investigate second language learners of the use of collocations in essay writing. The goal is to compare students' performance at different proficiency levels and compare their progress over a period through corpus-based research.

1.3 Purpose of the Study

Individual differences between second language learners is still an issue in second language learning research (Carlsen, 2012). Global measures of proficiency, such as institution status or

years of schooling, are often based on external criteria and are not reliable indicators of learners' actual proficiency (Callies, 2015).

The purpose of studying collocation production at different levels of proficiency is to understand how learners acquire collocations and how their use of collocations changes over time. This information can be used to develop more effective teaching and learning materials, to identify the specific collocations that learners struggle with, and to develop more accurate assessment tools.

1.4 Research Questions

1. What type of collocations is frequently used among learners of three different years of study?
2. What the degree of fixedness for identified collocations found in learners' learners writing at each year of study and across the academic year?
3. What progress have students made in their collocation production during their studies? does the learners' first language (L1) factor into their production of collocations in English?

1.5 Organization of this Thesis

This introduction is the first of seven chapters in the thesis. A literature review follows, which is divided into two parts: Chapters Two and Three.

The second chapter discusses the different approaches and ways to identify and define collocations. Firstly, it provides an overview of the various understandings and definitions of collocation. Afterward, the paper discusses the various approaches used to identify collocations in corpus linguistics.

The third chapter reviews relevant studies of L2 learners and their use of collocations. A description of research perspectives on collocations is presented, followed by a discussion of similar studies of L2 learners that have studied issues related to the production of collocations, such as learners' language proficiency and types of collocations.

The fourth chapter discusses the research design of the study prior to discussing the context and participants of the study. Additionally, it illustrates the methods used to collect and sample data. The final section of Chapter Four demonstrates the analytical procedure of the written texts, which consists of three steps: extracting candidate collocations from learners' texts; identifying these extracted collocations in the corpus; and measuring their fixedness.

Chapter 1

In Chapter Five, data analysis is discussed, starting with case studies from the three levels of Saudi learners examined, followed by analysis of each level and their production of collocations.

Chapter Six discusses the issues raised by the findings described earlier in the chapter. The conclusion of Chapter Seven discusses the implications, contributions, and limitations of the study and suggests areas for future research.

Chapter 2 Collocations

Collocations are an essential unit of second language learners' spoken and written language production (Wray, 2005). This research aims to investigate second language learners' collocations use in their writing to see their performance at different proficiency levels and compare their progress over a period through corpus-based research. Before getting started with the study itself, it is important to understand collocations in general and collocations used by second language learners. Thus, this chapter reviews the literature in two sections. The first part presents some collocation definitions followed by a discussion of collocation types. The second part reviews the literature on the importance of collocations in second language learning. The third part discussed collocation assessment and how it reflects learners' lexical proficiency.

2.1 Definitions and Approaches of Identifying Collocations

The term collocation has been generally used to refer to a phenomenon in which certain words tend to co-occur regularly within a language. Collocations are made up of different words and are fixed to a certain degree, lexically and grammatically (Nesselhauf, 2005). According to (Sinclair, 1991) collocation occurs when two words or more appear close space of each other within a text, and a group of words is considered as a collocation depending on the frequency they are found together in a text sample (Jones and Sinclair, 1974). Other scholars defined collocations as:

Collocations are frequently recurring two-to-three syntagmatic units, which can include both lexical and grammatical words (Henriksen, 2013).

A group of two or more words that occur frequently together, and it is not restricted to two or three-word sequences (Shin and Nation, 2007).

Usage- determined lexical combinations that are characterized by restricted co-occurrence of elements (Paquot and Granger, 2012).

As defined by Henriksen (2013), collocations are frequently repeated units that can include both lexical and grammatical words. Nevertheless, the limitation is that collocations are restricted to sequences of two or three words. It is important to keep in mind that collocations can be of any length, as long as they are frequently used together. Unlike Henriksen, Shin and Nation's (2007) definition of collocations contains no restriction on the number of words in a collocation. However, it does not identify any specific characteristics that distinguish collocations from other kinds of word combinations. Among the three definitions of collocations, (Paquot and Granger, 2012) provide the most specific definition. Two key characteristics of collocations are identified:

Chapter 2

they are usage-dependent and they have limited co-occurrence of elements. The purpose of this definition is to identify collocations and to understand how they work. Paquot and Granger's definition of collocations, therefore, is the most accurate, comprehensive, and specific.

In general, scholars agree that collocations are all about the close relationship between two or more linguistic units that occur frequently. In the following example, She burst into tears after receiving the terrible news; there are six co-occur words: burst into tears, receiving the terrible news, and thus are considered as collocations. Kjellmer (1984) used six criteria to measure collocations degree of lexicalization and they are :absolute frequency, relative frequency, length of the sequence, collocations distribution over texts, collocations distribution over text categories, and their structural complexity. However, not all scholars identify collocations as above as the criteria are not always applicable, comprehensive or applicable; in fact, views on collocations took two approaches: phraseological approach and textual approach (Carter and McCarthy, 2006).

2.2 The Phraseological Approach

The phraseological approach defines collocation by paying particular attention to lexical, syntactic and semantic limitations that collocations have (Nesselhauf, 2005). Researchers who follow this approach uses either native speaker judgment, or collocational references (corpora) or sometimes both to identify collocations. The main focus of this approach is teaching collocations to L2 learners, and this date back to the 1940s when educators and researchers in the Soviet Union and Japan were interested in forming collocation lists to teach to L2 learners believing this way L2 learners' lexicon will develop (Cowie, 2002). Howarth (1998) proposed an approach to phraseological categories, influenced greatly by Russian lexicology (Arnold 1986):

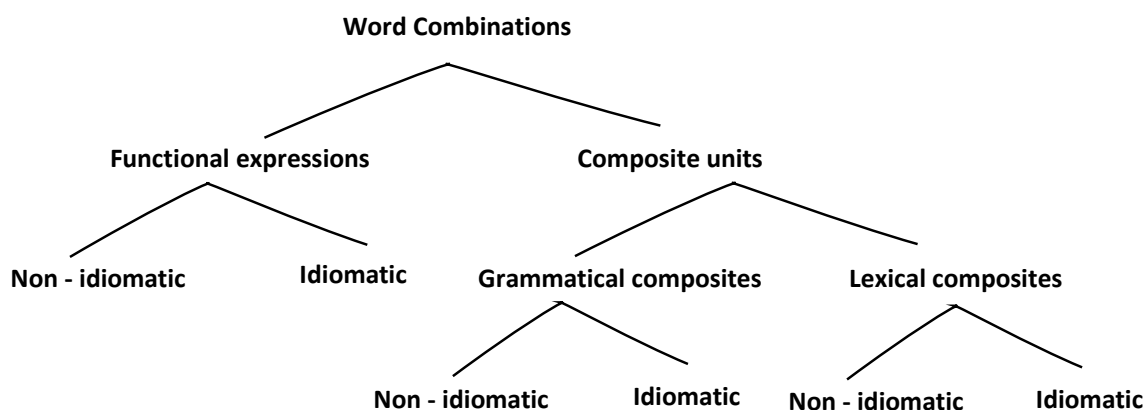


Figure 1 Phraseological categories (Howarth, 1998)

According to Howarth (1998) there are two types of word combinations: functional expressions and composite units. A functional expression is a word combination that serves a specific purpose in a conversation. The proverb "A stitch in time saves nine" is a functional expression that warns people about the importance of resolving small problems before they become large ones. A composite unit, on the other hand, refers to a word combination considered to be a single word even though it is composed of several words. As an example, the catchphrase "Just do it" is a composite unit often used to encourage people to take action.

A.P. Cowie is considered one of the most important representatives of the phraseological approach. He regards collocations as an abstract word combination that has presence in texts.

One of his key ways of defining collocation is to mark them off from other word combinations like idioms and free combinations. To do this he divided word combination into two main types, composites that have a primarily syntactic function e.g. Nounal composites "the middle ground." , and formulae that has a primarily pragmatic function e.g. Good night. Collocations are considered composites for their syntactic function and what differentiate them from other composites is two criteria: transparency and substitutability. Transparency here means whether the word combination has a literal or a non-literal meaning, and substitutability refers to the fixedness of the combination and to what degree the substitution of the combination elements is acceptable (Howarth, 1998).

This approach distinguishes three language combinations. First, In free combinations, elements of a unit are utilized in a literal sense (Nesselhauf, 2003) and carry a transparent meaning. For example, where verb and noun are not restricted and can be freely substituted (e.g., do a task). Second, real collocations (restricted collocations) as the name states, they are combinations that mostly do not allow lexical substitutions. As the noun is not restricted, but the verb is, it can only be combined with specific nouns. Since they are not free and are not restricted, they are sometimes called restricted collocations (e.g., perform a task). Lastly, idioms all elements of the word combination are restricted, and substitutions are not possible at all or only to a very limited degree. (e.g., bring (someone) to task) (Howarth, 1998).

A study by Howarth (1998) explored restricted verb-noun collocations (e.g., reach a conclusion) in academic written corpora that contain advanced native and non-native MA students writing. The research found non-native students used less than half restricted collocations comparing them to native writers. Also, he found that about 6% of collocations produced by advanced non-native students are unconventional. The findings show that advance MA non-native students struggle most with restricted collocations comparing to the other two collocation groups (i.e., free collocations, and idioms).

Chapter 2

In another study that adopted the phraseological approach (Nesselhauf, 2003) investigated L1 German advanced L2 English learners' verb-noun collocation (e.g., take a break). As (Howarth, 1998), Nesselhauf classified collocations into three groups: free collocation is one that is relatively free in its form and meaning. For example, the collocation "make a decision" can be used with a wide range of verbs, such as "choose," "select," and "decide." Semi-fixed collocations: These are collocations with a more restricted meaning and form. For example, the collocation "take a decision" can only be used with the verb "take" and has a more specific meaning than the collocation "make a decision." Lastly, Fixed collocations: These are collocations that are completely fixed in terms of their form and meaning. For example, the collocation "kick the bucket" can only be used to mean "to die." The results are as the following: 79% of learners' errors with restricted collocations, followed by a tied percentage with free combinations with 23 % and idioms with 23% (Nesselhauf, N. 2003).

The literature on the phraseological approach has provided evidence that although collocations are lexically variable, they are still limited to one or more features. One of the limitations was highlighted by (Stubbs, 1995) who argued that using natives' intuitions to identify collocations is not reliable because while natives provide some examples of collocations they cannot estimate their accurate frequency.

2.3 The Textual Approach

British linguist J.R. Firth was the first to use the term collocation in descriptive linguistics as a source of meaning (Ding, 2018). A famous quote of his states that, "you shall know a word by the company it keeps" (Firth, 1958). Firth considered a word's meaning derives from the words with which it co-occurs:

Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of night is its collocability with dark, and of dark, of course, collocation with night. (Firth, 1958)

In the Firthian sense, collocation could be interpreted as empirical statements about the predictability of word combinations (Evert, 2008). Some British linguists such as (Halliday, 1966; Sinclair, 1991; Stubbs, 1995) took Firth's concept of collocation and developed it, often referred to it as the Neo-Firthian school.

Sinclair defined collocations as "the occurrences of two or more words within a short space of each other in a text." (Sinclair, 1991). The "space" here refers to the distance between the 'nods'

and words surrounding it (up to 5 on the right and 5 on the left) (Gablasova et al., 2017a). Also, it is important to point out that different word forms can attribute to the same collocation. Therefore, it induces the definition of collocation as a relationship between lexemes that do not necessarily appear next to each other but in span of up to five words. (Halliday, 2005) explained this by mentioning the following example: in the statement, “he argued strongly, I don’t deny the strength of his argument, his argument was strengthened by other factors.” Here strong, strongly and strength followed or preceded by argued or argument would all be considered instances of the same collocation strong argument (Halliday, 2005). Strong argument is adjective +noun collocation, and the node here is argument:

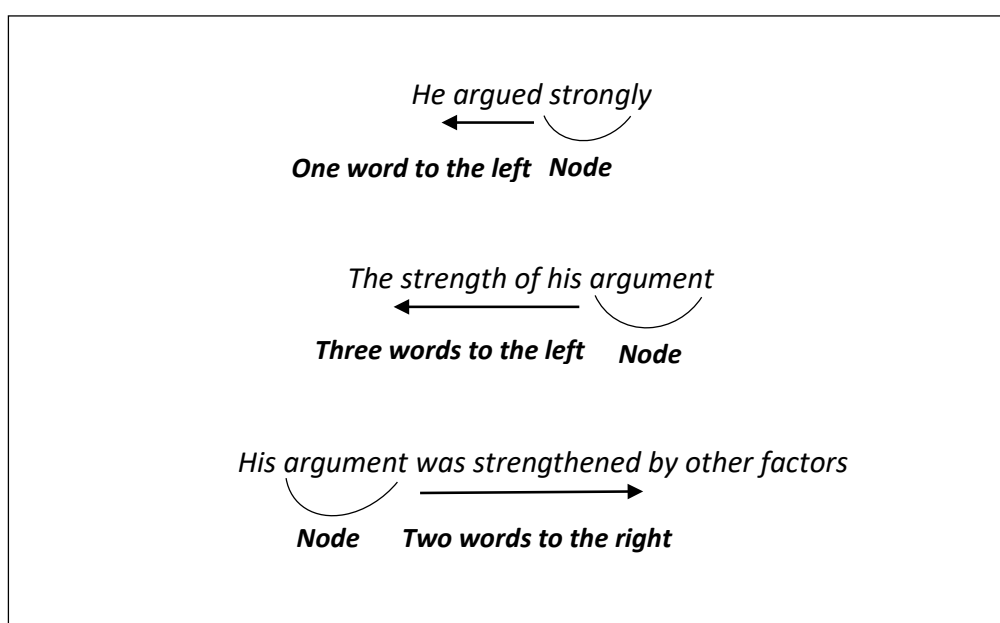


Figure 2 Breakdown of (Halliday, 2005) example.

Unlike the phraseological approach, the textual approach (also known as a frequency-based approach) define collocations based on their frequency found in the corpus. It identifies a collocation when two or more words occur within a short distance of each other in the text. Thus, this approach looks at collocations from a functional angle and consider based on what phrasal language users choose to achieve meaning (Barfield, 2013).

The frequency of linguistic features is an important variable in language learning research as it leads to many discoveries associated with psycholinguistics process that includes representing and producing language (Ellis, 2012; Rebuschat and Williams, 2012). Since collocations in the current study will be selected based on the frequency-based approach, this approach has to be explained in more detail.

Chapter 2

The frequency-based approach takes into consideration lexical items co-occurrences more than a probability of chance in order to identify the restrictiveness of two- word combinations. It engages statistical measures based on data taken from corpora to detect types of word combinations found with a high frequency in comparison to individual words (Wolter and Gyllstad, 2013). Corpus-based language studies have distinguished between two major criteria to identify collocations, absolute frequency and strength of association between two or more content words (Gablasova et al., 2017a). As the name implies, absolute frequency merely focuses on counting words co-occurrences. On the other hand, association measures (AMs) combine information about the frequency with other collocational properties that can be expressed mathematically. Therefore, it is considered one of the advantages of the frequency- based approach because it uses many ways to calculate the strength of association, ranging from raw (absolute) frequency, log-likelihood to Mutual Information (MI) (Brezina et al., 2015). A brief discussion about different measures and the importance of these will be introduced in the next section.

Statistical Measurements of Identifying Collocations

Raw Frequency

Raw frequency is a basic statistic for calculating overall repetition in language. It has received some criticism by researchers, as it is not the best way to locate language regularity and predictability used in corpora (Gablasova et al., 2017a). Findings in some corpora would appear more frequently than other corpora, due to the particularity of the context and number of speakers/writers.

(Stubbs, 1995) stated that many statistical calculations compare the frequency of observed frequency (O) of a pair of word to its expected frequency (E) in a random hypothetical corpus. The pair is only considered to be a collocation if the observed co-occurrence frequency is higher than the expected frequency (Evert, 2008).

(Evert, 2008) argues that a collocation is a combination of words whose observed frequency exceeds its expected frequency. Thus, the words in the collocation are not randomly distributed, but rather are meaningfully connected to one another. He mentioned an example of a very frequent combination is to as found in the Brown Corpus (Kučera and Francis, 1967). However it cannot be accepted as a collocation since its observed frequency (O) is equal to its expected frequency (E). The expected frequency of a combination is calculated using the standard formula: $E = f_1 f_2 / N$ (Evert, 2008). F1 stands for the frequency of the first word component in the corpus, f2 for the frequency of the second word, and N for the corpus size. So if we to directly calculate the expected frequency (E) of the pair is to in the Brown Corpus: $E \text{ (is to)} = 10,000 * 26,000 /$

1,000,000 = 260 (Farooqui, 2016). An additional factor k representing the span size is used in the expected frequency for surface co-occurrence following the formula $E = k f_1 f_2 / N$ (Evert, 2008).

MI-score

Another way to calculate collocation found in a corpus is through Mutual Information (MI) score. Through this statistic, researchers would be able to compare the frequency of one words sequence to the total frequency of each of the individual words. When the number of frequencies of individual words gets higher, it would be more likely that the sequence of the words occurred randomly. Thus, it will not be considered as a collocation (Biber, 2009).

Series of studies conducted by (Siyanova and Schmitt, 2008) used corpus-based frequency data and mutual information (MI) to investigate L2 production and processing of adjective-noun collocations. The first set data was generated from a non-native subcorpus of the International Corpus of Learner English (ICLE) consisting of essays written by 31 advanced Russian university students studying English as a foreign language. The second set of data was generated from a native corpus consist of essays written by L1 English university level students called the Louvain Corpus of Native English Essays (LOCNESS). The researchers extracted 810 adjacent adjective–noun collocations from the learner corpus, and 806 adjective-noun combinations from the native corpus. They consulted the British National Corpus (BNC) to determine the frequency and calculate the MI value of each native and non-native collocations. The frequency information helped in categorizing collocations into five bands based on the number of their occurrences: 0, 1–5, 6–20, 21–100, and > 100. The results show that 45% of the learners' collocations were frequent and strongly associated with English word combinations. The comparison revealed very little difference between native speakers and non-native speakers in the use of appropriate collocation. However, the authors insisted that this comparison does not necessarily suggest that L2 learners at an advanced level can produce develop fully native-like knowledge of collocations for a number of reasons. First, collocation exposure may be different for L2 learners. Through formal and informal education, native speakers are exposed to collocations from an early age. When L2 learners do not learn English in an immersive environment, they might not be exposed to collocations to the same extent. Second, L2 learners may not understand collocations as native speakers do. In general, native speakers have a strong sense of what collocations are correct and which are incorrect. As opposed to L2 learners, who may not have the same level of intuition, they may require explicit instruction and practice in order to become proficient. Beside, L2 learners may not be able to process collocations at the same speed as native speakers. Native speakers are able to process collocations quickly and automatically. L2 learners, on the other hand, may need to take more time to process collocations, and may make more errors as a result. In spite of these challenges, L2 learners can still improve their collocation skills through explicit

Chapter 2

instruction, practice, and exposure. To calculate MI score we follow the formula adopted from (Gablasova et al., 2017a) (Appendix S2):

$$\text{MI score} = \log_2 \frac{O_{11}}{\frac{R_1 \times C_1}{N}} = \log_2 \frac{O_{11} \times N}{R_1 \times C_1}$$

" O_{11} is the observed frequency of the collocation, N is the number of tokens in the whole corpus, R1 is the frequency of the node in the whole corpus and is C1 the frequency of the collocate in the whole corpus." (Gablasova et al., 2017a).

MI score follows the logarithmic scale that is used to reduce the range of value that the MI score acquires in order to demonstrate the ratio between the observed frequency of the collocation O_{11} and the expected frequency $\frac{R_1 \times C_1}{N}$ which relies on the frequency of the node and the collocate in the whole corpus (R1 and C1) relative to the size of the corpus (N) (Gablasova et al., 2017a). There are some issues of applying MI to extract collocations. For example, when the two words are not independent, MI can be misleading. E.g., the words "the" and "book" are very frequently used together. However, this does not mean that the two words are strongly associated with each other. Due to the common nature of both words, the word combination "the book" is very frequent. Also, there can be a correlation between MI and the frequency of two words. As there is a tendency for more frequent words in a corpus to have higher MI scores with other words, even if they are not strongly associated. Beside, The size of the corpus can have an impact on MI. Despite the same strength of association between two words, MI scores will tend to be higher for larger corpora (Gablasova et al., 2017a).

T-score

T-score has been widely known as "a measure of certainty of collocation" (Hunston, 2002) as researchers claim it indicates the strength of co-occurrence and tests the null hypothesis of the strength of association measures (e.g., z-score, chi-squared and log-likelihood tests) (Wolter and Gyllstad, 2011). The t-score provides confidence that the association between node and collocate is genuine by selecting occurrences that appear together less frequently than they would be expected to appear alone (Stubbs, 1995). Corpus size is a key factor to t-score; the larger the corpus is, the more significant a large number of co-occurrences this is due to the amount of evidence that is being taken into consideration. Using the t-score directly to compare collocations between different corpora is difficult because of the problem associated with the size of the corpus (Hunston, 2002). This is why some researchers have stressed on the fact that there is a close connection between t-score and raw frequency, especially in ranking as they are "very

similar to rankings based on raw frequency” (Durrant and Schmitt, 2009). To calculate t-score, we follow the formula adopted from (Gablasova et al., 2017a) (Appendix S2):

$$T - score = \frac{O_{11} - \frac{R_1 \times C_1}{N}}{\sqrt{O_{11}}}$$

“T-score is calculated as the frequency of collocation that we observe (O_{11}) minus the frequency of collocation we would expect by chance ($\frac{R_1 \times C_1}{N}$) divided by the square root of the observed frequency ($\sqrt{O_{11}}$). In the cases in which the expected frequency is small (close to 0), the value of the t-score will be approximately equal the squared root of the observed frequency, because $\frac{O_{11} - 0}{\sqrt{O_{11}}} = \frac{\sqrt{O_{11}} \times \sqrt{O_{11}}}{\sqrt{O_{11}}} = \sqrt{O_{11}}$ ” (Gablasova et al., 2017a).

Log Dice

Despite the many ways of identifying AMs, t-score and MI-score are found to dominate in recent research (Gablasova et al., 2017a). There is no ideal way to measure the strength of word combinations and using of this t-score and MI-score or other measures seem to be somewhat random (Fernández and Schmitt, 2015). Before choosing the proper calculation to use any of the AMs (Gablasova et al., 2017a) asserted that it is important to understand three things.

We need to understand (1) the mathematical reasoning behind the measure, (2) the scale on which it operates, and (3) its practical effect (what combinations of words get highlighted and what gets hidden/downgraded).

Log Dice is a statistical measure based only on a frequency of words w_1 and w_2 and the bigram $w_1 w_2$; it is not affected by the size of the corpus (Sketchengine.co.uk, 2018). Also, one of the ways to find AMs and it is introduced as an alternative to the MI-score. However, there has been much research on it in language learning according to (Gablasova et al., 2017a). A new study by (Zakharov, 2017) examined and evaluated different AMs to extract collocation found one of the best measures in extracting collocations for the node word (beat) is logDice. They noted that MI.l-og_f accuracy with 44 collocations is bigger than that of the logDice with 50 collocations because it has more true collocations in the top part of the list. To calculate logDice we follow the formula adopted from (Gablasova et al., 2017a) (Appendix S2):

$$\log Dice = 14 + \log_2 \frac{2 \times O_{11}}{R_1 + C_1}$$

“14 defines the theoretical maximum of the score from which a certain number is subtracted; the binary logarithm will always be negative or zero because of the ratio of

Chapter 2

$\frac{2 \times O_{11}}{R_1 + C_1}$ will be smaller than or equal to 1. $\frac{2 \times O_{11}}{R_1 + C_1}$ represents the harmonic mean of the proportions of $\frac{O_{11}}{R_1}$ and $\frac{O_{11}}{C_1}$ (Gablasova *et al.*, 2017a).

2.4 A Complementary Approach to Identifying Collocation in this Approach

Collocation simply means a combination of words that often go together in a language to produce natural-sounding speech or writing. For example, one in English can say: They served me strong coffee and a slice of delicious moist cake, it is not normal to say powerful coffee or humid cake. Whilst all four of the previous words would be recognised by learners; it takes a greater degree of competence with the language to combine them correctly in productive use. In recent years numbers of studies investigating collocations have grown in both learner and native language. However, the definition of the term ‘collocations’ is often hazy, and their way of identifying collocations is not the same. This section will present a more operational definition of collocations, and the approach followed in identifying collocations adopting the frameworks of (Laufer and Waldman, 2011; Farooqui, 2016).

The previous section discussed the two known methods for identifying collocations: the textual approach and the phraseological approach. The two approaches to defining collocations outlined above are not in opposition but should rather be viewed as complementary. As Nation (2001: 317) suggested, from the perspective of language learning, collocations should be considered as “items which frequently occur together and have some degree of semantic unpredictability.” Evert (2008) also stressed the close connection between the two approaches. Many collocations identified through corpus analysis have phraseological significance. Conversely, many collocations that have phraseological significance will stand out in corpus analysis. The approach taken in the present thesis is for the combination of both approaches. The term ‘collocation’ is operationalised here as:

“A non-idiomatic pair comprising two open class lemmas which occurs in a corpus (within a window of ± 5) above chance ($f > 5$, $\log \text{dice} > 4$ $MI > 3$ and $t\text{-score} > 2$) and which exhibits specific usage restrictions.”

Here the ‘window’ or ‘collocational window’, is the distance between collocate and the node (reference word). (Brezina *et al.*, 2015; Gablasova *et al.*, 2017a) indicated that the collocational window could go up to five words (5L 5R).

This definition employs both statistical and phraseological criteria. On the statistical side, the following criteria are applied:

- a. Only lexical collocations consisting of two parts of speech will be considered: verb + noun, adjective + noun, and noun + noun collocations.
- b. Lemmas are the form of unit analysis. A lemma is the basic form of a word as found in dictionaries form of a word. The point of using lemma is to be able to consider all forms of a word combination (e.g., quick way, quick ways, quicker way, and quickest way).
- c. The span between co-occurrences is up to five words to the left of the word of interest (node) and five to the right.
- d. The combination should co-occur at least five times in total across at least five different texts to be considered a collocation.
- e- Log Dice and MI are used as a measure of the strength of association with an MI minimum score of 3 and log Dice minimum score of 4. The t-score is used as a measure of the significance of collocations with a minimum score of 2.
- f- The pair has a transparent, non-idiomatic meaning that is clearly deducible from the senses of the individual words.

Thus, the frequency-based approach will be applied first in locating significant collocations in the corpora and then combined with the phraseological approach to investigate the factors that underlie over/underuse of the shared set of collocations among corpora.

2.5 Collocation Types

Researchers have categorized collocations into two main types: grammatical collocations and lexical collocations. Grammatical collocation refers to a phrase consisting of a dominant word (noun, adjective, and verb) and a preposition or grammatical structure such as an infinitive or clause, e.g., accuse (somebody) of. They typically are verb patterns with a bare infinitive or – ing form.

Table 1 Grammatical Collocations (GC) (Benson et al., 1997)

Collocation Types	Examples
Noun + Preposition	kind of

Chapter 2

Noun + to + Infinitive	years to come
Noun + that-clause	We reach an agreement that ...
Preposition + Noun	on purpose
Adjective + Preposition	tired of
Adjective + to + Infinitive	easy to learn
Adjective + that-clause	She was afraid that she would fail the
V + Prep	to wait for

In contrast, lexical collocations refer to a phrase that does not contain prepositions, infinitives or clauses. Instead, they consist of nouns, verbs, adjectives, and adverbs proposed as seven lexical collocation types proposed by (Benson et al., 1997) as seen in the table below.

Table 2 Lexical Collocations (LC) (Benson et al., 1997)

Collocation Types	Examples
verb + noun (denoting creation and/or activation)	make an impression
verb + noun (denoting eradication and/or nullification)	reverse a decision
adjective + noun	best regards
noun + verb	blood circulates
noun + of the + noun (units associated with a noun)	a bouquet of flowers
adverb + adjective	strictly accurate
verb + adverb	apologize humbly
Noun + noun	Office hours

The study of lexical collocations is important for language research because they help us understand how language is used in daily life. We can learn about common patterns of word usage in a language by studying lexical collocations. By utilizing this information, natural language processing (NLP) systems, such as machine translation software and text generation software, might be able to improve their accuracy. It can be done by providing them with information about common word usage patterns. By using this information, NLP systems can be trained to recognize and correctly use lexical collocations. It is also possible to use lexical collocations when teaching a

foreign language. We can help learners improve their communication skills in a foreign language by teaching lexical collocations. The study of lexical collocations can also be used to study the development of language. By observing how lexical collocations change over time, we can gain an understanding of how language evolves over time (Laufer and Waldman, 2011). It is worth to note that , most collocation errors, even among advanced-level writers, tend to occur with verb + noun collocation that has a delexical verb with the highest frequency verbs (i.e., take, give, do, make and have), e.g., make a mistake (Nesselhauf, 2003). Delexical verbs are essential to university-level English language learners as they are frequent in academic English (Altenberg and Granger, 2001).

2.6 The Nature of Collocation Knowledge

The lexical proficiency of a language learner is a measure of the level of knowledge he or she has of words in a foreign language. Global trait models are commonly used to assess lexical knowledge, which takes into account both breadth and depth of lexical knowledge. The first dimension is the breadth of lexical knowledge (i.e., the number of words a learner knows). The second dimension is the depth of lexical knowledge (i.e., pronunciation, and grammatical properties of words) (Anderson and Freebody, 1979). However, these two dimensions have received criticism for excluding some lexical properties related to accessing core lexical items. Such properties include word concreteness, imageability, and familiarity and account for additional elements of words that allow for quicker lexical processing or retrieval (Crossley et al., 2015). (Meara, 2005) Focused on a third dimension, lexical organization, which concerns how vocabulary knowledge is interconnected in a person’s mind. Lexical knowledge has also been categorized as receptive/passive (related to word recognition) or productive/active (which is related to word production). This framework is considered the most comprehensive. It was first proposed by (Nations, 1990) and revised 10 years later by (Nation, 2001). The next section will discuss the framework in more detail.

Receptive Vs. Productive Collocational Knowledge

Collocation recognition is a difficult task for language learners. Researchers have referred that part of the problem has to do with collocation processing. (Nation, 2001) Stated that knowing a word involves three major aspects: form, meaning, and use. Under each aspect, features can be either related to receptive mastery (R) or productive mastery (P) as can be seen in (Table 2.2) adopted from (Nation, 2001).

Table 3 What is involved in knowing a word? (Nation, 2001: 27)

(Note: R = receptive, P= productive)

Chapter 2

Form	Spoken	R What does the word sound like? P How is the word pronounced?
	Written	R What does the word look like? P How is the word written and spelled?
	Word parts	R What parts are recognizable in this word? P What word parts are needed to express this meaning?
Meaning	Form and meaning	R What meaning does this word form signal? P What word form can be used to express this meaning?
	Concept and referents	R What is included in the concept? P What items can the concept refer to?
	Associations	R What other words does this make us think of? P What other words could we use instead of this one?
Use	Grammatical functions	R In what patterns does the word occur? P In what patterns must we use this word?
	Collocations	R What words or types of words occur with this one? P What words or types of words must we use with this one?
	Constraints on use (Register, frequency...)	R Where, when, and how often would we expect to meet this word? P Where, when, and how often can we use this word?

Nation (2001) Has pointed out that receptive knowledge is easier for learners to acquire than productive knowledge. To understand a word, learners tend to only focus on a few distinctive features for a word while they need more knowledge of a word to produce it.

2.7 Importance of Collocation Knowledge in Second Language Learning

Many researchers have agreed on the fact that language is formulaic in nature, and phraseological competence is essential for language users in order to achieve native-like fluency (Wray and Perkins, 2000; Hoey, 2005). The importance of vocabulary knowledge was emphasized by (Wilkins, 1972), indicating that without grammar, it is difficult to pass on information, but without vocabulary, it is impossible to pass information. For second language learners' collocations are an important unit of development in spoken and written language production (Wray, 2005). According to (Friedberg et al., 2017) it is essential for students to acquire collocations that characterize their discipline as it will help them in gaining knowledge through independent reading and help them participate in meaningful discussions. Therefore, as learners progress, they are expected to produce collocations more accurately, which is a bit problematic for second language learners (Ellis, 2008). Research has shown the application of phraseological units such as collocations and phrasal verbs play a major role in a learner's communicative competence (Wood, 2010) because it has an impact on language proficiency in three aspects: complexity, accuracy, and fluency (Housen and Kuiken, 2009).

Collocation also plays an important role in taking on or rejecting a group identity (Wray, 2005). This is clearly confirmed in academic writing, where a writer from a particular discipline such as medicine or chemistry needs to demonstrate their knowledge of the collocations used in their field (Gledhill, 2000). Hence, collocations are important for second language learners not only for fluency and native-like competence but also constitutes a vital means for the writer to become one of the groups of an academic community.

Hsu's (2002) study was conducted among Taiwanese students undertaking English as a foreign language. The objective of the study was to examine whether teaching lexical collocations was imperative in terms of improving the learners' proficiency in collocation use and language fluency. In addition, the study also evaluated the learners' progress to assess for the factors that determine their ability to learn lexical collocations. Qualitative research methods were utilised in the study whereby corpus analysis of class notes, pre and post-workshop tests and other writings from the participants were analysed. A total of 14 participants including eleven females and three male student who were taking part in a workshop were included in the study. All the students were freshmen. Seven of them were majoring in English while the rest were majoring in Banking and Finance. Hsu (2002) hypothesized that a direct emphasis on lexical collocations positively enhanced creativity in collocation use both in written and spoken discourses. The findings suggest a positive correlation between acquired collocations and proficiency. The main limitations of the

Chapter 2

study included a small size of the sample use that may affect generalizability of the findings in the larger population.

Men (2015) examined the factors behind the increasing vocabulary and collocation learning among the Chinese EFL learners. Furthermore, the study explores the relationship between the increasing vocabulary and L2 collocation learning primarily focusing on the syn (onym) set. The study was conducted as corpus-based cross-sectional study among the selected Chinese EFL learners. Men (2015) hypothesized the increasing vocabulary and rising proficiency among the Chinese EFL is associated with collocation knowledge. The findings support this hypothesis as a lack in L2 verb + noun collocation knowledge among the learners correspondingly increased lexical verbs learnt and minimized errors. However, a limitation with this study is that the specific sample size used as well as the proficiency levels of the participants was not provided.

2.8 Collocation Fixedness

Collocation fixedness or a fixed collocation refers to a construction or phrase that is decompositional and whose parts recurrently merge. Fixed collocations have several features, with the first being that the constituents co-occur repeatedly with each other, for example, pretty sure. Martyńska (2004) notes that such co-occurrence represents a 'restricted' collocation that is not only fixed but also inflexible. Other features of collocation fixedness include the fact that their constituents are specifically bound to one another, they do not freely co-occur with other lexical units as only a few, if any, could be substituted with other units that co-occur. Lastly, the usage of fixed collocations is unpredictable, especially on the basis of the known meaning or function of the co-occurring lexical units. These features are consistent with the assertion by Benson et al. (1986) that in relation to fixedness, collocations are lexical units that are fixed and prefabricated. Larson (1998) is in support of this argument and further acknowledges that knowing the words that go together is a crucial step towards understanding the meaning of a text and further translating it accordingly. Therefore, collocation fixedness offers a prolific ground for L2 learners by helping them understand what different texts mean.

Despite the importance of collocation fixedness in second language learning, there is the question of the best way to measure how much collocation knowledge L2 learners have. Research on collocations utilised by L2 learners has found that while learners do use them, there are differences in their use in comparison to natives, and this is both in quality (such as diversity, appropriacy, and accuracy) and quantity (which pertains to using fewer or more of the collocations) (González Fernández and Schmitt, 2015). Measuring and establishing the fixedness of learners' collocations can be informed by usage-based theories of language, whereby

frequency plays a crucial role. According to Ellis (2002), knowledge of a particular language is influenced by the exposure to the same and frequency of utilising certain constructions. The findings from the study on incidental acquisition from graded readers by Webb, Newton and Chang (2013) supported the statement by Ellis (2002) as it was shown that Taiwanese university learners learned more collocations when the frequency of exposure rose up to 15. Therefore, it can be concluded that learners who frequently use the fixed collocations are better versed with these constructions.

Another method to measure knowledge of fixed collocation, as argued by Webb (2008), is the use of c-test formats which provide the first letters of the target word. According to the author, even though this approach is utilised to assess productive knowledge of vocabulary, it could be utilised to measure receptive knowledge, albeit to some extent (Webb, 2008). With regard to this approach, it is pertinent that learners have a larger receptive vocabulary as this would assist them know more of the words that make collocations or constructions compared to learners with a limited reception of vocabulary. Another related approach is the t-score that focuses on collocates of high frequency which includes grammatical words, and further tests certainty just like in raw frequency, but as Gablasova, Brezina and McEnery (2017) argue, this measure is not sufficient to assess collocations fixedness. Gablasova, Brezina and McEnery (2017) further expand on the MI score which tests the collocation's strength of low collocates of low frequency, i.e., lexical words, with the authors informing that this approach is suitable in studying the fixedness of collocations. Nonetheless, the researchers acknowledge that the MI score is an inadequate measurement for testing L2 learners' collocations and their fixedness due to their level of proficiency (Gablasova, Brezina and McEnery, 2017). In addressing this shortcoming of MI score as it relates to stress on "rare words" or sophisticated collocations, Paquot (2017) suggests the need to embrace other statistical measures such as log-likelihood, MI3 and Dice coefficient, all of which resemble the t-score with regard to emphasising on functional and high-frequency words. Therefore, it is pertinent for teachers to consider these measures in the assessment of collocation fixedness among L2 learners.

González and Ramos (2013) presented a contrasting but promising approach of measuring collocation fixedness, highlighting that the collocational richness of learners can be measured using four parameters namely, density, variety, number of errors and also sophistication. With this approach, the four aspects inform an understanding of the extent to which L2 learners have improved their knowledge of the second language. Nonetheless, Siskova (2012) criticises this methodology by arguing that one of the major difficulties in assessing collocation richness is the fact that it can only measure words as a combination of letters that are separated by space. However, with this approach, it is impossible to account for word combinations to form

Chapter 2

collocations or idioms; thus this is a shortcoming that ought to be considered before embracing the measurement method.

Another method used to find fixedness of learners' collocations is Log Dice which is one of the different forms of association measures (AMs) alongside t-score and MI score (Gablasova, Brezina and McEney, 2017). Log Dice, even though it is an alternative to MI score, is a measure that is yet to be fully explored in language learning research (LLR). The measure takes the average appropriate for ratios of two proportions that show the possibility of two words co-occurring in relation to the frequency of the words in the corpus. In its practical operation, Log Dice highlights exclusive combinations (even though not rare) whereby combinations showing a high Log Dice, i.e., more than 13, include phrases such as zig zag and coca cola (Gablasova, Brezina and McEney, 2017). In further clarifying the applicability of Log Dice in measuring collocations fixedness, Gablasova, Brezina and McEney (2017) explain that the method is preferable to the MI score in the event the LLR construct needs to highlight exclusivity between words for collocations with a well-delimited scale and which lack bias for low frequency. The study by Khoja (2019) found that Log Dice has a practical application in that it facilitates research on collocations constructed by learners. This argument is consistent with that by Granger (2018) who maintained that Log Dice does not simply detail information about collocation and non-collocations but also offers detailed insight and interpretation of results obtained from such LLR. With regard to the role of teachers in assessing their learners, Log Dice is particularly valuable as the scores obtained enable comparison between learners in two distinct proficiency levels as well as between distinct forms of collocations that learners produce (Khoja, 2019), and this is consistent with the suggestion by Paquot (2017) on the need to embrace statistical measures. Nonetheless, Khoja (2019) concluded that with regard to the measure's relevance and value in identifying the strength of collocations, the definition and approach to collocation, the research context, and the researcher's interest are key factors that influence the type of results obtained. In other words, due to differences in these factors, it is not uncommon to attain different findings, i.e., some studies could criticise Log Dice terming it as an inappropriate measure. In this light, it is pertinent that researchers understand the strengths and shortcomings of each approach of assessing collocation fixedness before its adoption.

Collocation Fixedness and Language Development

For the past two decades, scholars have focused on studying collocations and their relations to L2 learning. For example, Sadeghi (2010) explored the essence of collocation in teaching and learning vocabulary and noted that L2 learners tend to rely on their native language during communication or translation. While this strategy could assist some learners, especially

beginners in language learning, it is the major source of collocational errors in L2 acquisition as even equivalent lexical items do not readily share the same sense in two different languages for reasons such as cultural differences as reflected in every language's vocabulary and word choice. Accordingly, Sadeghi (2010) argues that neglecting collocation in L2 classrooms should be a major concern for teachers as the lack of knowledge of patterns of fixed and other forms of collocations exposes them to numerous collocational errors which are often more disruptive in communication compared to grammatical errors. The argument is consistent with the statement by Firth (1957) that knowledge of the right collocations forms a part of the competence or fluency of the native speaker. As such, L2 learners should also have a rich stock of conventional collocations which are characterised by different levels of restrictedness or fixedness. In other words, when L2 learners are conversant with a wide range of conventional fixed collocations, their fluency in the language increases, hence evidence of language development.

In addition, from his critical literature review, Qader (2018) concluded that knowledge of collocations, especially lexical collocations as well as collocational usage in the appropriate order, ultimately improve oral fluency and assist learners to sound natural just like native speakers. In another detailed review of literature, Gablasova, Brezina and McEnery (2017) explored collocations in the research on corpus-based language learning and after identifying, comparing, and interpreting the available evidence, concluded that collocations are prominent in the understanding of learning and use of language. Utilising evidence from the Italian context, Fioravanti et al. (2021) acknowledged that word combinations that comprise collocations, idioms and binomials among others are invaluable in the learning as well as the use of L2. In particular, improving the knowledge of learners to such items positively influences the learner's fluency to second learning. Hence, with regard to the emphasis of the current review, collocation fixedness has positive implications on language development. However, proponents of the two approaches employed in word combinations, i.e., phraseological or frequency-based approach, consider collocations to be particularly problematic for L2 learners. For example, as mentioned in Laufer and Waldman (2011) study.

Prefabricated units that involve collocations have a crucial function in language pedagogy and fluency, which ultimately indicates the learner's competence in the second language. Lewis (1997) prioritises collocations in language pedagogy and supports the argument that native speakers are fluent in their language due to existence of prefabricated units (in this case, collocations) in their mind. According to Lewis (1997), fluency in a second language increases with the acquisition of a large pool of fixed or semi-fixed prefabricated units or items which are the basis for establishing any linguistic creativeness or novelty. In this light, L2 learners who are taught how to increase their store of fixed collocations develop fluency in the new language and as such, their

Chapter 2

communication could easily attain the natural sound. The outcomes from the study by Mirsalari (2019) are in support of this argument pertaining to the positive role of collocation fixedness in language development. Precisely, the authors used Pearson correlation combined with paired samples t-test for the analysis of data on developmental patterns of lexical collocation among Iranian learners. Mirsalari (2019) concluded that collocations are beneficial not only in oral but also written production among L2 learners as they help them avoid improper phrases or lengthy sentences when articulating their ideas. With regard to the focus of this section, then it can be concluded that collocation fixedness as learners improve in their proficiency levels is pertinent for language development. Similarly, Siyanova and Martinez (2015) explored the production of noun-adjective collocations by Chinese learners who were part of a course for the intensive Italian language. The researchers found that learners in the advanced learning level could easily produce better or improved L2 collocations (Siyanova and Martinez, 2015). This study implies that once learners move from beginner to advanced levels, they are better placed to produce high frequency and strong collocations. In this light, it is evident that there is a possibility of learning development which makes it possible for the learners to reach the level of native individuals when learning L2. Therefore, collocation fixedness, which clearly improves as learners advance high the proficiency level, is an important indicator of language development among L2 learners, hence it should be emphasised. From the above review, it is evident that fixed or semi-fixed collocations help in language development, thus should be an important part of teaching L2 learners.

Even when learning languages other than English, it is evident that the language proficiency level of learners positively impacts their level of producing collocations. A number of authors including Siyanova-Chanturia (2015) acknowledge that collocations are indispensable parts of L2 knowledge and are crucial to non-native learners and speakers of a certain language in that they assist in the accurate and fluent mastery of the language. Learners' knowledge on fixed collocations is highly correlated with their proficiency level, i.e., a beginner who has little to no knowledge, intermediate and advanced levels (are well versed with collocations). In a more recent study that focused on Djiboutian postgraduate students' collocation knowledge, Abdi and Arifin (2020) found that such knowledge is positively correlated with the writing production of the learners. The implication is that with increased proficiency in fixed collocations, a learner attains the advanced level of L2. The usage-based classroom study by Li, Eskildsen and Cadierno (2014) is, however, one of the most suitable investigations into the link between collocations and learners' proficiency levels considering the authors trace the progression of an English as a Second Language (ESL) learner (Carlos) from beginner to high intermediate during three and a half years in an English classroom. The researchers concluded that the inventory of motion constructions is productive whereby emerging patterns build on previous experience (Li, Eskildsen and Cadierno,

2014). In other words, the vocabulary and language experiences gained as a beginner (stored combinations) help the learner to develop up to advanced levels

The findings on the positive influence of collocations on proficiency are consistent with the argument by Mirsalari (2019) that having a store of collocations that are fixed or semi-fixed helps L2 learners to improve their language proficiency. Similarly, Siyanova-Chanturia (2015) is in support of the statement, further arguing that, unlike advanced learners, beginners or those learners with no previous L2 knowledge, have much smaller stores of single words or even phrases, and as such, they lack proficiency in the language. Beginner learners make use of noun+ adjective combinations as they increase their proficiency, with the increase being reflected in both quality and quantity of the resulting combinations (of fixed and semi-fixed collocations), for instance, “tanned skin”. The researchers suggest that about five months might be adequate for novice learners to gradually begin to come up with or construct more idiomatic output that is almost similar to that of native speakers (Siyanova-Chanturia, 2015). From this study, there is a possibility of collocational performance development, and this facilitates learners to attain the level of native speakers when learning L2, regardless of it being oral or written. The findings are consistent with those by Farooqui (2016) who concluded that advanced-level EAP students could come up with collocations that are the same as native speakers. Through the use of frequency, Farooqui (2016) compared written samples of L2 learners and native speakers to different scholarly articles and journals of professionals from a computer science corpus. The comparison highlighted similarities in the collocation productions of the two categories, especially with regard to noun collocations when assessed against noun collocations of the expert writers (from the journal articles). After follow-up interviews with the participants, the researcher established that factors influencing the accurate collocations production include genre, sub-discipline and topic of the text, and this further supported the argument that learners in the advanced level could attain native speakers’ level. However, Farooqui (2016) provided further insights explaining that the L2 learners must, however, have the same expertise or knowledge in a given genre and discipline as this help them to come up with relevant and accurate academic collocations. Therefore, as learners progress up the proficiency levels, they improve their understanding of L2.

Other scholars such as Laufer and Waldman (2011) and Huat (2012) have also explored the link, but go a step further by investigating the formation of verb-noun collocations among the three proficiency levels. However, Huat (2012) had crucial findings as they related to the qualitative development of these collocations by Malaysian EFL learners. After analysis of the beginner, intermediate, and upper-intermediate (advanced) levels, Huat (2012) concluded that there was a sequence of development with regard to how the learners produced the verb-noun collocations, with the two groups in the higher levels producing collocations that were the same in pattern and

Chapter 2

frequency. In particular, beginners used verbs with non-specific meanings, i.e., pick some flowers while intermediate and advanced level learners produced collocations with particularly specific meaning, for instance, pluck some flowers (Huat, 2012). Therefore, it can be concluded that beginners used general collocations whereas advanced learners could use fixed collocations, and this shows the development in proficiency among the levels.

Fixedness of collocations can be identified as per native-speaker corpus such as the British National Corpus (BNC), which helps in examining the production of distinct types of collocations such as verb-noun, adjective-noun, and noun-noun collocations by learners. According to Nesselhauf (2003), the production of collocations of verb-noun is considered the frequent combinations and learners produce them easily. Nonetheless, learners find it challenging to produce the phraseological combinations that co-occur as per their restricted and fixed meanings (including restricted collocations and idioms). Howarth (1998) explains that the struggle does not just affect learners at the beginner level as it is also not uncommon for advanced learners to only differentiate between idioms and free combinations but fail to grasp the collocations fit between the two (restricted collocations). These findings show the complexity of collocations considering learners from all proficiency levels are affected or find them difficult to understand.

In a task-based study, Martyńska (2004) also examined the other forms of collocations, i.e., verb-noun (i.e., make an appointment), adjective-noun (such as high-heeled shoes), and adverb-adjective among Polish learners. The researcher found that learners could successfully produce almost 50 per cent of the collocations from each of the listed types. Nonetheless, it was evident that L2 learners still struggle with these collocation types as they only completed recognition tasks like multiple choice questions better and accurately compared to production tasks. Similarly, it was easier for the learners to construct verb-noun and adjective-noun collocations than it was for adverb-adjective ones (i.e., totally exhausted) and noun-verb collocations like car breaks down. The study by Chiu and Hsu (2008) had contrasting findings as it showed that Taiwanese learners struggle with collocations that involved adverbs and not others. Therefore, teachers involved in instructing learners on fixed collocations must focus on all the different types of collocations to ensure they help in increasing the proficiency of the learners, i.e., from intermediate to advanced levels.

2.9 Second Language Learners Issues with Collocations

Research has shown that second language learners face difficulties when producing collocations. There are many reasons for this assumption, one of them is learners creating native-like collocations as they try to be idiomatic in their language production (Nesselhauf, 2003). This is

because they rely on creativity and assume that collocations could be altered to an equivalence of semantically similar items, e.g., plastic operation instead of plastic surgery (Wray, 2005). Another problem that learners have is limited use of collocations comparing to native speakers. In some cases, learners tend to overuse a certain set of collocations. For example, collocations constructed with core verbs (be, have, make,) or particular amplifiers (very, completely, highly, strongly) (Granger, 1998). (Nesselhauf, 2005) Pointed out the overuse of the same set of collocations may indicate that is cognate with learners first language. If a learner's first language collocation incongruous with the second language it would cause a negative first language transfer (Wolter and Gyllstad, 2013).

While examining the literature about second language learners' difficulties with collocations, three themes were found. First, learners have difficulties with verb + noun, adjective + noun and noun- noun collocations. Second, learners are slower in recognizing collocations comparing to native-speakers. Third, there are multiple factors that contribute to learners collocation errors. The discussion below highlights the literature that covered three collocation types with a brief examination of each study.

Verb + Noun Collocations

Nesselhauf (2003) looked into the use of verb + noun collocations produced by advanced L1 German speakers, using the German subcorpus of ICLE. After the author extracted verb + noun collocations from the corpus, the researcher accepted learners' collocations if they were found in two out of three English- English dictionaries. When a collocation is not found in the dictionaries, they are presented to two native speakers of English to judge them using a five-stage scale from clearly-acceptable to clearly-unacceptable. Overall, 1,072 collocations were extracted from the learner writing, and around 24% of the collocations were problematic. One of the most common errors involved learners using the wrong verb, particularly with the delexical verbs make and do. For example, learners wrote, make one's homework instead of do one's homework.

Wang and Shaw (2008) wanted to examine verb + noun collocations transfer from L1 focusing on delexical verbs: have, do, take, make. Two hundred written texts were collected from university-level students from two different L1 backgrounds; 100 Chinese and 100 Swedish. All participants were asked to write an essay on the same topic in 30 minutes without accessing any reference tools. The authors extracted collocations manually and then used two criteria to measure their acceptability. The first criteria were native English speaker (the second author of the research), and the second criteria are checking collocation frequency in British National Corpus (BNC). The findings show both chose similar sets of verb + noun collocates: L1 Chinese 299, L1 Swedish 298. However, the two groups produced around 20% of similar error-prone collocations.

Chapter 2

Adjective + Noun Collocations

Li and Schmitt (2010) conducted multiple case-study approaches to investigate learners' use of adjective-noun combinations in their academic writing. Four L1 Chinese postgraduate students studying English Language Teaching abroad took part in the study. Researchers collected eight essays and one dissertation for each participant and conducted a learner corpus consisting of 36 academic writing assignments of three academic terms. Thus, the resulting corpus contained 149,587 tokens divided into three subcorpora (one corpus for an academic term). For comparison, the researchers used the BNC academic written corpus as it consists of texts by proficient writers. The procedure involved extracting all adjective-noun combinations from the learner corpus. Study results indicated that learners had a limited understanding of adjective-noun collocations. Collocations were often used incorrectly or inappropriately, and some collocations were overused. It was also difficult for the learners to understand the meaning of collocations they had not encountered in the past. In addition, the study found that the learners' knowledge of adjective-noun collocations improved over the course of the academic year. Despite these improvements, learners still had limited knowledge of collocations at the end of the year.

Takač and Lukač (2013) investigated the use of adjective + noun collocations by L1 Croatian learners of English, using the Croatian Corpus of English Learner Essays (CELE) which consists of 72,598 words from 298 argumentative essays. The researchers used a log-likelihood technique to identify 73 adjective + noun collocations in the corpus. They compared learners' use of adjective-noun against findings from BNC and ICLE. Their analysis indicated that the following adjectives were overused: bad, big, different, good, important, negative, strict. In addition, Croatian learners of English tend to use highly frequent collocations that account for general-use adjectives, rather than specific adjectives, e.g., a good person rather than a kind person or a responsible person. The authors believe that the use of general-use adjectives could happen due to learners' limited vocabulary and L1 transfer. However, the authors did not report what percentage of the produced collocations was problematic.

Noun + Noun Collocations

Parkinson (2015) investigated noun + noun collocations using three subcorpora of the ICLE corpus. The corpus contains argumentative essays written by learners of English from different L1 backgrounds, and the author chooses three; Mandarin, Spanish, and Tswana. The study compared the influence of learning contexts on learners' use of noun + noun collocations. Also, the researcher wanted to examine if L1 had any influence in learner use of these collocations in English. Findings show that ESL learners' use of noun + noun collocations in writing is more accurate compared to EFL learners. On the influence of L1 in collocation use, the study compared

writers of L1 Mandarin with L1 Spanish writers noting that Mandarin language permits noun-noun phrases, on the other hand, Spanish does not allow noun-noun phrases. The findings revealed that L1 Mandarin writers significantly produced more noun + noun phrases while writing in English compared to Spanish writers. A few examples of noun-noun collocations that ESL learners accurately used include, university degree, government policy, and business decision. Among the noun-noun collocations used by EFL learners less accurately were, data analysis, problem solution, and team work.

In sum, several patterns emerge from the studies examined above. First, second language learners tend to produce highly frequent collocations similar percentage of error-prone collocations regardless of their L1. Even advanced learners of a language may misuse collocations by replacing words that are semantically related to the correct word with words that they think sound more native-like. For example, an advanced learner of English might say "data analysis" instead of the more common "data processing" because they think that "analysis" sounds more sophisticated. However, this is not the correct collocation, and it makes the learner's writing sound unnatural. Also, learners have a smaller set of collocations and keep repeating the same collocations and transferring items from L1. Moreover, most studies reported a number of produced collocations and errors and explored what types of errors were made (lexical, grammatical, and semantic).

2.10 Collocation Deficiency: Overuse, Underuse, and Misuse

Overuse and Underuse

L2 learners tend to use words they know, even if they don't fit perfectly into the context, according to the available choice principle. Conversely, the restricted choice principle states that L2 learners will avoid using words they do not know, even if they are more appropriate (Ellis, 2008). Besides the inefficient use, L2 learners tend to underuse or overuse some collocations. A study on the verb-noun collocations generated by both native and L2 learners of English discovered a few numbers of verb-noun collocations generated by the non-native English speakers compared to the original English speaker counterparts (Men, 2015, pg. 55). The number of intensifying adverbs ending with the letter "L" tends to be lower among L2 learners than among other intensifying adverbs. The reason for this is likely to be that these adverbs are common in the L1 of the L2 learners. A second language learner from a Spanish-speaking background may be more likely to use the intensifying adverb "muy" (very) than "really.". Regarding the use of lexical forms, non-native speakers also recorded a lower use than the original speaker peers, whose compositions demonstrated otherwise.

Chapter 2

The failure to use broader native-sequel expressions frequently leads to the lack of diversity in composition, making the non-native writing sound odd or foreign (Men, 2015, pg. 43). The lack of diversified collocation use is also featured by the underuse and overuse of some collocations. In the spoken generation of routines, sequence, and formulaic by L2 learners, studies found highly notable underuse of certain vagueness and the overuse of some. Regarding the writing performance, it was noted that L2 students generated a much lower proportion of idiomatic phrasal verbs than their native peers. They also tend to underuse more idiomatic collocations and overuse open collocations when using adjective-intensifier combinations in their writings. The forms of over utilised collocations are always associated with the linguistic combinations in the students' native dialect.

The small tokens of precast that advanced learners have in their reserves and the degree to which they are utilised and over utilised. The main reason for the underuse and overuse situation that arises in Non-native speakers' collocation is that students often hold on to fixed expressions and phrases that they feel confident in using. The fixed expressions and phrases tend to become their safe bets or islands of reliability. Thus, their heavy reliance on familiar collocations results in the overuse and avoidance of the ones they are not sure to use results to underuse. These behaviours of L2 students are hardly surprising because underuse and overuse of collocation are unavoidable situations in the interlanguage development process, as is also the case with the use of lexis and grammatical structures.

The use of collocations by L2 learners will typically be less varied than that of native speakers. This is because L2 learners are less likely to be exposed to the full range of collocations in their target language, and they may also have difficulty learning the idiomatic meanings of some collocations (Men, 2015, pg. 45). Which is the main criticism faced by contrastive interlanguage analysis, to the level that there is a tendency of oversimplifying generalisation of students' underuse and overuse when their dialect is directly compared with native peers. To put it in other ways, underuse and overuse is hardly a particular issue of collocation. What is more crucial in non-native collocation research is to study the forms of misuse and discover the fundamental challenges encountered with collocation education.

Alangari (2019) conducted corpus-based research among advanced ESL learners in Saudi. The first part of the study adopts a phraseological approach to examine the use of the verb-noun and adjective-noun lexical collocations among both the native speakers and Arab learners. The second and third part of the study uses a novel approach. In the first study, a sample of 130 speakers of Arabic language produced between 2013 and 2015. Alangari (2019) hypothesized that Arabic students, tend to use the phrasal verb clause structure more often. The limitation with the study

is that following section of the fifteen clauses structures, intransitive clause structures were excluded.

Collocation Misuse

Past studies on non-native collocation noted a considerable proportion of unfitting collocation use. An investigation on the verb-noun collocation in a collection of compositions by advanced L2 learners of English discovered that almost a third of the collocation were misused. It was concluded that these students had notable challenges in choosing the right verbs in verb-noun collocations. Even though non-native speakers generated a considerable percentage of collocation mistakes, previous investigations suggested that not all types of collocation pose similar challenges to L2 students. They faced significant problems in generating verb-noun collocations compared to other types of collocations such as adjective+ noun. Students tend to acquire adjective-noun collocation easily and earlier, while the verb-noun collocations are hard and acquired later (Men, 2015).

However, as learners continue to master and advance their L2 understanding, they also improve their use of collocation. L2 learners increased their accuracy on adjective-noun collocations as they became more and more competent in their second language. This is confirmed by the massive gap between advanced learners who have better control of adjective-noun collocations and beginners who have much less command of the same. Indicating that as they continue to develop their competency, their adjective-noun collocation understanding advances. Despite their better performance on adjective-noun collocations, the phenomena generally pose unique learning challenges beyond doubts even for competent L2 students. Studies have concluded that the learning challenges face by this group include the arbitrary limitations in word combinations (Men, 2015).

Non-native speakers (NNS) are more likely to make mistakes with collocations that are of average restrictiveness, compared to more restricted collocations. This may be due to the fact that NNS are less familiar with the collocational relationships between words, and their mother tongue may also be influencing them.

Among NNS, collocations such as "make a decision" and "take a risk" were more likely to lead to mistakes than collocations such as "fall in love" and "get married" (Men, 2015). Moreover, the researchers found that NNS tended to use collocations that are more common in their mother tongue, even if they are not used in English.

According to another study, NNS' interlanguage (the language they use while learning a new language) is characterized by both linguistic and pragmatic errors (Siyanova & Schmitt, 2009).

Chapter 2

According to the study, NNS were more likely to make linguistic errors than pragmatic errors. It is important to note, however, that both types of errors can make NNS' language sound strange or unnatural.

Although the research on NNS's collocation errors is at an early stage, it has been found that NNS face a number of challenges when learning collocations. These challenges include the limited exposure of NNS to collocations in the target language, the influence of the NNS' mother tongue, and the difficulty of learning the collocational linkage between words. While these challenges are present, NNS can improve their collocation skills by reading and listening to as much of the target language as possible, paying attention to the collocations used by native speakers, using collocation dictionaries and thesaurus, and practicing the use of collocations themselves.

Studies of unnatural production in second language (L2) learning are motivated by the fact that advanced learners of English may not make any practical or systematic errors yet still sound strange or unnatural. Usually, L2 learners will transfer lexical items from their first language (L1) and directly translate them into their second language. It is possible for the L1 to have a different collocation pattern than the L2, resulting in collocation errors.

One study found that Arabic-speaking learners of English were more likely to make collocation errors than English-speaking learners of Arabic (Khamis, 2015). In the study, it was found that this was due to the fact that Arabic and English have different collocation patterns. For example, the Arabic collocation "to make a decision" is translated as "to decide" in English, but the English collocation "to take a risk" is translated as "to risk" in Arabic. This can lead to errors such as "I decided to risk my life" in Arabic, which would be translated as "I made a decision to risk my life" in English.

It has also been found that L1 transfer can also lead to errors in pragmatics, which is the use of language within a particular context (Siyanova & Schmitt, 2009). For example, Arabic speakers may use direct speech more often than English speakers, and they may use different politeness strategies. This can lead to misunderstandings and errors in communication.

Gries (2018) examined the use of corpus analysis in linguistic research in his 2018 study.

According to Gries's study, linguists tend to either overuse or underuse learner corpus research.

The overuse of learner corpora can lead to a focus on errors and negative aspects of language learning, while the underuse can lead to a lack of understanding of the challenges that language learners face. Gries suggested that the current lack of sophistication in linguistic research methodologies can be attributed to the overuse or underuse of learner corpus research.

According to him, linguists need to be more careful when using learner corpora and develop more

sophisticated research methods. Gries's study is primarily limited by the fact that it was a secondary review. Gries' findings may therefore have been influenced by biases or errors in the original research. Despite this limitation, Gries's study represents an important contribution to the field of linguistics. The article raises important issues regarding the use of corpus analysis in linguistic research, and it provides a framework for developing more sophisticated research methodologies.

Collocation Assessment

Achieving lexical proficiency is an essential element of language learning and development. For second language learners, lexical proficiency is an important indicator of their academic achievement. Also, lexical proficiency is a representation with the overall language proficiency. Several studies have reported that there was a sizable correlation ($r_{CA} = .73 - .83$) between learners' general English proficiency measures and collocational proficiency (Eyckmans, 2009; Nizonkiza, 2012).

In a more recent study, (Crossley et al., 2015) research focused on analytic features of lexical proficiency and how these features can be used to predict holistic ratings of lexical proficiency. The models of lexical proficiency they used strongly supported the notion that collocation accuracy on the part of both writers and speakers is the strongest predictor of lexical proficiency. They found that automatically assessing lexical proficiency may be difficult until such time that collocation accuracy can be measured accurately. Thus, they strongly suggested the importance of using human judgments to understand lexical proficiency as it provides us with access to lexical features that are more contextual in manner and assess the accurate and naturalistic use of words within a text. Also, they insisted on the importance that human raters should have access to context (at least in the case of spoken data) to provide reasonable assessments of proficiency. This will provide interlocutor data (data receives from their interlocutors, such as, grammar, and pronunciation) , which may indicate lexical recycling (the process of using the same word or phrase in different contexts) and priming (exposure to one stimulus can affect the response to another), both of which may influence human ratings.

Collocation Assessment Methods

Researchers have used different methods to investigate collocation knowledge among L2 learners. Some of those methods focused on the receptive knowledge of collocations by using tests such as multiple-choice items and cloze test and the other methods focused on the productive knowledge. Since this study is concerned with productive knowledge of collocations, it will only discuss methods researchers have used to test learners' productive knowledge of

Chapter 2

collocations. Three different techniques found in the literature to assess learners' productive knowledge of collocations: composition, translation, and prompted recall test.

Methodologies	Studies
Translation	(Laufer and Girsai, 2008)
Prompted recall test	(Fernández and Schmitt, 2015), (Pellicer-Sánchez, 2017)
Composition	(Farooqui, 2016) (Weigle and Goodwin, 2016), (Wijitsopon, 2017)

Translation Method

In the translation method, learners are asked to translate collocations from L1 to L2 and vice versa. The translation test could vary in its form; one way is to translate collocations directly from as seen below in figure 3 Another way collocations are embedded in full sentence and learners are required to translate the entire sentence, or sometimes learners have to answer a different type of question such as fill-in-the-blanks, or MCQs.

Translate the following English words and phrases into Hebrew or explain them in

English:

Meet the expectations _____

Laudable _____

Figure 3 Translation Task (Laufer and Girsai, 2008)

Prompted Recall Test Method

This method requires learners to use their memory of collocation knowledge to recall the appropriate word sequence to answer a given question. Researchers provides learners with a list of words and learners have to supply the word with some possible collocations using their background knowledge. They usually use tasks such as fill-in-the-blanks.

Fernández and Schmitt (2015) used this method to investigate productive collocations knowledge among Spanish learners of English. The purpose of this study was to explore how collocations frequency and measures of language usage relate to collocations knowledge. The study involved 108 male and female L1 Spanish learners of English from a various range of proficiency levels (beginner-advanced) and different age range (18 -64) years old. The researchers choosed to test learners collocation knowledge by forming a recall test based on 50 common collocations found on COCA. The test instructions were in Spanish and learners were required to provide 2- word

English collocations to summarized or complete information after they read a given statement in Spanish. For example:

El novio de mi amiga es muy alegre y siempre ve el vaso medio lleno.

He seems to have a p_____ a_____ towards life, and I envy him for that.

Figure 4 Fill-in- the-gap task (Fernández and Schmitt, 2015)

On average, participants answered 56% of the questions correctly. The test was followed by a questionnaire about learners' ways of engaging and using English as a second language. On learners' knowledge of collocations, the results indicated that activities outside the classroom such as reading, watching TV and using social media has a more positive effect than a number of years they spent studying English.

In 2017, Pellicer-Sánchez conducted a study to investigate the incidental acquisition of adjective–pseudoword collocations while reading. The study involved 41 English language learners from diverse L1 backgrounds. Participants were randomly assigned to two groups. The first group read a story that contained six target collocations, each repeated eight times. The second group read a story that contained the same six target collocations, but the collocational adjectives were repeated only four times each. One week later, participants were tested on their collocation knowledge (form and meaning) using five recall and recognition tests, which were administered in the form of paper-and-pencil tests and interviews.

Try to fill in the blanks with the word that you think appeared frequently in the story before the words in bold. Select one of the five options. Only one is the correct option. If you don't know the meaning of the word, please select option 'e.'

They lived in a/an ____ holter

- a) small
- b) old
- c) poor
- d) dirty
- e) I don't know.

Figure 5 Collocation recognition test (Pellicer-Sánchez, 2017)

Chapter 2

The question above was taken from the fifth test, and the aim of that test is to assess learners' ability to recognize the form of the collocation by asking them to select the right collocate among five options provided (the correct collocate, three distracters, and an 'I don't know' option). The researcher intended to use distracters relative to the story they read and placed adjectives of the same word class of the correct collocate. The results of the study suggest that learners can incidentally acquire knowledge of the form of collocations from reading.

Composition Method

In this method, learners are asked to write a piece of text in an essay format. After that, researchers analyze students' written output manually or electronically to extract collocations and examine them often building corpora and using corpus analysis tools to do that. This is a common method in literature to examine collocation knowledge; listed below are some studies that employed this technique.

Weigle and Goodwin (2016) carried out an example corpus-based study to support a validity argument on tests distinguishing between low and high proficiency writers in their use of multi-word units. The researchers used 332 same topic essays written by Georgia State Test of English Proficiency (GSTEP) examinees between 2009 and 2012 to build up their corpus. The GSTEP test consists of three source texts: two reading passages and one prompt for the writing task. The researchers hypothesized that high-proficiency writers would use collocations interchangeably, while less proficient writers would use longer strings of words when they found them in source texts. The findings of their study prove that low proficiency depends on source texts in using academic collocations. Examples of collocation found in less proficient writers include (because it is, for example in, we can see). On the other hand, high proficiency writers are more dependant in the application of multi-word units, and many of the collocations found in their writing are mentioned in the Academic Formulas List by (Simpson-Vlach and Ellis, 2010). Examples of multi-word units found in high proficiency writers include (the ability to, the fact that). Their research confirmed the fact that as learners' proficiency level increase, the less they depend on using verbatim forms from source texts in their writing.

In her Ph.D. study (Farooqui, 2016) carried out corpus-based research in order to examine the use of collocations in native speakers of English and second language learners' of English. The argument she brought up is second language learners (non- native) use a limited set of collocations in their academic writing comparing to native speakers. The researcher compiled 600,000 words corpora of 55 students' dissertations of the Computer Sciences major and compared it to a reference corpus compiled 600,269 of 63 Computer Sciences research articles. The results indicated that both native and non-native speaker overuse noun collocations. The

researcher mark such results are due to the students' lack of knowledge of collocations. Towards the conclusion, the author urges educators to alert students to the use of collocations in their academic writing. In her suggestions for further research, the author suggested investigating this matter more including the different levels of proficiencies.

Wijitsopon (2017) Their study focused on one of the linguistic mechanism concept theories called "local textual function" that manifest a controversial relationship between lexical items and texts. The researcher argued this concept has been tackled in many text types, but no one before them applied it to learner writing. In their study, they applied text-lexicogrammar theory to analyze quantifiers found in learner corpora. Two major corpora were used in this study to compare results a learner corpus and a native corpus. First, the learner corpus consists of English argumentative essays composed by L1 Thai learners. Second, the native corpus that consists of argumentative written by English language speakers called the Louvain Corpus of Native English Essays (LOCNESS). Findings from the study shed light on the textlinguistic dimension of the quantifiers in native and non-native speaker learner usage and have pedagogical implications.

In summary, each collocation assessment method provides insights into how collocations can be assessed. There are advantages and disadvantages to each method. Despite being easy to administer and score, multiple-choice tests can be limited in their ability to measure collocations because they only measure the most common ones. Translating tasks can provide a more comprehensive assessment of learners' knowledge of collocations, but are harder to administer and score. Completion and production tasks can also assess learners' knowledge of collocations more comprehensively, but they are more challenging to administer and score.

It is important to note that the method used to test collocations will depend on the research question and the resources available to the researcher. In this study I am going to investigate learners collocation production by compiling a learner corpora. A learner corpus is a more comprehensive, more natural, and more flexible method of examining collocation production than a collocation test. Learner corpora are comprehensive because they contain a large amount of data, which means that they can be used to measure learners' knowledge of a wider range of collocations. A learner corpus contains data that has been produced by real learners. Therefore, the collocations in the corpus are more likely to reflect the natural use of language by learners. Furthermore, learner corpora can be used to study a variety of research questions, not only collocations. For example, the researcher could also use a learner corpus to investigate how learners' collocation knowledge changes over time. Collocation tests, on the other hand, are typically designed to measure a specific research question, which means that they may not be able to be used to investigate other research questions (Fernández and Schmitt, 2015).

2.11 Summary

In this Chapter, I reviewed the relevant literature with regards to collocations and divided it into two main sections. The first section I started it by reviewing various collocation definitions and approaches to identify them with examples of each approach. Then, I presented collocation types and highlighted research done on three most collocation types produced by second language learners; verb + noun collocations, adjective + noun collocations and noun + noun collocations. After that, I discussed the nature of collocation knowledge while referring to literature on collocation knowledge across proficiency levels, and the importance of collocation knowledge for second language learners. The second section covered literature about the approach to investigating lexical proficiency, collocation assessment methods and an overview of writing an assessment.

Chapter 3 Learners' Proficiency

Most research studies have shown that L2 learners' collocation knowledge has a positive correlation with their overall language proficiency (Lee & Liou, 2003; Chen, 2002; Koosha & Jafarpour, 2006), although some studies have not shown such a correlation (Gitsaki, 1996, 1998; Hsien, 2003; Tong, 2004). Researchers focusing on the learning effects of collocation teaching have argued that collocation instruction has been effective for students of both higher and lower proficiency (Lee & Liou, 2003; Lin, 2002; Tseng, 2002). Tseng (2002) investigated the learning effects of 12 weeks of explicit collocation teaching for a group of high school students from Taiwan. Half of the students received explicit instruction, and the other half did not. The results showed that the group that received explicit instruction made significant improvement regardless of the students' proficiency levels. Lee and Liou (2005) divided their 32 subjects into two groups—advanced and intermediate—and had them learn collocation through bilingual corpus-based collocation practices. The result indicated that although higher achievers outperformed lower achievers in both the pre-test and the post-test, lower achievers improved more than higher achievers did. In the same vein, Lin (2002) investigated the effects of collocation on verb-noun collocation teaching among a group of students from senior high school in Taiwan and found that, after the instruction, students of all levels made more progress in receptive collocation tests than in productive collocation tests.

3.1 Collocation Knowledge across Proficiency Levels

Laufer and Waldman (2011) Study examined native Hebrew students use verb-noun collocations in English language writing at three proficiency levels. The researchers compiled a learner corpus, and they called it the Israeli Learner Corpus of Written English (ILCoWE). The corpus consists of 759 argumentative and descriptive essays written by Israeli learners of English of different proficiency levels. The corpus is made up of around 300,000 tokens in total; the basic subcorpus contains 200 essays and 41,621 tokens, the intermediate subcorpus composed of 252 essays and 47,117 tokens, the advanced subcorpus comprises 307 essays and contains 202,311 tokens. The authors used a native corpus (LOCNESS) as a reference and for comparison. The first part of the data analysis devoted to analyzing the native speakers' corpus in order to baseline the comparisons. This involved three main steps. First, they retrieved the 220 most frequently occurring nouns in LOCNESS using the word list function of WordSmith Tools. Then, they compared the collocation found in their learner corpus with native speakers (LOCNESS) to find the 220 most frequent collocations. After that, they created concordances for nouns and identified verb-noun combinations. Next, they verified the frequency and correctness of collocations using dictionaries

Chapter 3

and extracted 1,082 well-formed collocations from the learner corpora. After that, they analyzed each subcorpora to compare learners' collocation production in each proficiency group. The results show English learners at all levels of proficiency produced far fewer collocations than native speakers. Comparison between learners' proficiency level revealed that advanced learners use more collocations; however, interlingual errors found at all levels of proficiency.

Granger and Bestgen (2014) examined the use of collocations among second language learners. However, as opposed to comparing proficient and non-proficient learners, the study compared intermediate and advanced non-native writers. The study used corpus of learner texts. A total of 233 learner essays were used and assigned a grade based on the Common European Framework of Reference for Language (CEFR). However, since there are very few studies examining the development of phraseological competence, Granger and Bestgen (2014) relied on the methodology developed by Durrant and Schmitt. Granger and Bestgen's (2014) work was a non-primary research. In other words, they did not collect their own data, but rather relied on data collected by Durrant and Schmitt previously. Due to this, their findings may not be generalizable to all non-native speakers. Furthermore, Granger and Bestgen's study was relatively small in scale, with only 120 participants. Consequently, the findings of their study may not be definitive, and more research is required to confirm them. Although Granger and Bestgen's study has limitations, it provides valuable insights into the development of phraseological competence in non-native speakers. The results of their study suggest that intermediate learners may benefit from instruction on how to use collocations effectively, while advanced learners may benefit from opportunities to practice using collocations in creative ways.

Nguyen and Webb (2016) examined the EFL knowledge among the Vietnamese learners regarding both adjective-noun and verb-noun collocations. The study used a corpus analytical framework whereby textual material from the learners was assessed for the first 1,000 word frequency levels. Additionally, the study examined the impact of the five factors including part of speech, collocation frequency, node word frequency, congruency and mutual information score. Nguyen and Webb (2016) hypothesized that learners will have adequate knowledge of collocations especially for the first frequency levels. The hypothesis was not met as the results indicated that regardless of the frequency levels, the participants had very little knowledge of collocations which further decreased at each level. However, the findings suggested that there was a close relationship between the participants' knowledge of collocations and single-word items. The key predictor of receptive knowledge of collocations was the node word frequency. The study did not point out the number of participants included in the study and therefore, this limits the generalization of the findings.

Wolter and Yamashita (2017) used a combination of L1 congruency, collocational frequency, word frequency and L2 proficiency in understanding how they affect L2 collocational processing. The study used two groups of English non-native speakers and one group of English native speakers. Although the methods varied across the three studies, the collected data was useful in comparing collocational use among participants of different second language proficiency. Wolter and Yamashita (2017) hypothesized that compared to the native speakers, non-native speakers process congruent collocations relatively faster than English-only collocations. The study findings suggest that both the order of acquisition effects and age are vital in explaining the congruency findings.

In their study, Kamarudin et al. (2020) used a sample of 21 Malaysian ESL learners comprising of 11 female and 10 male students. The aim of the study was to examine the knowledge of collocation among the participants including both their productive and receptive knowledge of grammatical and lexical. The study further assessed the participants' overall performance in three types of collocations including adjective-noun, verb-noun and verb-preposition. The two key instruments employed in this study included the Productive Collocation Test and the Receptive Collocation Test. Compared to the previously examined studies, the main limitation of this study was the small sample size employed which reduced generalizability of the findings. Kamarudin et al. (2020) hypothesized that knowledge of collocations increases with advancement in language proficiency.

Lateh et al. (2021) assessed for collocational knowledge of ESL learners with varying levels of proficiency. Particularly, the study investigated the role of proficiency in terms of learners' ability to use collocation in writing. The research was based on a case study using a sample of 120 first-year students from a Malaysian public university between the ages of 19 and 22 from who essays were collected for corpus analysis. A comparison was carried out after analysis for lexical collocation frequencies and categories. Lateh et al. (2021) hypothesized that compared to low proficiency students, those with a higher level of proficiency use higher lexical collocations. Nonetheless, the findings of this research did not find any significant difference in terms of lexical collocations regardless of their level of proficiency in second language. However, in terms of writing, students who were more proficient used more lexical collocation categories.

Chapter 3

3.2 L1 Interference

One of the major issues second language instructors see in students writing is L1 interference on L2 performance. Language interference is a phenomenon that occurs when a learner of a second language (L2) transfers rules from their first language (L1) to the L2. This can lead to errors in the L2, as the learner may not be aware of the different rules that apply in the L2. Haugen (1969) defined language interference as "the influence exerted by one language upon the other in the speech habits of bilinguals as a result of their regular use of both languages." In other words, language interference is the result of the learner's brain trying to process two languages at the same time.

Transferring rules from the L1 can be commonly used in early language learning stages. This is because learners are still developing their knowledge of the L2, and they may not be aware of all of the differences between the L1 and the L2. As learners progress in their studies, they will become more aware of these differences, and they will be less likely to make errors due to language interference.

Nevertheless, researchers find L1 transfer as a method of learning a second language (Karim and Nassaji, 2013). Sometimes, second language learners transfer rules from their L1 to communicate in L2. It is important to note that not all transfers are negative. As a matter of fact, positive transfers are possible. It occurs when learners transfer some of the rules from their L1 (first language) to their L2 (second language). As a result, learners will not make any errors in the L2 context, and their forms will be accurate (Odlin, 1989). For example, English and Spanish both have a subject-verb-object word order. So, if a Spanish speaker is learning English, they are likely to transfer this rule and use the correct word order in their English sentences. This is an example of positive transfer because it helps the learner to produce correct language. Negative transfer occurs when learners transfer rules from their L1 that are different from those in their L2. As a result, errors may occur in the second language. For example, in Spanish, the verb "to be" is conjugated differently depending on the subject. So, if a Spanish speaker is learning English, they may incorrectly use the Spanish conjugations for "to be" in their English sentences. This is an example of negative transfer because it leads to errors in the L2 (Almukhaizeem, 2013).

Guo et al. (2021) found that learners who were exposed to collocations were better at understanding the meaning of words in context than learners who were not exposed to collocations. This is because collocations provide learners with a sense of how words are used in real-world language. Alijani et al. (2020) argue that it is impossible to understand the meaning of a word without knowing the context in which it is used. This is because the meaning of a word can change depending on the context in which it is used. For example, the word "run" can mean to

move quickly, to operate, or to participate in a race. The meaning of the word "run" will depend on the context in which it is used. L1 interference is a phenomenon that occurs when a learner of a second language (L2) transfers rules from their first language (L1) to the L2. This can lead to errors in the L2, as the learner may not be aware of the different rules that apply in the L2. Collocations can be affected by L1 interference. For example, a learner who speaks Spanish may be more likely to use the collocation "to run out of" with the verb "acabar" (to finish) in English, even though the correct collocation in English is "to run out of" with the verb "run."

Writing in Arabic Vs Writing in English

Writing is an important tool of communication with the reader. In writing, the writer can convey messages, pass knowledge, finish his/ her work, take exam, and overall learn (Archibald, 2004). However, writing is a very complex process as it needs the producing cognitive processes and employing thinking skills (Klassen, 2002). Writing is considered an extremely challenging skill to master for both native and non-native speakers, and is regarded the last step of language acquisition (Al-Khairiy, 2013, Hammad, 2014, Javid and Umer 2014). According to (Javid and Umer, 2014), although composition skill is last step of acquisition it is the most important skill for learners as they have to complete written tasks to prove what they have learned.

The reasons behind writing complexity because it involves three different stages; pre- writing, writing and editing in order to reach the final piece. In addition, writing contains many elements a writer should take into consideration such as, punctuation spelling organization. Hence, proficiency in L1 writing does not necessarily mean proficiency in L2.

In a study that compared internal logical development of the paragraph in English and other languages (Arabic, Korean, French, and Russian). The researcher described how ESL students from various language backgrounds organize their English paragraph. He claimed that while English paragraph is linear in development, Arabic writing is characterized by a series of parallel constructions. Also, in Arabic ideas are largely related to each other through coordination, both clausal and sentential confirming the difference between English rhetoric and Arabic rhetoric (Kaplan, 1966).

Another study looked into the contrast between the two rhetoric has compared (translationally equivalent) and (contextually -equivalent) parallel texts in Arabic and English. The findings indicated that there are significant differences between texts in the two languages at a syntactic level: the Arabic sentence is longer than the English sentence, and it contains more clauses than the English one. Also, there are differences in sentential punctuation underlie differences at a semantic level and that both syntactic and semantic differences reflect differences at a rhetorical level (Mohamed and Omer, 1999).

Chapter 3

Other researchers who have investigated factors behind second language learners struggle to acquire writing skills have listed them under multiple categories. For example, (Ezza, 2010) mentioned writing difficulty for second language learners could be because of cross-cultural differences and/or teaching method being instruction-oriented. Alghizzi (2012) pointed out many reasons to second language struggle in writing. It could be due to L1 language interference, method of teaching and assessment, weakness of L1 composition, and different writing conventions and features.

English writing can be conceptualized in various ways. For instance, it can be viewed as local versus general knowledge. More so, it can be taught through the product writing method or process writing method. It can also be conceptualized as functions according to other views. Regarding general knowledge against local knowledge, there are arguments between researchers of English writing on if to focus on either the former or the latter while educating learners using distinct composition styles. The researchers who adhere to the inner-directed methods or cognitivism think that the composition process is comprehensive. Thus, it is more of a general rather than local knowledge. (Alshammari, 2016, p.g.8). On the other side, those who follow the outer-directed approach or socialist are against the notion that students acquire composition techniques along with the dialect as an internal process and assert that it is an eternal process.

Socialists think that writing is linked to the discourse of the community and the inner-directed theorists are unaware of it. Studies have outlined an explicit difference between the two methods by crediting the local understanding to a discourse community. The composer becomes competent when they have the adequate understanding to write as a member of a discourse community. Hence, the outer-directed theorists claim that book writers should be initially restricted to the discourse society. They indicated that the initial stage in the composition process is that low-level composers should master writing in their discourse society and investigate all the conventions in that field. The local understanding method criticises the universal or general method's ideas that the community does not affect the ability of the composers.

On the contrary, proponents of the universal understanding methodology assert that all individuals can meet comprehensive strategies in various scenarios regardless of their backgrounds. They think that composition possesses general guidelines that involve the dialect in general and writing language in particular. Three assumptions support this approach; first, the masters have more effective universal styles in writing compared to beginners. They believe that comprehensive understanding is stronger than local understanding and can be transferred from one domain to the other. Even though some studies have discovered that thinking is a precise

domain-context bound and cannot be generalised to other domains as local understanding proponents may suggest.

Research has elaborated that distinct domains are not separable and have some similar strategies of argument. Consequently, the cognitive technique is universal, yet the context binds it. Thus, comprehensive understanding is critical for the entire theory of expertise, and often some experts sometimes resort to universal strategies when they face beginner challenges in their field. Suggesting that comprehensive knowledge complements local knowledge and is not conflicting methodologies. In support of this idea, studies used an example of two neighbouring countries in dispute, whereby a country leader has an issue with a neighbour country with aggressive intentions towards his country.

Knowing the limitations of the strengths of his army, he is confident of the inevitable invasion. Thus, he plans to defeat his enemy through politics and thinking rather than military options. The only card that he has to play is being smart. He believes his brilliant capability requires some political lessons to identify a political plan to eliminate the problem. The most important question is if his intelligence can defeat the foe or not. In other words, can the universal understanding or highly cognitive capability translate itself to other domains, or does it solely exist in a particular domain? However, there are no straightforward answers to such problems or which of the two methodologies, local or universal, is right.

Consequently, prior to choosing what will happen, an individual needs to consider some essential variables. For instance, are the chess expert's plans linked precisely to chess games, and accordingly, will he not be able to translate them? Does the chess master succeed by using distinct strategies linked to finding solutions, or is he just an intuitive chess player? Finding solutions to such problems could provide a catch on his percentage of success in his objectives. It is argued that applying universal or specific study is decided by the nature of the subject and class, and one method should not be generalised to categories. By dissimilating the contradictions, the two approaches can be merged.

Universal understanding is a mindful, context-free process and considers comprehensive plans. The more experience the student obtains, the more the scale moves towards the specific understanding, and the more the plans become restricted and precise with a context-based method. Thus, it suggests that a comprehensive understanding of composition can be applied in novice writing and vice versa is true. Therefore, is wrong to generalise one methodology and ignore the other. Instead, both forms of understanding can be used regarding the ability of the students and the nature of the class. In general, the two approaches overlap and cannot be used in a specific domain.

Chapter 3

Process versus Product Writing Approaches

The product or traditional methodology concentrates on the final composition product, which sees the education on composition as the process of evaluating and assigning writing parts. It does not focus on the thinking process behind generating the final output as viewed in the process composition methodology (Alshammari, 2016, pg. 11). The main objective of the educator in the traditional writing approach is correcting and assigning grades on the learners' papers. However, the teacher can analyze learners' papers and categorise the series of mistakes in what is referred to as the error analysis methodology. During training, the educator typically instructs learners to read plays, essays, poetry, and novels and analyze them in written essays. The other type of traditional approach plan is that the educators request learners to copy a composition and generate a similar example. Then the instructor evaluates the outcomes before offering the learners a similar assignment of generating a comparable piece of the composed work.

The traditional methodology concentrates on the outcome, and little attention is given to writing strategies or the thinking process by which learners develop their compositions. Guidelines in this method focus on conventions such as introductions, thesis sentences, and probable paragraph structures. The approach is broadly used in Saudi Arabia in English sections and the English Malaysian school system because it consumes less than the process methodology. In Saudi Arabia, the traditional methodology is the primary approach to teaching writing in the Arabic dialect. The educators favour spending more time analyzing learners' papers rather than engaging students in the thinking process. The objective of the approach is to offer learners lexical understanding, and thus students are required to copy composition templates with distinct subjects. In addition, the approach focuses on the proper use of syntax and vocabulary.

According to this approach, learners imitate the template and try to compare the writing model given by the educator (Alshammari, 2016, pg. 13). The teacher's responsibility is colossal because they give feedback and ensure that the learners generate the needed follow-up response. Traditional compositions have four phases: controlled writing, familiarisation, free composition, and guided writing. For instance, the instructor makes one of his objectives to educate learners how to describe a house. Initially, he offers the learners with writing template, which depicts the suitable preposition, and the names of the rooms used in the description of the house. In the subsequent phase, learners generate simple statements obtained from statements utilized in the template given to them. The instructor may physically paint parts of the statements so that learners know how to substitute some vocabulary.

In the next phase, the learner may utilize the image of the house to generate a passage of instructed composition. Lastly, they can compose paragraphs about any house or write essays

about their houses in the free writing phase. The first step in this approach is copying until the students master the technique and, afterwards, write freely. This methodology is greatly influenced by the behavioristic approach where the instructor offers the stimulus, and the student generates the response. The behaviorists view education as a copying process and external variables rather than internal one's influence students. They believe that learning primarily occurs when students receive stimulus from environmental exposure.

There are two distinct types of traditional writing theory. The first is the form conducted in the ESL classroom, where the concentration is on the drill-and-practices activities. The second arises in the non-ESL classroom, where four elements are identified.; the subject and the objective of the compositions assignment is assigned or controlled, learner writing is graded for linguistic and syntactic accuracy as well as conceptual content, learners are then given prompt mistakes response and corrections in the prioritized techniques, and learners may be requested to make adjustments in the prioritized techniques. The ESL learners will have similar composition guidelines as the native speakers when their competency of the dialect advances, thus the variations in guidelines between ESL and non-ESL are linked to priority only.

Arabic writing borrows its approach of writing from the Holy Quran and is hugely influenced by the Quranic techniques. For example, its argumentive writing completely differs from English (Alshammari, 2016, pg. 18). Contrary to the western approach, which relies on a logical system that is beyond words, Arabic argumentation relies on the fact that it is expressed with a few words. There are four elements of Arabic argumentation that differentiate it from the western style of writing. First, Arabic writing is mainly characterised by repetition both in the Arabic dialect and in writing. The recurrence in writing is not solely limited to writing but incorporates all the dialect elements such as lexical, morphological, and phonological, semantic, and syntactic stages. The arguer expresses facts by making them present in the discussion, by paraphrasing, doubling, repeating, and calling them to attention.

The English writing style is linear, where it begins with the subject statement and is succeeded by proof that confirms the claim and final results to the conclusion that supports the subject statement. However, Arabic is a Semitic dialect based on compound series of parallel fabrications, both negative and positive, which is why Arab students tend to make vast use of coordination in a manner that proficient English speakers see as an overuse. Additionally, Arab students transfer their rhetoric techniques from their native dialect and do not achieve their receivers' expectations in the target society while writing in English. Therefore, in connection to logic and structure, they simultaneously gain rhetoric expressions and reasoning forms when they obtain their first dialect.

Chapter 3

Generally, the most prevalent source of error of L2 Arab students is the influence of L1. Therefore, when they process English syntactic structures, they are likely to borrow some styles, such as over-generalisation and simplification, from their native language. On the contrary, certain researchers have argued against the notion of the negative transfer from L1. They believe that the mistakes displayed by L2 students while composing are because of their inadequate understanding of their intended language, English. Based on their studies, they discovered no serious relationship between students' L1 and their mistakes in the composition of L2. More so, their discoveries did not indicate any severe variations in the rhetoric performance between original Arabic speakers and their English counterparts.

Socialists think that writing is linked to the discourse of the community and the inner-directed theorists are unaware of it. Studies have outlined an explicit difference between the two methods by crediting the local understanding to a discourse community. The composer becomes competent when they have the adequate understanding to write as a member of a discourse community. Hence, the outer-directed theorists claim that book writers should be initially restricted to the discourse society. They indicated that the initial stage in the composition process is that low-level composers should master writing in their discourses society and investigate all the conventions in that field. The local understanding method criticises the universal or general method's ideas that the community does not affect the ability of the composers.

On the contrary, proponents of the universal understanding methodology assert that all individuals can meet comprehensive strategies in various scenarios regardless of their backgrounds. They think that composition possesses general guidelines that involve the dialect in general and writing language in particular. Three assumptions support this approach; first, the masters have more effective universal styles in writing compared to beginners. They believe that comprehensive understanding is stronger than local understanding and can be transferred from one domain to the other. Even though some studies have discovered that thinking is a precise domain-context bound and cannot be generalised to other domains as local understanding proponents may suggest.

Universal understanding is a mindful, context-free process and considers comprehensive plans. The more experience the student obtains, the more the scale moves towards the specific understanding, and the more the plans become restricted and precise with a context-based method. Thus, it suggests that a comprehensive understanding of composition can be applied in novice writing and vice versa is true. Therefore, it is wrong to generalise one methodology and ignore the other. Instead, both forms of understanding can be used regarding the ability of the

students and the nature of the class. In general, the two approaches overlap and cannot be used in a specific domain.

According to the traditional approach, learners imitate the template and try to compare the writing model given by the educator. The teacher's responsibility is colossal because they give feedback and ensure that the learners generate the needed follow-up response. Traditional compositions have four phases: controlled writing, familiarisation, free composition, and guided writing. This methodology is greatly influenced by the behavioristic approach where the instructor offers the stimulus, and the student generates the response. Behaviourists view education as a copying process, and external variables rather than internal ones influence students. They believe that learning primarily occurs when students receive stimulus from environmental exposure. Generally, Language (L1) transfer is a concerning issue that many language students experience, particularly in the initial stages of language learning. Arabic students continue to face issues when they borrow language items from the first dialect, then translate them and use them in the second language.

3.3 L2 Exposure

Obeidat and Mahadi (2019) assessed the knowledge of collocations of 60 students at Indiana University of Pennsylvania—30 native speakers of the target language and 30 non-native speakers of the target language. A gap-filling test and a written exam similar to the TOEFL were used. Both tests revealed that the non-native group performed much worse than the native group, corroborating earlier results. Zhang (2013) observed that collocational knowledge improves both written communication fluency and the ability to produce appropriate collocations. Furthermore, no information was provided on respondents' length of stay in the United States or the amount of time spent engaged in language learning. Unlike prior studies, Alijani et al. (2020) research's focused on the reliability and validity of a collocational knowledge exam that included two gap-filling tests and a multiple-choice test. To achieve this aim, University of Hawaii students from East Asia were required to take a battery of exams. The linguistic ability of the participants ranged from basic to advanced.

Furthermore, the study looked at the link between participants' understanding of collocations and their overall English skills. The first and third subtests showed good reliability and validity, whereas the second subtest, which looked at verb-proposition collocations, had low validity and reliability. The research discovered a link between collocation knowledge and general linguistic ability.

Chapter 3

In 2018, El-Dakhs et al. demonstrated that learners' total language competency is connected to their understanding of lexical collocations. In his research, he found 81 male Saudi English majors who represented the four Saudi EFL student categories: juniors, seniors, and freshmen. To assess their understanding of verb-noun collocations, a 50-collocation exam was administered. A written test and an institutionalized TOEFL exam were also used to assess participants' overall language ability.

In contrast to El-Dakhs et al. (2018), AlZahrani (2021) study found a positive relationship between participants' collocation knowledge and total language skill. In other words, persons who are fluent in English have a high standardized knowledge, while those who are less fluent have a low level. As a result, AlZahrani revealed significant disparities in students' knowledge of collocations at different academic levels.

Finally, Alijani et al. (2020) investigated three elements of optional knowledge: the use of collocations as a measure of language competency, the tactics employed by L1 Arabic English learners to produce acceptable collocations, and the mixing of Arabic and English phrases. Sixteen Arabic verb khasara ('break') colloquialisms were assessed on two groups of advanced EFL university students utilizing a multiple-choice and a free translation exercise (38 graduate students and 32 undergraduate students). Participants employed various communication tactics while translating the Arabic word khasara into English, including avoidance, paraphrasing, literal translation, assumed synonymy, overgeneralizations and analogies, substitutes, and literary style imitation, according to the findings. Even proficient L1 Arabic speakers of English struggled with English collocations, according to the study's findings. According to the study, conversational English should be studied more extensively at both the high school and university levels. Sonbul and El-Dakhs (2020) corroborated prior research findings by analyzing learners' generation of English collocations. To do this, the researcher gave a list of themes to 42 Arabic-speaking university students studying English. As a homework assignment, participants were required to compose an essay. The participants wrote 42 essays on one of the problems offered to the researchers. As a result of this evaluation, individuals demonstrated a poor degree of collocational competence. Because of their wide use of collocations, including lexical and grammatical collocations, 64 per cent and 61 per cent of the word combinations they utilized were incorrect, respectively. They claimed that some of the inaccuracies were caused by the influence of Arabic, their native language. According to the findings of the study, teaching collocations is crucial for foreign language English learners.

Lu and Dienes (2020) explore whether L2 learners usually have the collocation of the L2 language beforehand. Primarily, the study examines both the conscious and unconscious acquisition of the

collocation knowledge. The study used an experimental approach whereby a total of 68 Chinese undergraduate students took part. Researchers divided participants into two groups. The first group was asked to rate the typicality of English collocations without any training. The second group was given training on collocations before rating their typicality. Lu and Dienes (2020) hypothesize that unconscious structural knowledge, unlike conscious structural knowledge, focuses on acquiring statistical regularities. This means that unconscious knowledge is acquired through exposure to patterns in language, while conscious knowledge is acquired through explicit instruction. For example, if you are learning a new language, you may unconsciously learn the statistical regularities of that language, such as which words are more likely to be used together. You may also learn the conscious knowledge of the language, such as the grammar rules. Lu and Dienes (2020) argue that unconscious knowledge is more important for language acquisition than conscious knowledge. They believe that unconscious knowledge allows us to learn language more quickly and efficiently. The study findings indicate that learners acquired the knowledge of collocations both consciously and unconsciously. The key limitation of this study is that the level of L2 proficiency among the two groups of participants was not taken into consideration when designing the experiments. Proficiency as a variable could have affected students' knowledge of collocations irrespective of whether or not they were trained on the 24 English collocations.

Alsakran (2011) explore both the essence of both the productive and receptive knowledge of collocations among advanced Arabic ESL learners. Primarily, the study examined whether or not the language environment influences a learners' ability to acquire collocations both in written and spoken second language. A cross-sectional study was conducted from a sample of 68 participants comprising of 30 Arab students undertaking Intensive English programme at the Colorado State University and 38 Saudi students studying at that the Institute of Public Administration. Alsakran (2011) hypothesized that there would be a significant difference in performances among the participants in the three types of collocations including verb-preposition, verb-noun and adjective-noun. The findings for the comparative assessment of the student in Saudi Arabia and the United States showed significant difference suggesting that environment has a significant impact on learners' acquisition of the L2 collocations. Also, the study found a significant statistical difference in the three types of collocation. Generally, the Arabic-speaking learners had a poor knowledge of collocations than their counterparts. The main limitation in Alsakran's (2011) study was the small sample size used which limited generalization of the findings to the larger population.

Pellicer-Sanchez (2015) investigated the incidental acquisition of L2 collocations. Specifically, the study examined how adjective-pseudoword collocations that learners encounter while reading contribute to their acquisition of collocational knowledge. The cross-sectional study used a total

Chapter 3

of 41 participants all of whom were second language learners within a classroom setting. All the participants were required to read through a written text which contained a total of six targeted collocations. The two groups of participants in the study include the 8-repetitive group and the 4-repetitive group. Pellicer-Sanchez (2015) hypothesized that it was possible to learn collocations incidentally from reading. The hypothesis was confirmed by the findings whereby incident learning was found to be as effective in learning collocations at a similar rate as other lexical components. The study also did not find any significant implications of frequency manipulation in acquisition. The main limitation of the study was the failure to control the impact of variables such as environment, and level of proficiency among the L2 learners as this could have as well impacted on the level of incident learning.

Ramadhan (2017) explores the influence that various features of collocations and its related knowledge and development have on learners' development. The study used a longitudinal research design in a sample of 252 Kurdish high school students learning English language as a foreign language. While the study essentially examines the same aspects of the collocation use across learners of varying proficiency. The use of longitudinal research by Ramadhan (2017) is unique because it provides a hybrid definition of collocations whereby both the phraseological and frequency-based views are used. The study employed two tests; one for learners' receptive knowledge and another for measuring the participants' productive knowledge. Application of longitudinal design was important in determination of collocations use over time while providing clear focus and validity. However, the design had a negative impact of repeated testing. In their study, Kamarudin et al. (2020) used a sample of 21 Malaysian ESL learners comprising of 11 female and 10 male students. The aim of the study was to examine the knowledge of collocation among the participants including both their productive and receptive knowledge of grammatical and lexical. The study further assessed the participants' overall performance in three types of collocations including adjective-noun, verb-noun and verb-preposition. The two key instruments employed in this study included the Productive Collocation Test and the Receptive Collocation Test. Compared to the previously examined studies, the main limitation of this study was the small sample size employed which reduced generalizability of the findings. Kamarudin et al. (2020) hypothesized that knowledge of collocations increases with advancement in language proficiency.

Khoja (2019) examined the extent of collocation use in writing among foundation-year students using corpus-based approach whereby a total of 16 written sample texts were produced by the participants. The two levels of participants used in the research included pre-intermediate and intermediate ESL learners. However, the specific number of both pre-intermediate and intermediate participants who took part in the research was not stated in the methodology. Khoja

(2019) hypothesizes that Saudi learners tend to produce a high number of collocations that are acceptable. Secondly, the study hypothesizes that there is no significant differences in collocations use between the pre-intermediate and intermediate levels of second language learning.

Lu and Dienes (2020) explore whether L2 learners usually have the collocation of the L2 language beforehand. Primarily, the study examines both the conscious and unconscious acquisition of the collocation knowledge. The study used an experimental approach whereby a total of 68 Chinese undergraduate students took part. The participants were divided into two main groups. The typicality of English collocations was examined by the first group with no training while the second group was trained on the collocations before undergoing the text. Lu and Dienes (2020) hypothesizes that unlike conscious structural knowledge, unconscious structural knowledge focuses on acquisition of statistical regularities. The study findings indicate that learners acquired the knowledge of collocations both consciously and unconsciously. The key limitation of this study is that the level of L2 proficiency among the two groups of participants was not taken into consideration when designing the experiments. Proficiency as a variable could have affected students' knowledge of collocations irrespective of whether or not they were trained on the 24 English collocations.

Chen et al. (2021) investigated the acquisition of collocation competence in second language learners based on a corpus analysis. Specifically, the study hypothesizes that ESL learners develop competence in collocation in the process of growing ESL proficiency. The study used a large sample size of L2 texts amounting to 5200 essays. The large sample size was pivotal in allowing for generalizability of the findings. An external criteria used for assessment levels among the learners was the standardized English proficiency tests. The study findings indicated that as learners' proficiency improved, exclusivity also increases. Lastly, the findings showed that as proficiency grows. Learners tended to use less deviant collocations.

Chen and Baker (2014) explore discourse features in the development of a second language. Primarily, the study examines lexical bundles from corpus data ranging from 1990 to 2002. However, the specific sample size (number of learner essays) utilised was not indicated and therefore this could negatively affect both generalizability of the findings and reproducibility of the research. The study hypothesized that compared to proficient ESL writers, learners writing at lower proficiency levels share more features with conversation. On the contrary, proficient ESL writers are more close to academic prose. The variation in proficiency therefore informs on collocation knowledge and use in ESL writing.

Chapter 3

Koc (2021) explores the frequency and development of collocations based on students' essays. The study focused on Turkish EFL learners with varying levels of written and spoken proficiency. The corpus-based analysis focused on a sample of 75 students. The participants were divided into three sub-groups of 25 students each. The categorization was based on the students' levels of the proficiency including intermediate, upper-intermediate and low advanced EFL level. Koc (2021) hypothesized that learners' use of verb-noun and adjective-noun collocations in essays predicted further improvement of collocations use with increasing proficiency. The study findings demonstrated that a significant statistical difference existed across the three student groups in terms of collocations use. The study's main limitation is that it only considered EFL proficiency levels and overlooked other variables.

3.4 A Usage-Based Approach to Second Language Acquisition

Usage-based approaches are a number of cognitive linguistic models that emphasize that actual language use has a significant influence on linguistic structures (Tyler, 2010). They assume that language learning is fundamentally usage-driven, which means that language is learned through "frequency-biased contextualized exposure to and use of meaningful linguistic constructions to objectively observable cultural and interactional behaviour in the target discourse community." (Cadierno and Eskildsen, 2015). In other words, this approach assumes that language learning is bottom-up, statistical, and frequency-informed. In terms of second language acquisition, learners' track frequency of linguistic input (word and patterns) then, and they constantly categories, consolidate, and create language structures.

Ellis (2002,2003) has stated that collocations are developed from a psychological mechanism known as "chunking", which was used to explain how short-term memory is tied to "chunks" of information for fast processing. Ellis argued that collocation acquisition goes through the same process. When a combination of two or more words co-occur frequently, they are registered as a chunk and processed as a one structure. This process of learning chunks is thought to be facilitated by the way that our short-term memory works. Short-term memory is a limited-capacity memory system that can hold about 7 items of information for a short period of time. However, if we repeatedly encounter the same group of words, they can become stored in our long-term memory as a single chunk. This means that we can access them more easily, and we don't have to keep re-learning them from scratch.

The ability to chunk words together can have a number of benefits for language learners. First, it can help to improve fluency. When we can access chunks of words as single units, we can produce

and understand language more quickly and easily. This is because we don't have to stop and think about the individual words in each chunk. Also, chunking can help to improve communication ability. This is because chunks are often associated with specific meanings. For example, the chunk "to run a marathon" means to participate in a long-distance race. This means that when we use this chunk, we are able to communicate our meaning more precisely.

(Durrant and Schmitt, 2009) did carried a study to see if usage-based models could also be applied to adult second language learners. English native and non-native writers took part in the study to investigate their use of collocation. Collocations were identified using the textual approach.

The researchers wanted to obtain short and extended pieces of writing for data analysis. For non-native learners, they choose research assignments texts that learners produced as project work for courses in English for Academic Purposes. They explain the reason behind their choice of this text is because they suspected that longer text would show the extent of collocation use and statistically robust trends may only emerge in longer stretches of writing. They also analysed a set of shorter essays from the International Corpus of Learner English (ICLE) to see if the analysis could work for less extended texts. For native extended texts, they used assignments from students on the MA degree in Applied Linguistics at a British university, and essays from the current affairs magazine Prospect. The shorter native texts were obtained from the Louvain Corpus of Native English Essays (LOCNESS). The study found that non-native writers rely heavily on high-frequency collocations, however, they do not use infrequent strongly associated collocations sufficiently comparing to native speakers. Their findings are consistent with usage-based models of acquisition while accounting for the impression that non-native writing lacks idiomatic phraseology.

Wray (2002, 2008) discussed "dual-processing system" which is a comprehensive model to explain how language users use and process collocations. The model includes two types of processing mechanisms: analytic processing that helps speakers formulate novel utterances, and holistic processing which is relatively involve less effort and enables speakers to produce formulaic language patterns. Adult second language learners acquire language differently than first language children. Children learn a second language by focusing more on phrases, adult second language learners tend to separate fixed phrases and expressions into single lexical items, and thus make more mistakes while trying to reconstruct the lexical items into formulaic language. Wray suggested that this difference between child L1 learners and adult L2 learners is the result of a combination of social and cognitive factors. On the social side, adult learners are not necessary in immediate need to memorize helpful communicative sequences for communication especially if are not in L2 environment or they are learning in a classroom setting. Moreover,

Chapter 3

traditional language classrooms focus more on vocabulary learning and grammar drills, collocation and other formulaic sequences are not usually facilitated.

All views and models discussed above are originated from first language acquisition and highlights the significance of storing and processing frequently formulaic language. Applying these models to collocation learning indicates that collocations acquisition is different from native to learners. Natives acquire collocations as formulaic and familiar and frequent word combinations are processed faster in first language. This is because native speakers have been exposed to collocations more frequently and have had more opportunities to use them.

Gries (2009) mentioned number of factors that can contribute to the difference in collocation acquisition between native speakers and non-native speakers. These factors include: frequency, context and intuition. Native speakers are exposed to collocations more frequently than non-native speakers. This is because collocations are more common in native speech and writing. Also, collocations are often used in specific contexts. For example, the collocation "to run a marathon" is only used to refer to participating in a long-distance race. Native speakers are more likely to be aware of the contexts in which collocations are used. In addition, native speakers have a better intuition about how collocations are used. This is because they have had more opportunities to use them and to see how they are used by other speakers.

The usage-based approach to collocation learning suggests that non-native speakers can improve their collocation acquisition by being exposed to more frequent collocations and by learning the contexts in which they are used. In addition, non-native speakers can benefit from explicit instruction on how to use collocations correctly.

3.5 Learner Corpora in Second Language Learning and Assessment

Since the mid-1980s and over the past four decades corpus-based approach to linguistics and language education has gained prominence. The reason behind this interest is because corpus analysis can be illuminating "in virtually all branches of linguistics or language learning" (Leech, 1997). In fact, learner corpus data can be used for pedagogical purposes by incorporating the findings of second language acquisition (SLA) and foreign language teaching (FLT) research into language classroom or into teaching materials, or after undertaking surface research into the language learners' production (Díaz-Negrillo and Thompson, 2013).

It is essential to look at the differences between *reference corpus* (or native corpus) and *learner corpus*. They are the two most commonly used types of the corpus used in second language learning. Reference corpus is formed from spoken or written language produced by proficient or

native language users. Examples of reference corpus are Corpus of American English and the British National Corpus. On the other hand, learner corpus consists of written or spoken text produced by language learners. Examples of learner corpus are Cambridge Learner Corpus (CLC), and the International Corpus of Learner English. Learner corpora are built and used to help scholars analyze learners' text and understand how it is different from natives' language. Such analysis and understanding would help scholars and educators in many ways. First, through corpus analysis, they could observe how learners' language develop at each level of proficiency. Second, it could help in developing learning and assessment materials. Moreover; it would give a closer and deeper look at language production that would help greatly in second language acquisition research (Barker, 2010; Park, 2014; Akeel, 2016).

Learner corpora are considered an ideal source to Investigate collocation knowledge because of two major advantages. First, they contain extended written or oral discourse rather than decontextualized language units, e.g., phrases or sentences. Second, learner corpora typically contain texts resulting from pedagogical tasks that give learners the freedom to choose wording from their lexicon, such as, writing composition, and translation (Paquot and Granger, 2012).

Corpora in Language Testing and Assessment

In language testing and assessment, corpora have not been used until the twentieth century (Barker, 2006). Alderson (1996) suggested that corpora could be used in language tests throughout test construction, compilation and selection, scoring, and results. He added that the potential use of corpora in language tests would make a clear judgment about test takers' ability, knowledge, and performance since they are based on real language. In the United Kingdom, the first involvement of corpus usage in Language testing and assessment goes back to the 1990s when Cambridge University created a written learner corpus. Later, Cambridge became more interested in spoken language and learner's output. Both written and spoken learners' language was then archived at what is now known as Cambridge Learner Corpus (CLC)(Barker, 2006; Park, 2014).

Corpus linguistics and learner corpora have contributed a lot to second language research. Learner corpora have played a major part in the interlanguages description, and many of its outcomes has positively affected foreign language teaching and learning (Callies, 2015; Callies and Götz, 2015). In recent years, learner corpora and native corpora have got more attention from the language testing and assessment community this is due to the many useful uses of corpora (Barker, 2013; Park, 2014).

Chapter 3

Learner corpora are used by test writers to explore the collocational pattern in the learner and or native production so that common or less frequent pattern can be tested to distinguish between candidates at a particular level of proficiency. Additionally, the most frequent errors or misuses of specific collocation appearing can be used to provide suitable distractor items for multiple-choice questions. Corpus evidence whether from learner or native corpora is used alongside experienced questions writers' intuitions about what learners can be expected to know a certain level, so there not considered to replace human questions writers in the test writing process.(Fiona Barker, 2014).

Looking to real-life text samples produced by natives, proficient, or learners, corpora would be valuable for test developers especially to use in large-scale tests (Park, 2014; Mustafa, 2015). Through corpora, researchers can examine language produced by language users by studying its structures, functions, and uses (Barker, 2014). By conducting such investigation, corpora would assist in informing, validating, and developing ways to handle proficiency at different levels as in the Common European Framework of Reference for Languages (CEFR).

Corpora Use in Test Development

Corpora can provide test developers with test content which considered as the input, and the performance of the test takers which is considered as the output. These can build up archives for developers to enable them to reflect on many issues related to testing development (Barker, 2006). Developing a language test goes through six different steps. First, developers have to define what trait or skill they want to assess. Second, they should create a test plan that includes types of items tested and their response. Third, they write test items and run a trail. Fourth, assembling the test by combining test to form a complete test and conduct a pilot test. Fifth, test developers would administer and score the test. Lastly, the tests are reviewed for validation (Weigle and Goodwin, 2016).

Another useful use of corpora in language testing and assessment is to use it corpus data in revising and updating assessment materials. Many high stake language tests such as TOFEL, IELTS, and Cambridge use corpora for such purpose (Barker, 2006; Park, 2014). The Corpus of Contemporary American English (COCA) is a large-scale corpus that is monitored, and data is added gradually (COCA, 2017). This feature would help scholars access to the latest language collection and notice language change over time. In language test and assessment such hallmark would allow developers verify their tests is updated to the latest changes in language.

Corpora Use in Test Validation

According to (Fulcher, 2013) finding tests level of reliability could be either through running a pilot test (with participant whom the test is not intended for and see the statistical analysis results) or by doing a revision based on item analysis (item discrimination) to check their level of difficulty. An item is considered discriminated if only high-achieving students were able to answer it.

Corpora have proven to be a valuable empirical base for test validation this is why it is useful for large-scale tests (Park, 2014; Mustafa, 2015). Reference corpora contain language structure produced by expert users, and it could be useful for language test development in the sense that they contain well-formed grammatical context. Therefore, it can be used for tests validation. Unlike learner corpora which on the other hand would give test developers access to various forms of language structured by language learners of different levels of proficiency. Gaining access to numerous content of language production is an advantage to record language growth lexically and grammatically across time (Weigle and Goodwin, 2016).

One positive example of the value of using corpora as a tool for validation tests is a study done by (Mustafa, 2015). At the start of his study (Mustafa, 2015) mentioned that research has proven that most test given to students are not checked for their reliability. He argued that the lack of knowledge and carelessness are the reasons why educators do not consider the reliability of their tests. In his study, the researcher designed a reliable grammar exam without having to do the usual reliability test check (as mentioned above by (Fulcher, 2013)). Instead, he used corpora for that. The researcher used reference corpora to design two sets "TOFEL-like" grammar test. He tested and retested the pilot then compared the results with ETS TOFEL grammar test to give him a number close to exact to the ETS TOFEL version. This research is one proof of the usefulness of corpora in designing a reliable language test.

Corpora Use to Document Aspects of Learners' Language Development

Learner corpora could also be used to report lexical and grammatical growth of second language learners. This is useful for educators to describe differences between learners across time or levels of proficiency. Researchers taking part in the English Profile Programme (englishprofile.org) are working on a collaborative project for the CEFR to use corpus research techniques to provide grammatical and lexical structures at different levels of proficiency using data from Cambridge Learner Corpus (CLC). Insights of this project are posted as entries on their website based on specific features in learners' grammatical development called "Grammar Spotlight" (English Profile, 2015). For example, in the employment of present simple, they found B1- level students (lower intermediate) could form negative tag questions to request an opinion or seek agreement (e.g., "It's a bit strange, don't you think?"). On the other hand, B2 (intermediate) learners can use present simple in a wider range of speech act. (e.g., "I'd love to go to the concert with you, so I

Chapter 3

accept your invitation.” Information of this nature would help test professionals in developing tests that will identify learners’ level of proficiency.

In his study, Barfield (2006) assessed for knowledge and development in second language collocation. Additionally, the study sought to examine the factors that either hinder or promote knowledge in L2 collocation. The study adopted a statistical analysis method whereby a total of 188 students participated in a test whereby they responded to seven subjects. The study hypothesized that students exploit selective and context-independent conditions in terms of lexical knowledge which includes multiple lexical strategies, paraphrasing and associative networks. The study findings suggested the need to improve corpus analytical methods to statistical methods that are more practical in terms of exploring collocation knowledge development in ESL. The main limitation of this study is that it only focused on the specific collocations without putting into consideration alternative variables such as environmental factors and proficiency levels which could also impact on collocational and lexical knowledge.

3.6 Designing a Learner Corpus

Taking on a corpus study starts first by creating the corpus itself (Sinclair, 1991). Learner corpus-based studies design their corpus differently depending on their research aims and linguistic features investigated, despite all this, corpus compilation has specific criteria that should be considered. These criteria are mentioned in (McEnery, Xiao, & Tono, 2006; Wynne, 2005) and include key design principles such as sampling, representativeness, and balance.

Flowerdew (2004) stated that a corpus does not have a particular or ideal size. The author added that the size of a corpus depends on the needs and purposes of research. Nevertheless, it has to be of a sufficient size where a number of linguistic patterns are enough to validate a hypothesis. On the other hand, (Aston, 1997) indicated small-scale corpora or subcorpora should range from 20,000 to 200,000 words. All in all, there is no specific size for corpora and size will majorly depend on depends on the purpose (McEnery et al., 2006).

Representativeness is an important feature in the corpus, and it determines corpus quality. It can be referred to the range to which a sample contains a full collection of variability in a community (Biber, 1993). In other words, corpus contents have to represent the language that it samples if the outcomes based on its content can be generalized to investigated language variety (Leech, 1991). Representativeness does not only refer to the implication of all possible linguistic features found but also extends to include competence features in all its levels, for an instance, interlanguage development found in the corpus.

A balanced corpus usually covers a wide range of text categories which are supposed to be representative of the language or language variety under consideration. These text categories are typically sampled proportionally for inclusion in a corpus so that it offers a manageably small-scale model of the linguistic material which the corpus builders wish to study as in the example below (Table 3.1) from (Lancaster.ac.uk, 2018).

Table 4 Composition of the spoken BNC (Lancaster.ac.uk, 2018)

Region	%	Interaction type	%	Context-governed	%
South	45.61	Monologue	18.64	Educational/informative	20.56
Midlands	23.33	Dialogue	74.87	Business	21.47
North	25.43	Unclassified	6.48	Institutional	21.86
Unclassified	5.61			Leisure	23.71
				Unclassified	12.38

3.7 Frequency in Learner Corpora

The corpus-based approach provides an essential input on the frequencies of specific structures in language. It provides information about specialized collocations and patterns that are not available in textbooks or grammar books and can only be found in corpora (Boulton, 2012).

Corpus approaches provide researchers with a detailed description of interlanguage at different stages of development and that offer insights into the sources of variation in the language use and level of proficiency reached by language learners. (Gablasova *et al.*, 2017b).

The frequency of linguistics features occurring in contextualizes language use can provide us with further insights into the linguistic system developed by second language learners and uses. (Laufer and Waldman, 2011) Insisted on the valuable information that corpora provide to us and cannot be provided by any other means. For the reason that, it gives us information about frequency and frequency is an important feature in analyzing language use at different levels of proficiency it analyses the lexical-grammatical and pragmatic were there these features appear once or co-occurred (McEnery *et al.*, 2006; Granger, 2015).

Frequency information is no doubt a valuable source of evidence about language development and news. However, it's it would be difficult to interrupt evidence without a reference point. As the results of major corpus-based language studies employ a comparative design in which a corpus is compared to another corpus of a different learner language, or with a native corpus.

Chapter 3

This allows researchers to contextualize and evaluate corpus evidence making the findings more informative than if L2 corpus was considered alone (Granger, 2015).

3.8 The Focus of the Current Thesis

Having reviewed the literature of corpus-based studies on collocations use among second language learners, I can summarize learner corpus research findings related to collocations in three points. First, second language learners seem to have difficulties producing collocations. Also, comparing them to native speakers, second language learners seems to use fewer collocations. Moreover, second language learners appear to lack awareness of the limitations of collocations as well as the possibilities of compounding multi- words unit.

Most studies only focused on investigating second language learners collocation production at one proficiency level or learners performance at one point of time, and this does not answer questions related to language development over time (Belz, 2007). Using a corpus-based approach in this study would provide essential input on the frequencies of specific structures in language (Boulton, 2012).

Thus, I am interested in investigating second language learners' collocation use, as well as which factors underlie the over/underuse of the most frequent collocations and to discover how much learners' collocation production have developed through their academic year. The goal is to understand learners' performance and progression in order to identify measures that would be useful to educators, researchers, and examiners of learners' lexical ability at different levels of proficiency.

3.9 Summary

This chapter introduced corpora and how it played a major role in language research, pedagogy and more recently in language assessment. The second part focused on frequency in learner corpora and what input frequency could provide for language research. The third part identified relevant literature in regards to the uses of corpora in language assessment and ways to assess second language writing. The next chapter presents methodology and discussions on data collection and data analysis.

Chapter 4 Methodology

This research focuses on university-level students of English Language performance at different proficiency levels and compares their progress over a period through a corpus-based investigation. The main aim of this study is to discover individual differences between second language learners' use of collocations and the differences in collocation production between three years of study. In this chapter, I present the plan of methodology and address how to answer my research questions:

1. What type of collocations is frequently used among learners of three different years of study?
2. What the degree of fixedness for identified collocations found in learners' learners writing at each year of study and across the academic year?
3. What progress have students made in their collocation production during their studies? does the learners' first language (L1) factor into their production of collocations in English?

The methodology section consists of two main sections. The first part outlines the research design that will be deployed, and the second one describes the research method, including the data analysis approach.

4.1 Participants

The study is conducted at a university in Saudi Arabia. Participants are Saudi female university students at English Language and Literature program speak Arabic as their first language. The program offers courses in: English Literature, Translation Studies, and Linguistics. Each program consists of 8 levels (or 8 semesters) during four years of study, and each year has two levels (or 2 semesters). Students in all programs are required to take English language skills courses as core modules. Those modules include, Reading and Writing in level one, Advanced Writing in level two, and Professional Writing in level three.

Participation were taken from a major university I collected 354 samples 293 were considered, (104) participants from the first year, (98) from the second year, and (91) from the third year.

4.2 Research Instruments

As the present study was aimed at investigating L2 learners of English written production of collocations, the research tools selected are argumentative essay writing tasks adopted from the

Chapter 4

International English Language Testing System (IELTS) Task two. To pick the essay topic questions, I consulted three English language teachers (one English native and two non-native). I picked out ten essay topics that are relevant to learners and asked the teachers to choose three topics where students would show their collocation development. The first topic is:

Some people say that the only reason for learning a foreign language is in order to travel or to work in a foreign country. Others say that these are not the only reasons why someone should learn a foreign language. Discuss both these views and give your own opinion. (*Cambridge English IELTS 11 : authentic examination papers, 2016*).

The second topic is:

Some people think that universities should provide graduates with the knowledge and skills needed in the workplace. Others think that the true function of a university should be to give access to knowledge for its own sake, regardless of whether the course is useful to an employer. What, in your opinion, should be the main function of a university? (*Cambridge IELTS 7 : examination papers from University of Cambridge ESOL Examinations : English for speakers of other languages, 2009*).

The third topic:

Some people think that parents should teach children how to be good members of society. Others, however, believe that school is the place to learn this. Discuss both these views and give your own opinion. (*Cambridge IELTS 8 : examination papers from University of Cambridge ESOL examinations, 2011*).

All the topics above were chosen carefully in consultation with English language teachers. The goal was to have students write about topics they are familiar with or could form an opinion about. All essays are argumentative, and the topics are connected and reflect ideas and experiences of everyday life.

4.3 Research Design

An assessment of the progression in the use of collocations in writing among students, comparing and contrasting differences between different learner levels is a sophisticated task. In this respect, three methods have been commonly deployed in the studies that combine quantitative and qualitative approaches in analysing texts. The techniques include the cross-sectional approach, contrastive interlanguage analysis, and frequency-based approach

(Malamatidou, 2017). Combining these three approaches facilitates an in-depth understanding of the issue under scrutiny in the given study.

This study uses Khojah's (2019) research methodology, which is a contrastive interlanguage analysis (CIA). A CIA allows researchers to assess learners' linguistic development and collocations use by comparing their writing to that of native speakers. According to Granger (2015), CIA is a highly popular method when studying learner corpora. The comparative design of this has enabled us to uncover a comprehensive range of features distinctive to learner languages. In addition, we have evaluated the extent to which these features are generalizable across learner populations. This framework has been used in similar studies and proved reliable. For instance, Lam and Crosthwaite (2018) used the CIA method to investigate the ability of learner corpora to assess particular elements of learner language compared to native speakers' language. They acknowledged that the CIA facilitates the study of different language learners in varied contexts through comparison, investigation of the differences, and generalisation.

Appel (2020) consensually opines that the CIA approach is practical when deploying corpora to compare language use among English as Second Language (ESL) learners and native speakers and different learners' groups from diverse L1 backgrounds. Based on these views, framing the proposed study on the CIA framework will ease the evaluation of learners' written collocation compared to native speakers, as shown on the BNC. Besides, the use of the CIA will help compare two different learner's levels; intermediate and pre-intermediate use of collocation.

The study uses corpora to apply the frequency-based strategy to carry out these comparisons. This approach will allow checking the collocation learners produced concerning their association and frequency measurements in the BNC to verify their acceptability. This approach is also similar to Wanner, Ferraro and Moreno's (2017) classification of collocation study. Although Wanner, Ferraro and Moreno's (2017) approach to studying collocation involved assessing common words, specific theories, a group of words, he did not refer to corpora. In contrast, based on BNC, the proposed study seeks to assess four forms of juxtapositions; noun-verb, verb-noun collocation, adjective-noun collocation, and noun-noun collocation. The ones identified from students' texts will be examined for fixedness as well as other aspects such as complex forms. Additionally, interpreting frequency numbers and strength scores of collocation in the BNC based on corpora will aid validation of the comparison between different learner levels to show possible differences and similarities in their production.

The third strategy in this study will be the cross-sectional approach. This will aid the comparison between learner levels. Although Pérez-Paredes and Mark (2022) opined that the CIA is beneficial with longitudinal studies, Appel (2020) had a different view that corpora study using

Chapter 4

the CIA strategy can also be deployed in a cross-section approach as the two studies are similar. That is, cross-sectional enables the researcher to identify differences between levels over the same period; longitudinal research identifies the differences among students with similar proficiency levels over a given long period. This method is called 'quasi-longitudinal', where samples of learners involved are obtained from the same type, same as L1, but from different levels of proficiencies (Appel, 2020). In this study, learners will be from diverse proficiency levels but in the same context.

Additionally, they will be investigated over the same period. Samples of the learners from will be obtained simultaneously and then analysed using the same analytical approach. The three strategies will overlap in assessing the study phenomenon, with each method focusing on a different area. Nonetheless, all the research will focus on the 1st year, 2nd year and 3rd year students' use of collocation by comparing their written texts.

4.4 Data Analysis Method

This study will use three steps in analysing learners' sampled essays to assess their use of collocations. The first phase will involve extracting learner collocations from the essays manually. The second stage will involve identifying the extracted collocations based on their fixedness, aided by the Sketch Engine. The final step will focus only on collocations that will not have been recognised from the corpus.

Extracting Learner Collocations

There are various procedures for extracting learner collocations from written texts. Such strategies include using extraction software programmes used in Uhrig, Evert, and Proisl's (2018) study and manual extraction as Putri and Siahaan (2016) used in their research. This study will involve a few essays and seeks to observe learners' writing closely. Therefore, manual extraction is a suitable approach for extracting candidate collocations. Nonetheless, manual extraction is challenging and requires systematic application considering its limitations. Although it has been occasionally used in some previous studies, such as Putri and Siahaan (2016), such studies only focused on adjacent collocations and one form, adjective-noun.

The current study differs from Putri and Siahaan's (2016) study in that it seeks to investigate various forms of collocation in a broader span than adjacent collocations. The manual identification approach is slow and requires revisions. On the other hand, although extraction software can reduce mistakes that might occur due to manual extraction, manual extraction is still the most suitable for learners' level of writing. Besides, this approach is recommended for

handwritten samples. Markedly, learners involved in this study will be from different proficiency levels; thus, their texts might have spelling and grammatical faults that can best be explored and addressed through manual examination. Before finalising the data extraction process and identifying learners' collocations from their essays, the analysis will be tried using samples from all learners' levels. The initial testing will guide decision-making and flow systematically throughout the extraction process.

Extracting the collocation will begin by proofreading learners' written texts to identify and correct spelling mistakes. This will ease the process of extracting learners' collocations. However, this will be done without affecting students' lexical use. Brackets will be used to indicate corrected spelling. Nonetheless, grammatical errors will be left because they are the primary concern of the research in collocation production. Additionally, leaving grammatical errors will help maintain the data as original as possible.

Mainly, the study will concentrate on producing three forms of collocations, namely, verb-noun, adjective-noun, and noun-noun collocations. These will be identified as learner collocations during extraction. However, the extraction process will have specific measures for selecting these combinations from students' work. For instance, only common nouns used to describe the particular meaning and take countable and uncountable forms will be considered. Proper nouns will be omitted since, in the current study, they are primarily relate to the Arabic context. Such nouns are rare in native English corpora; hence they will make an insignificant difference in assessing English collocations. In extracting learners' adjective-noun (A) collocations, noun modifiers and verb complements will be included.

When identifying learner verb-noun (V) collocations, primary auxiliaries "have" and "do" lexical verbs, and open set nouns will be included. Fontaine (2017) used Halliday's categorisation of verbs into three main groups. These include lexical verbs such as write, modal auxiliary verbs such as "could" and "can", and auxiliary verbs. Unlike primary verb "be", the other forms of primary verbs have and do are challenging for Arab students because Arabic does not contain equivalent verbs in collocations. Therefore, most Arab students find them difficult when used in collocations. A similar case applies to the auxiliaries "be", "do", and "have" when deployed as helping verbs. Therefore, most students omit the primary verb "be" in their writing because it does not exist in Arabic.

Additionally, no pronouns, determiners, articles, conjunctions, punctuation, or quantifiers will be included in the extraction process because the study is only focused on collocations. Kolesnikova (2016) noted that grammatical words such as articles frequently occur; hence, they should not be considered in collocation assessment but considered in the word count.

Chapter 4

Additionally, propositional verbs such as 'live in' and 'sit in' will be excluded since this study will only focus on lexical combinations.

The last stage in the extraction process is whether or not the collocation should be examined as adjectives or in a broader span, such as *the book on the table is small*. Paulsen (2018) opined that taking into account linguistic boundaries when obtaining collocations can be challenging to establish the span. Ziafar and Namaziandost (2019) also noted similar problems associated with limits occurring even in individual words. The author underlined the challenge of identifying a definite point where to start or commence words like 'English-speaking' and determining whether the words should have hyphens or should be two separate words. In other words, authors used different word spans based on their methodologies and aims. For instance, Ziafar and Namaziandost (2019) considered four to five-word windows as suitable spans for collocations. Nevertheless, a collocation word count span and how to limit the span is the researcher's decision.

A five-word span limit can be used. However, as Demir (2017) is concerned, this might be ineffective, leading to many linguistically uninteresting combinations that might be frequent but not meaningful collocations. Considering different learners' language proficiency levels and the possibility that incorrect use of punctuations and run-on sentences might frequently occur, analysing sentences might be difficult. Similarly, smaller windows such as two or three-word span will not be considered as this might result in the omission of relevant arrangements. Likewise, using a five-word window limit might not be ideal in the proposed study. This word span might help include various collocations from students' essays; it might generate uninteresting combinations, making it complicated when choosing collocations. Nonetheless, if the span does not exceed a sentence, there is a high possibility of finding suitable collocations.

Although adjacent collocation is essential, collocations in a broader span will need to be explored to understand learners' collocation use better. In their study, Demir (2017) obtained collocations manually and restricted their extraction to nearby collocations, precisely, words that appear next. Only the nearby one to the modified noun will be obtained when the identified collocation has two adjectives. Nonetheless, this approach has some limitations that will need to be addressed. For instance, if a learner writes *clean house*, the juxtaposition is adjacent and will be deemed adjective-noun collocation. However, if the same learner writes *a clean and nice house*, such collocation is not considered adjacent and will be omitted in this study. This implies that using lexical groupings shorter than a sentence will lead to loss of the collocations that might not be adjacent but have an understandable meaning with noun node. In this respect, a five-word span will be deployed with a clause instead of a sentence. Using clause boundary limit will allow

the identification of many lexical groupings in the learners' texts and more remarkable collocations, which will be omitted if only adjacent collocations are considered.

A clause consists of two sections; the predicate and subject (Ziafar and Namaziandost, 2019). The current study will use this definition but include conjunctions. Therefore, sentences will be broken into short clauses, and learner collocation arrangements will be identified. The process will begin with the noun node and then the identification of collocates. That is, nouns and adjectives, five words before and after. After extracting learner combinations from all the sampled essays, they will be keyed into tables based on their forms and then recognised in the corpus.

4.5 Identifying Collocations in the Corpora

Learners' written texts will be checked against COCA and BNC to check whether or not their collocations make acceptable collocations. The two corpora form a well-rounded and comprehensive research tool that provides a great insight into modern English language usage, independent of the variety of English being used, as they provide a comprehensive and well-rounded look at English as a modern language. Since each corpus consists of a variety of texts from different genres and time periods, it offers a unique selection of texts that makes it a valuable resource. The global use of the English language continuously shapes the way it is used in English-speaking countries. By using both of these corpora, we are able to gain a better understanding of how the situation is overall. The BNC will be selected to represent British English users' language use. Additionally, the BNC is commonly used and proven to be a reliable reference corpus. The BNC has more than 95 million modern British English words. The software also contains data from written and spoken English texts presented by varied genders and stages and in different domains and topics. The tool is also readily accessible online in various forms mainly used for linguistic research, primarily the Sketch Engine. Sketch Engine will suit this study because it involves all relations measurements to investigate juxtaposition strength, including LogDice and t-score. This study will use the LogDice scale though other measurements might be referred to. COCA is also used to complement the BNC in assessing American English as a corpus-related instrument.

4.6 Distribution of Collocations in Learner Corpora

Relative frequency, also referred to as IPM (instances per million), is the frequency at which an item can be found (hits) compared to the number of tokens in a million. It is used to compare frequencies between corpora of different sizes. In order to facilitate comparisons and

Chapter 4

interpretation, relative frequency is used instead of raw frequency in order to provide an easy way to understand the distribution of frequency, which makes it easier to compare and interpret. Additionally, log dice scores are used to measure the association between collocations. ANOVA was used to carry out a statistical analysis of the data, which was then compared to other data sets. Two additional steps were taken in order to compensate for the non-normal distribution of data, in order to make it more accurate. The first step involved approximating normality using the log10 transformation as the first step. In addition to the Kruskal-Wallis ANOVA and the One-way ANOVA, a non-parametric test of normality was used in order to determine whether this data was normal. In order to test whether the continuous data (source) are equal after the data has been compiled, we have used the Kolmogorov–Smirnov (K-S test) nonparametric test. For the parametric tests Mean and Standard Deviation are used while mean and median ranking are taken for the non-parametric tests. Strength and direction of relationship between year of study (proficiency level) and relative frequency of collocations was measured using **Spearman's correlation**. "Spearman's rank-order correlation calculates a coefficient, r_s or ρ (pronounced "rho"), which is a measure of the strength and direction of the association/relationship between two continuous or ordinal variables."

4.7 Degree of Fixedness for the Identified Collocations

LogDice scale of rating collocations was used to verify collocations based on low, medium, and high scores. Different researchers use different bands of threshold scores to identify collocations at each of the three levels. For instance, Gablasova, Brezina and McEnery (2017) used a band of 3.3/6/7/11 for the three levels respectively. (A or B)

Ellis et al. (2008) had a band of 3.3/6.7/11 which corresponded to low, medium, and high respectively, whereas Granger and Bestgen (2014) had non-collocation with 3*, low collocation with 3-4.99, medium collocation with 5-6.99, and high collocation for 7 and more.

Hence, in dividing the scale into levels, collocations with scores 0.1-3.5 are rated as low, 3.6-7 are rated medium, 7.1-10.5 are rated high, and 10.6-14 are rated advanced.

In order to classify collocations according to the degree of fixedness, this research follow (Khoja, 2019) threshold. The scale divides collocations based on their LogDice score. Collocations with scores 0.1-3.5 are rated as low, 3.6-7 are rated medium, 7.1-10.5 are rated high, and 10.6-14 are rated advanced. Apportioning the scale into many levels will ease the classification process and capture all the proficiency levels. Additionally, utilising this scale will add value to the deployment of the association dimension LogDice by showing and categorising similarities or differences between presented collocations and learners' levels as opposed to merely individual production of

collocation and non-collocations. Khoja (2019) acknowledged this as a practical approach to trace L2 students' progress in collocation production across a scale of association measurement.

After identifying the collocations at each learner level as well as their levels of fixedness in the BNC, the descriptive statistics technique will be deployed to compare and describe the results of the study. Noteworthy, repeated forms of collocations will be considered as one example when presented in a single text. Still, when the same collocations re-occur in a different text, it will be identified as another example. Therefore, regardless of the number of times a single learner writes it in the text, it will be considered one example but counted again if it appears in another context. This is because the study mainly concentrates on assessing the collocations used by students instead of the frequency of collocations used.

An analytical process comprising extracting and identifying the combination of collocations from students' written essays will help examine how the learners can produce fixed and acceptable collocations and determine the forms of collocations. Besides, since this procedure will be used in all the learners' levels, it will also facilitate comparison. Nevertheless, the comparison between the levels might be related to various issues such as places such as less idiomatic arrangements and where it has been specifically referred to as discussed in the literature review. The discussion of such instances will be supported by various tools, including the Arabic-English dictionary, the Oxford Arabic Dictionary for examining collocations, and the Arabic-English Lexicon for using verbs.

Additionally, this analysis process will help investigate the degree to which the students can produce fixed and acceptable collocations and identify the forms of collocations.

Building Corpus using Sketch Engine

Sketch Engine

The Sketch Engine (Kilgarriff et al., 2004, 2014) is a commercial web based tool for corpus analysis developed by Lexical Computing Ltd. In addition to the corpora searching tool, the users are provided with corpora in many languages including Arabic. Along with the usual features of such tools (e.g. concordance, word lists, key words, collocation, and corpus comparison), Sketch Engine has some unique features such as Word 11 Sketches that provide summaries of a word's grammatical and collocational behaviour, Word Sketch Difference to compare and contrast words visually, and WebBootCat, which lets users create specialised corpora from the Web.

Sketch Engine is an online tool that allows users to analyse texts. It is a tool that helps exploring how language works through algorithms analyses of authentic texts in order to identify what is

Chapter 4

typical in language and what is rare, unusual or emerging usage. Sketch Engine allows generate word list of the most frequent words, combinations, and comparisons between corpora.

After collecting data, I went through students' writings and with a help of data entry professional I transcribed students writing into txt files classified them according to their level and university and uploaded them to Sketch Engine for analysis. As mention in data collection section, and after eliminating some samples I built a corpus 553 participants and consist of 74,996 words.

Collocations Extraction

To extract collocations from the corpus, it is important to add is to add tags or labels to facilitate locating collocations. One way is using part of speech (POS) tagging which is the process of marking up a word in a text as corresponding to the specific word–class combinations minding the word's definition and its context—i.e., its relationship with adjacent and related words in a phrase, sentence, or paragraph. This study is going to target three lexical collocations that fall into the following types of POS combinations: verb + noun (e.g., take advice), adjective + noun (e.g., heavy traffic), and noun + noun (e.g., car park). The corpora will be first tagged manually using paper and pencil before typing texts in to the computer. Then I will use an automated online tagging service developed by (Sketchengine.co.uk, 2018). The website designed the “tagest” tool that could include tags for the most common parts of speech (N for noun, V for verb, A for adjective, etc.). Also, it could go into more detail and distinguish between nouns in singular and plural, verbal conjugations, tenses, aspect, voice and much more. What makes POS tagest tool offered by Sketch Engine special is that it allows researchers to “develop their own very specialized tagsets to accommodate their research needs.” (Sketchengine.co.uk, 2018).

The second phase was carried out by checking and identifying which POS was prominent for top nouns located in corpora.

are committed to pro bono. As	we work closely PP VVP RB	with pro bono partners and o
more important than ever that	we work together PP VVP RB	with other organisations on o
re that most people succeed if	they work hard	The global economic crisis l
air grades and sub	He has worked hard PP VHZ VVN RB	items that directly
performed in make		trying to see how
it, and our Irish predecessors.	We have worked hard PP VHP VVN RB	to achieve the consensus nei
re can achieve great results, if	we work together PP VVP RB	. I am grateful for your attent
be close to family and friends.	He has worked hard PP VHZ VVN RB	to get back the use of his arm
ensing to commercial entities.	She works actively PP VVZ RB	with investigators in negotiati

Figure 6 Concordance with POS tags displayed (Sketchengine.co.uk, 2018)

Word Sketches

Word Sketch is a tool offered in Sketch Engine (Sketchengine.co.uk, 2018) that has a “very complex and sophisticated algorithm” help individuals by identifying collocations and calculating frequency of every word combination to determine their strength or typicality. In order to extract most frequent collocation, you have first to insert the lemma and choose what part of speech that you would like to find what associating words the often come with.

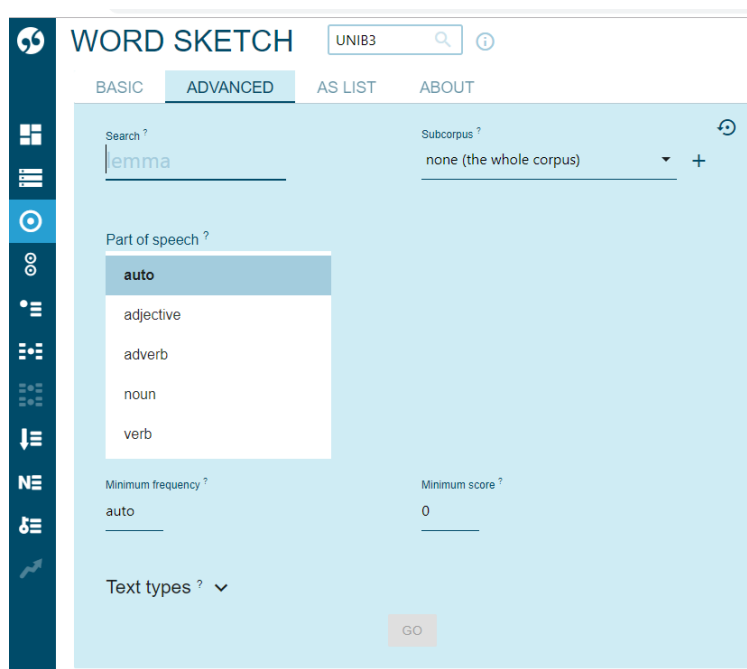


Figure 7 Word Sketch Interface

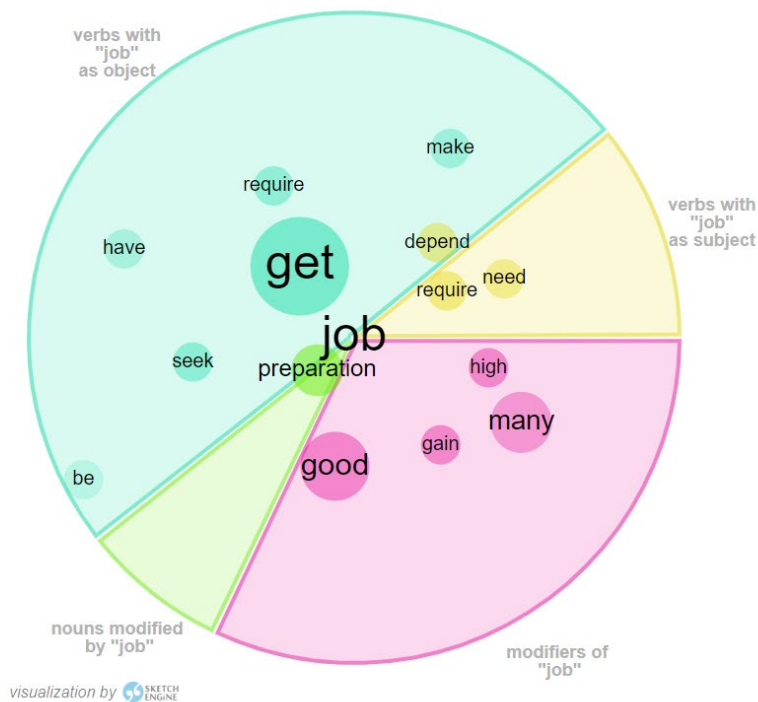


Figure 8 Combinations that has the word "job"

For example, I tried to look for what combinations are associated with the word “job” and as shown above is a diagram of most frequent word combinations divided into four categories modifiers, nouns, objects or subjects of a verb. Based on the diagram above the most frequent and strong collocations are:

POS	Collocation	Frequency	Log Dice Score
Verb + Noun	Get a job	9	12.88
Adj + Noun	Good job	4	11.87
Pro + Noun	Your job	4	10.34

Figure 9 Example of Combinations Strength using Word Sketch

Methodological Limitation

This process for extracting learner collocations combination has one limitation. As Khoja (2019) noted, co-occurrence of collocations can be assessed through three criteria: textual, syntactic, and surface co-occurrence. Surface co-occurrence alludes to word span from one collocate to another; syntactic co-occurrence means that the syntax collocates share, while textual co-occurrence refers to the unit in which collocates occur (Smadja, 2021). In most cases,

collocation blends appear in a five-word span within specified clause boundaries. This addresses the textual and surface co-occurrence. However, the clauses in the learner's text might not reflect the bond of the two collocations straightforwardly. This is important because this study will investigate the collocation of two words recognised based on corpus strategy.

To avoid this limitation, the extraction methodology will be adapted by further considering the syntactic correlation of two collocates in the learner's collocations combination. This implies that noun-verb and verb-noun juxtapositions can involve verb-object and subject-verb, but some prepositional phrases are excluded. The syntactic relations within adjective-noun collocations will be noted when the collocates are modifiers, and the two nouns co-occur in a list. Limiting the extraction in this order will help have a systematic, replicable process and a more focused investigation.

Summary

This chapter has outlined the approaches that will be adopted in this study. The study will use three triangulate approaches; the CIA method, the cross-sectional approach, and the frequency-based approach to foster methodological credibility. Such combination will be necessary because the learners' written texts will be obtained manually and examined qualitatively when comparing different learners' levels.

This study will use qualitative as well as quantitative methods to strengthen the analysis further. The qualitative approach will involve statistical tests for collocation associations. The context requires the application of an analytical process followed, beginning from the extraction process of learner collocations combinations. While manual extraction might have some limits, such as less accuracy than software tools, it will add to the study's validity.

Chapter 5 Findings

This chapter examines the data in relation to each research questions. Section (5.1) addresses the first research question “What type of collocations is frequently used among learners of three different years of study?”. Section (5.2) examines the second research question: “What the degree of fixedness for identified collocations found in learners’ learners writing at each year of study and across the academic year? . Section (5.3) discusses the third research question: “ What progress have students made in their collocation production during their studies? does the learners' first language (L1) factor into their production of collocations in English?”

5.1 Extracting Learners’ Collocations

As previously mentioned in the methodology chapter Sketch Engine is used to extract and identify learners’ collocations. This was done in two steps. First, collocations are identified as acceptable collocations they occurred five times or more in the reference corpora BNC & COCA . his is to demonstrate common occurrences and their conventional availability according to learners' meanings. Second, extracted learners’ collocations should score more than zero on the LogDice. After that, all identified collocations are classified as low, medium, high, or advanced levels of fixedness in order to illustrate the level of fixedness.

Overall results show verb- noun collocations are the most used type of collocations found in students writing with overall 147 instances are found that takes 50% of overall collocations. Followed by adjective-noun collocations as 121 instances are found that takes 41.2% of overall collocations. Noun-noun collocations came last with 26 instances are found in students writing and it takes only 8.8% of overall collocations.

Looking specifically to collocation production of type and year of study, see Table 5. It shows throughout the study verb- noun collocation was the most used collocations in Year 1 (54.3%) and in Year 2 (52.7%). Where in Year 3 verb- noun collocation production was equivalent to adjective-noun collocations

Table 5: Crosstabulation for year of study and type of collocation

			type			Total
			Adjective + noun Collocation	Noun + noun collocation	Verb + noun collocations	
year	1.00	Count	28	4	38	70
		% within year	40.0%	5.7%	54.3%	100.0%
		% within type	23.1%	15.4%	25.9%	23.8%

2.00	Count	42	10	58	110
	% within year	38.2%	9.1%	52.7%	100.0%
	% within type	34.7%	38.5%	39.5%	37.4%
3.00	Count	51	12	51	114
	% within year	44.7%	10.5%	44.7%	100.0%
	% within type	42.1%	46.2%	34.7%	38.8%
Total	Count	121	26	147	294
	% within year	41.2%	8.8%	50.0%	100.0%
	% within type	100.0%	100.0%	100.0%	100.0%

As stated before, relative frequency of collocations is considered to see the difference between collocations produced across years of study, irrespective of collocation type. Figure 10 shows the variation between Year 1, Year 2, Year 3. However, and due to extreme values of relative frequencies, it was not clear any difference. Hence, it was necessary to make a transformation by taking Log10 to reduce the variation within each year and make the data normally distributed and carry the parametric test ANOVA. .

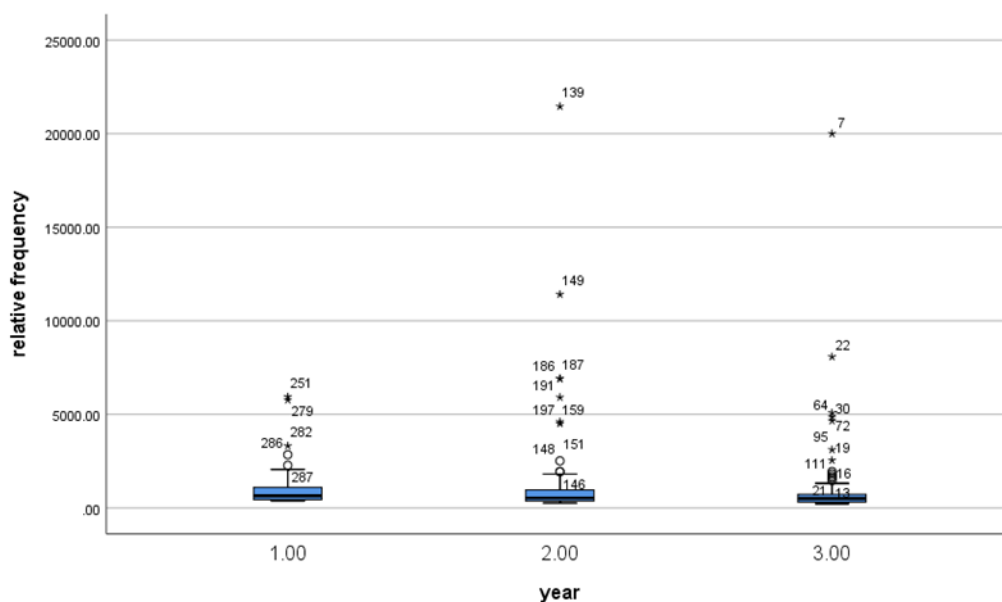


Figure 10 Distribution of relative frequency between the three years

Chapter 5

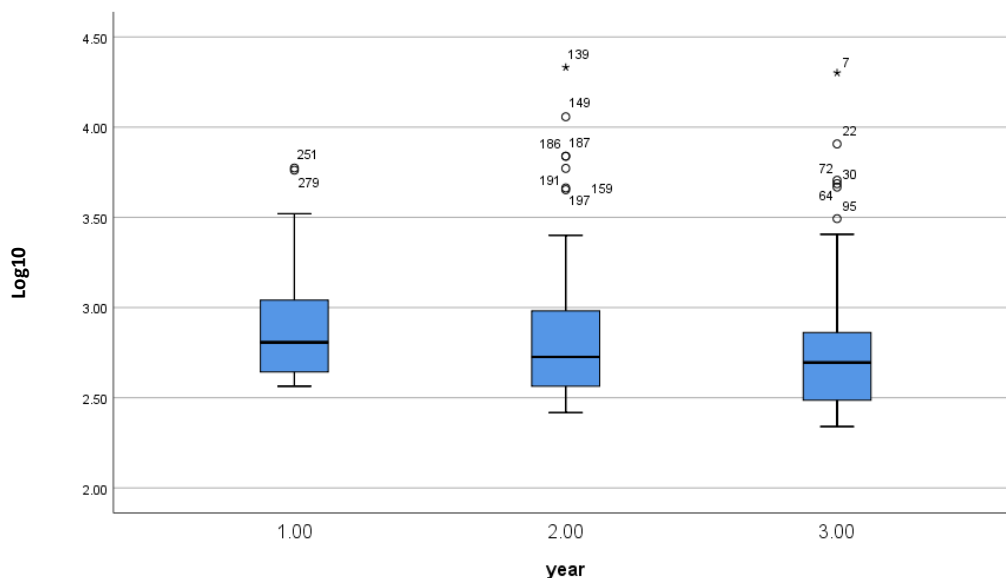


Figure 11 Distribution of Log10 relative frequency between the three years

Table 6 Antilog10 summary statistics of relative frequency and ANOVA between the three years

	N	Mean	Std. Deviation	ANOVA	
				F	p-value
Year 1	70	758.57	1.96	3.732	.025
Year 2	109	676.08	2.34		
Year 3	114	549.54	2.25		

For antilog10 of relative frequency, retrieve the actual value of relative frequency, indicated that Year 1 showed a slightly higher value Log10 relative frequency (mean=758.57) than Year 2 (mean=676.08) and Year 3 (mean=549.54), and Year 2 tended to be slightly higher than Year 3. Table6. One way-ANOVA indicates that there is significant difference in relative frequency between the three years ($F=3.73$, $p\text{-value}=.025$).

5.1.1 Verb- Noun Collocations

To see verb-noun production among students in the three years of study, K-S test is used to compare relative frequency of collocations. The test showed that the relative frequencies or Log10- relative frequencies are not normally distributed. See Appendix B (tables 44,45,46). Hence, Kruskal-Wallis ANOVA (non-parametric test) is used to compare the relative frequency of verb+ noun collocation. Median is used rather than mean to present average relative frequencies. It is

observed for relative frequency that Year 1 (median=659.63) was higher than Year 2 (Median=532.70) and Year 3 (Median=509.09).

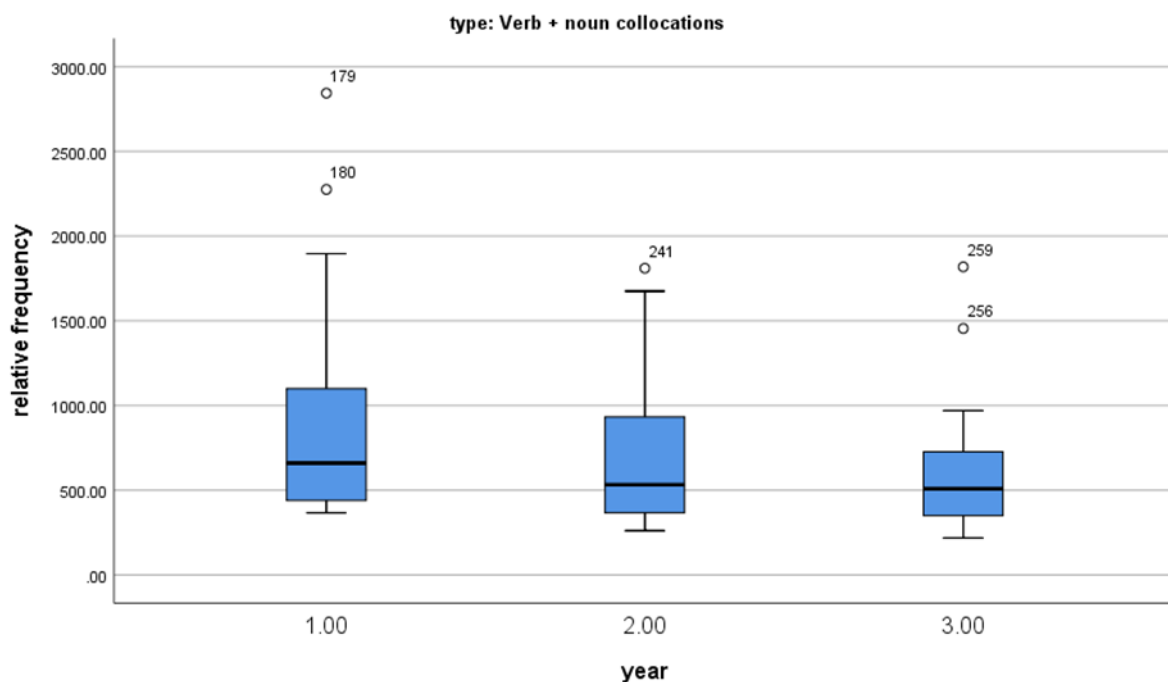


Figure 12 Relative frequency of verb+ noun collocation for year one, two and three.

Table 7, there was significant difference ($X^2=9.93$, $p\text{-value}=.012$) between the relative frequency of verb+ noun Collocation between the proficiencies.

Table 7 Summary statistics and ANOVA between the three years for relative frequency of verb + noun Collocation.

	N	Median	Mena rank	Kruskal-Wallis ANOVA	
				X^2	p-value
Year 1	38	659.63	89.74	9.93	.012
Year 2	58	532.36	73.81		
Year 3	51	509.09	62.49		

Comparing association measures between log dice scores of verb + noun collocations. It is noted that the average looked the very close between Year 1 on ($M=10.80$, $SD=1.56$), Year 2 ($M=10.98$, $SD=1.28$) and Year 3 ($M=10.65$, $SD=1.50$), Table 8 and Figure 13. Using the ANOVA test, show that there is no significant difference in log dice score ($F=.748$, $p\text{-value}=.475$).

Chapter 5

Table 8 Summary statistics and ANOVA between the three years for Log Dice score of verb + noun Collocation.

	N	Mean	SD	ANOVA	
				F	p-value
Year 1	38	10.80	1.56	.748	.475
Year 2	58	10.98	1.28		
Year 3	51	10.65	1.50		

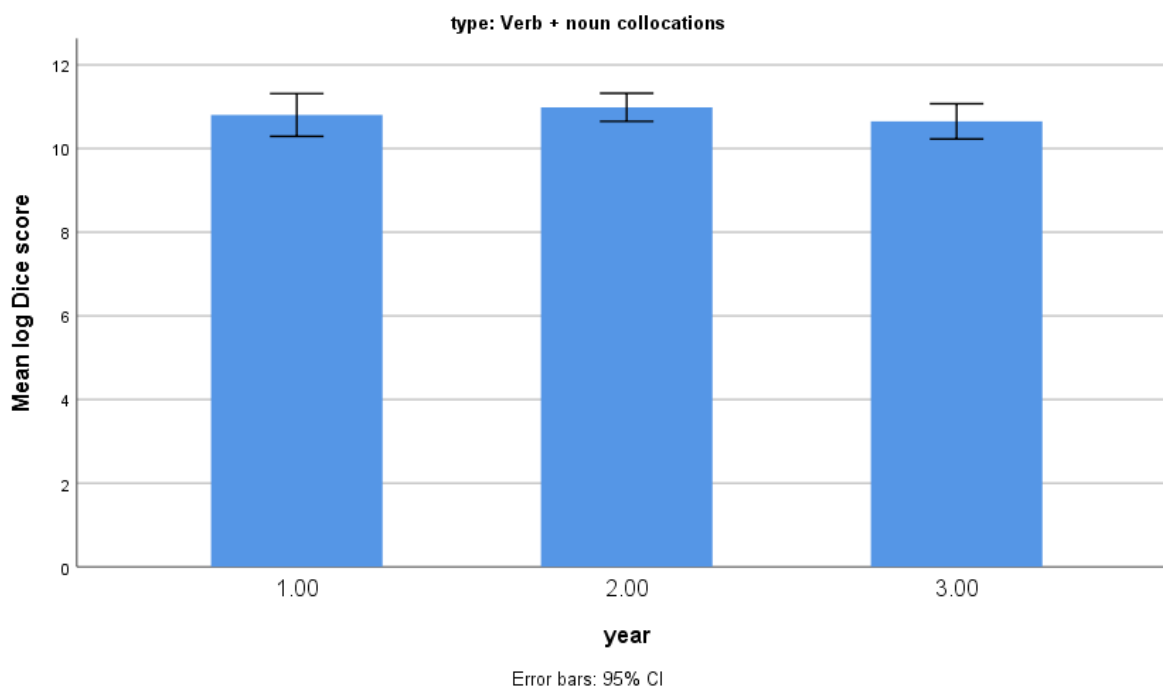


Figure 13 Mean log dice score of verb+ noun collocation for year one, two and three.

As mentioned previously, Spearman rank correlation is going to be used to see strength and direction of relationship between years of study. When applying Spearman's correlation, It shows that there was negative correlation between the of verb + noun collocation and years of study ($r=-0.245$, $p\text{-value}=.003$). (table 48 see appendix B)

5.1.2 Adjective- Noun Collocations

Similarly, to verb + noun collocations, adjective-noun collocations are analysed using K-S test is to compare relative frequency of collocations across years of study. The test showed that the relative frequencies or Log10- relative frequencies are not normally distributed. see the appendix.

Hence, Kruskal-Wallis ANOVA (non-parametric test) is used to compare the relative frequency of adjective+ noun collocation. Median is used rather than mean to present average relative frequencies. It is noticed that Year 1 (median=560.26) and Year 2 (median=5599.66) had a higher relative frequency of adjective + noun collocation compared to Year 3 (Median=437.77). Using Kruskal-Wallis ANOVA, Table 9, there was no significant difference ($X^2=5.83$, $p\text{-value}=.054$) between the relative frequency of adjective + noun collocation between the years of study.

Table 9 Summary statistics and ANOVA between the three years for relative frequency of adjective + noun Collocation.

	N	Median	Mena rank	Kruskal-Wallis ANOVA	
				X^2	p-value
Year 1	28	560.26	69.75	5.83	.054
Year 2	41	559.66	64.98		
Year 3	51	437.77	51.82		

Comparing association measures between log dice scores of adjective + noun collocations. It is noticed that there is no more difference between Year 1 ($M=10.76$, $SD=1.42$), Year 2 ($M=11.01$, $SD=1.43$) and Year 3 ($M=11.17$, $SD=1.23$), Table 10 and Figure 14. Using the ANOVA test, there was no significant difference in log dice score ($F=0.849$, $p\text{-value}=.431$).

Table 10 Summary statistics and ANOVA between the three years for Log Dice score of adjective + noun collocation.

	N	Mean	SD	ANOVA	
				F	p-value
Year 1	28	10.76	1.42	.849	.431
Year 2	41	11.01	1.43		
Year 3	51	11.17	1.23		

Chapter 5

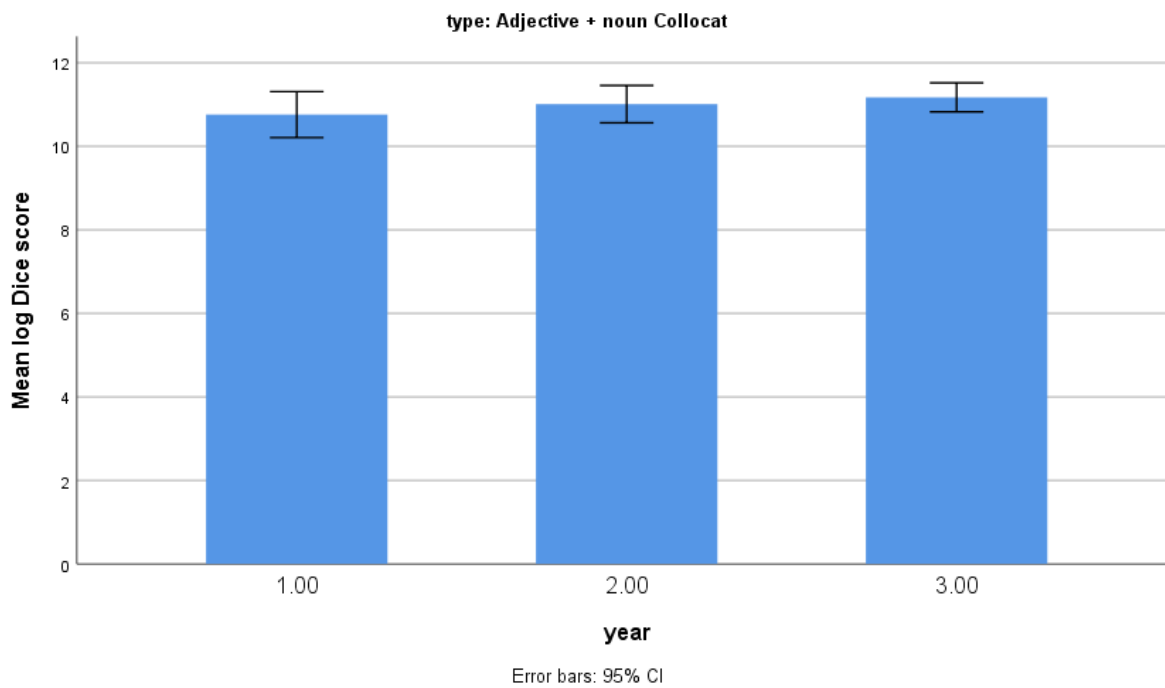


Figure 14 Mean log dice score of adjective+ noun collocation for year one, two and three.

Spearman's correlation test of strength and direction of relationship between years of study showed a significant negative correlation between adjective + noun collocations and Years 1,2, and 3 ($r=-0.218$, $p\text{-value}=.017$). (table 48 see appendix b)

Table 11 Spearman's correlation between relative frequency of adjective + noun collocation and Years 1,2, and 3

type	year	relative frequency
Adjective + noun Collocation	Correlation Coefficient	1.000
	p-value	.017
	Correlation Coefficient	-.218*
relative frequency	Correlation Coefficient	-.218*
	p-value	.017
	Correlation Coefficient	1.000

5.1.3 Noun-noun Collocations

Similarly, to verb + noun collocations, noun-noun collocations are analysed using K-S test is to compare relative frequency of collocations across years of study. The test showed that the relative frequencies or Log10- relative frequencies are not normally distributed. see the appendix.

Hence, Kruskal-Wallis ANOVA (non-parametric test) is used to compare the relative frequency of noun+ noun collocations. Median is used rather than mean to present average relative frequencies. It is noticed for relative frequency that Year 1 (median=1164.64) was higher Year 2 (Median=369.70) and Year 3 (Median=521.91), and also year one showed high variation in relative frequency, Figure 15.

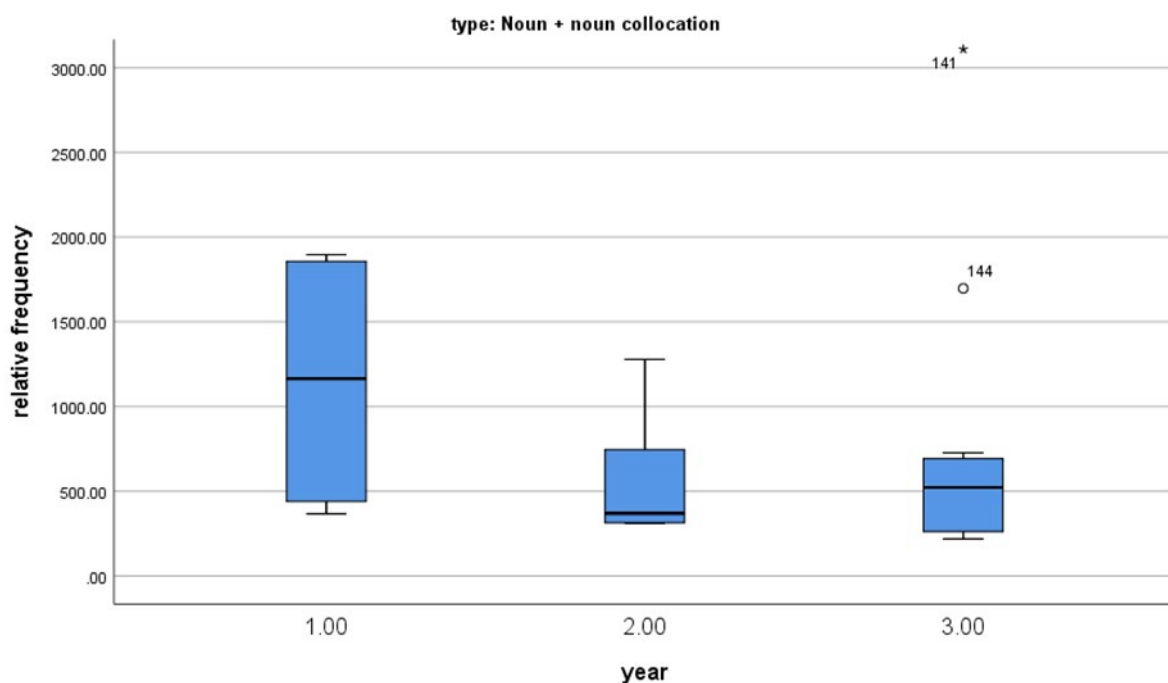


Figure 15 Relative frequency of noun+ noun collocation for year one, two and three.

Based on Kruskal-Wallis ANOVA, Table 12, there was no significant difference ($X^2=1.67$, p -value=.433) between the relative frequency of noun+ noun collocation between the proficiencies.

Table 12 Summary statistics and ANOVA between the three years for relative frequency of noun + noun Collocation.

	N	Median	Mena rank	Kruskal-Wallis ANOVA	
				X^2	p-value
Year 1	4	1164.64	18.00	1.67	.433
Year 2	10	369.70	13.00		
Year 3	12	521.91	12.42		

Comparing log dice scores for the association measures. It is noticed showed higher average in Year 1 ($M=13.68$, $SD=0.378$) than Year 2 ($M=12.74$, $SD=1.82$) and Year 3 ($M=11.93$, $SD=1.40$) in

Chapter 5

noun-noun collocations, Table 12 and Figure 15. Using the ANOVA test, there was no significant difference in log dice score ($F=2.227$, $p\text{-value}=.131$).

Table 13 Summary statistics and ANOVA between the three years for Log Dice score of noun + noun collocation.

	N	Mean	SD	ANOVA	
				F	p-value
Year 1	4	13.68	.378	2.227	.131
Year 2	10	12.74	1.82		
Year 3	12	11.93	1.40		

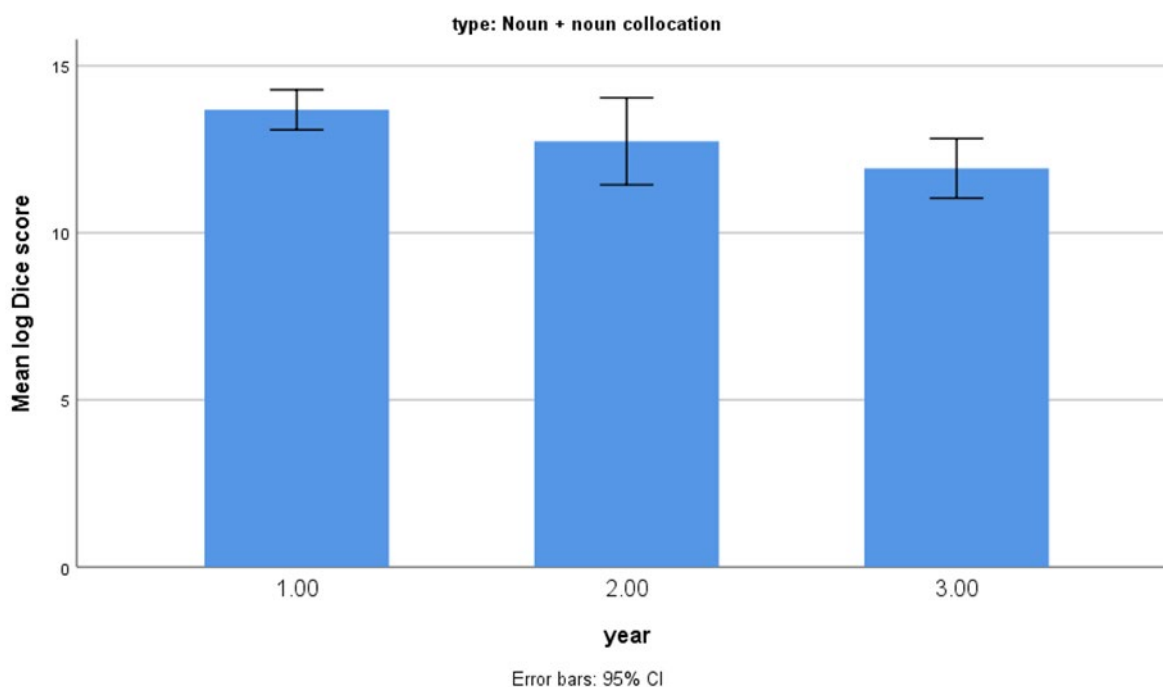


Figure 16 Mean log dice score of noun+ noun collocation for year one, two and three.

Using Spearman's rank-order correlations to assess the relationship between noun+ noun collocations and Years 1,2, and 3 as seen in (table 48 see appendix B), it shows negative correlation between relative frequency of noun+ noun collocation and years of study ($r=-0.201$, $p\text{-value}=.0326$), but this correlation was not significant. (table 48 see appendix B)

5.2 Degree of Fixedness for the Identified Collocations

Sketch Engine is used to search for extracting learners collocations. This was done in two steps to confirm when score in the BNC (British National Corpora) of five citations or more will first be deemed an acceptable collocation. Furthermore, learners should score more than zero on the LogDice. After that, all identified collocations are classified as low, medium, or high levels of fixedness in order to illustrate how well they produce acceptable collocations.

In order to classify collocations according to the degree of fixedness, this research follow (Khoja, 2019) threshold. The scale divides collocations based on their LogDice score. Collocations with scores 0.1-3.5 are rated as low, 3.6-7 are rated medium, 7.1-10.5 are rated high, and 10.6-14 are rated advanced.

5.2.1 The First Set of Data Collection

a. Collocations Extracted from Year 1

Verb-Noun Collocations

As mentioned before verb-noun collocations is the most used collocation overall in all years of study followed by adjective- noun collocations.

In terms of collocation findings per year of study the findings show learners at year 1 produced more verb-noun collocations than any other kind of collocations (54.3%). It is then followed by adjective-noun collocations (40%). Noun-noun collocations came last and scored only (5.7%).

Table 14 Degree of Fixedness for Verb + Noun collocations identified from Year1 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Communicate with people	9.32	High
Communicate with the world	7.04	Medium
Get a job	7.79	High
Have a job	4.91	Medium
Have fun	11.72	Advanced

Chapter 5

Have a language	1.82	Low
Have reason	5.21	Medium
Know a language	3.05	Low
Know a lot	5.86	Medium
Know culture	4.43	Medium
Learn English	5.22	Medium
Learn something	9.14	High
Learn thing	8.00	High
Make friends	6.19	Medium
Read a book	9.74	High
Speak a language	9.29	High
Speak English	7.74	High
Study language	5.18	Medium
Talk with people	3.88	Medium
Use English	6.62	Medium
Use language	6.90	Medium
Watch a movie	6.34	Medium
Work in a country	3.35	Low

Table 13 above shows collocations extracted from Year 1 learners in the first set of data collection. The table also shows the level of fixedness and accessibility of collocation according to the threshold score. According to the proposed classification presented in the methodology section, the identified collocations were listed as follows. Most of the collocations identified were high fixedness, scoring between 7.79%-9.74%. A total of six collocations scored a medium level of fixedness 3.88%- 7.04%. The last four collocations had a low score on the level of fixedness scale 1.82%-3.25%.

Adjective-Noun Collocations

When it comes to adjective- noun collocations they were harder to identify especially in BNC. One explanation by Lorenz (1999) is that native speakers use more infrequent adjective- noun

combinations. Table below shows adjective-noun_collocations extracted from Year 1 learners in the first set of data collection.

Table 15 Degree of Fixedness for Adj + Noun collocations identified from Year1 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Different Country	2.48	Low
Different Language	2.94	Low
Different Reasons	7.17	High
English Language	5.23	Medium
First Language	5.29	Medium
Foreign Language	10.01	High
Good Job	0.84	Low
Good Way	7.39	High
Important Language	2.77	Low
Important Thing	4.82	Medium
Many People	7.38	High
Many Reasons	2.66	Low
Many Ways	3.40	Low
More Languages	1.90	Low
Most People	7.47	High
New Job	7.54	High
New Thing	1.70	Low
Other Country	2.75	Low
Other Cultures	4.80	Medium
Other Languages	3.45	Low
Other People	2.42	Low
Other Reason	3.67	Medium
Second Language	6.77	Medium

Chapter 5

In accordance with the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. The majority of colloquial expressions 11 combinations scored low on LogDice, ranging from 0.84-3.45 The level of fixedness for 6 collocations was medium 3.67-6.76. Also, 6 collocations scored high 7.17-10.01 on the level for was.

Noun-Noun Collocations

Noun- noun collocations are the least type of collocation found in learners writing. Two collocations are found in first set of data for year 1 as shown in the table below.

Table 16 Degree of Fixedness for Noun + Noun collocations identified from Year1 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Point of view	11.11	Advanced
Mother tongue	7.75	High

Learners managed to use this collocation in similar method to native speakers. One student wrote *"In my point of view I think learning languages is so important to communicate with others."*. in the BNC *"Point of View"* appeared in a similar fashion, *"I'm just telling you my thoughts, trying to put across my point of view for a change."*

b. Collocations Extracted from Year 2

Verb-Noun Collocations

Table 17 Degree of Fixedness for Verb + Noun collocations identified from Year2 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Get a job	7.79	High
Have a job	4.91	Medium
Have a language	1.82	Low
Have a point	4.12	Medium
Have reasons	5.21	Medium
Know a language	3.05	Low
Know culture	4.43	Medium

Learn a culture	4.70	Medium
Learn a thing	8.00	High
Learn English	5.22	Medium
Love language	3.29	Low
Make a person	4.20	Medium
Make friends	6.19	Medium
See the world	5.19	Medium
Speak a language	9.29	High
Speak English	7.74	High
Study English	2.25	Low
Study language	5.18	Medium
Take a course	6.67	Medium
Use language	6.90	Medium
Watch a movie	6.34	Medium
Improve skills	8.28	High
Learn language	8.89	High

Table above shows verb -noun collocations extracted from Year 2 learners in the first set of data collection. The table also shows the level of fixedness and accessibility of collocation according to the threshold score. Most of the collocations identified were high fixedness, scoring between 7.1-10.5 on LogDice. A total of six collocations scored a medium level of fixedness on LogDice, 3.6-7. The last four collocations had a low score on the level of fixedness scale.

Adjective-Noun Collocations

When it comes to adjective- noun collocations they were harder to identify. Table below shows adjective-noun_collocations extracted from Year 1 learners and the first set of data collection.:

Table 18 Degree of Fixedness for Adj + Noun collocations identified from Year2 the first phase of data collection

Chapter 5

Collocation	LogDice Score	Level of Fixedness
Common language	7.27	High
Different country	2.48	Low
Different culture	3.81	Medium
Different language	2.94	Low
Different reasons	7.17	High
Different way	4.77	Medium
Easy way	3.57	Low
English language	5.23	Medium
First language	5.29	Medium
Foreign country	9.55	High
Foreign language	10.01	High
Foreign people	4.64	Medium
Helpful ways	6.41	Medium
Important language	2.77	Low
Important thing	4.82	Medium
International language	6.18	Medium
Many people	7.38	High
Many reasons	2.66	Low
Many ways	3.40	Low
Native speaker	10.97	Advance
New language	5.28	Medium
New thing	1.70	Low
Other country	2.75	Low
Other language	3.45	Low
Other people	2.42	Low
Other reason	3.67	Medium
Only reason	6.19	Medium

Own reason	5.04	Medium
Second language	6.77	Medium
Whole world	4.46	Medium

In accordance with the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. The majority of colloquial expressions (14 combinations) scored medium on LogDice, ranging from 3.67-6.77. The level of fixedness for collocations was Low 1.70-357. Also, 5 collocations scored high 7.17-10.01 on the level for was. 1 advance 10.97

Noun-Noun Collocations

There was only one noun-noun collocation identified in BNC

Table 19 Degree of Fixedness for Noun + Noun collocations identified from Year2 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Point of view	11.11	Advanced
Mother tongue	7.75	High
Tv Show	9.22	High

In accordance with the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. The majority of colloquial expressions (2 combinations) scored high on LogDice, ranging from 7.75-9.22. The level of fixedness for one collocation was advanced 11.11.

All noun-noun collocations extracted here are at a high- level of fixedness. One unidentified collocation was found 6 times in Year 2 writing: “Mother Language”. *Learners used this collocation when they want to talk about their native language. One learner writes “You have to know a new language but you have not to forget your **mother language** and used it in your daily life.” One explanation for the use of “mother language” instead of more native like collocations such as, “mother tongue”, “native language” or “first language” is that is a literal translation of Arabic collocation اللغة الأم that literary translate to “mother language”. It is the common collocation used in Arabic to talk about one’s native language. Interestingly, one student used both collocation in their writing: “In my opinion because in now times learning other or different*

Chapter 5

languages from your mother tongue is necessary. For example, to work in a company you must have 2 different languages beside your mother language.”

c. Collocation Extracted from Year 3

As mentioned before verb-noun collocations is the most used collocation overall in all years of study followed by adjective- noun collocations.

In terms of collocation findings per year of study the findings show learners at year 3 produced more verb-noun collocations than any other kind of collocations (54.3%). It is then followed by adjective-noun collocations (40%). Noun-noun collocations came last and scored only 5.7%.

Verb-Noun Collocations

As mentioned above students in their third year of study used more than any other kind of collocations. They depended more on using action verbs in their word combinations. As seen in table below:

Table 20 Degree of Fixedness for Verb + Noun collocations identified from Year3 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Communicate with people	9.32	High
Expand knowledge	4.04	Medium
Get a job	7.79	High
Have a language	1.82	Low
Have reasons	5.21	Medium
Know a language	3.05	Low
Learn a culture	4.70	Medium
Learn a language	8.89	High
Learn English	5.22	Medium
Love language	3.29	Low

Meet people	8.86	High
Read a book	9.74	High
Speak a language	9.29	High
Take course	6.67	Medium
Talk language	4.98	Medium
Use language	6.90	Medium
Watch a movie	6.34	Medium

In accordance with the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. The majority of colloquial expressions (8 combinations) scored medium on LogDice, ranging from 4.04-6.90. The level of fixedness for 3 collocations was Low 1.82-3.29. Also, 6 collocations scored high 7.79 -9.74.

Adjective-Noun Collocations

When it comes to adjective- noun collocations they were harder to identify. Table below shows adjective-noun_collocations extracted from Year 3 learners and the first set of data collection:

Table 21 Degree of Fixedness for Adj + Noun collocations identified from Year3 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Different language	2.94	Low
English language	5.23	Medium
First language	5.29	Medium
Foreign country	9.55	High
Foreign language	5.29	Medium

Chapter 5

Good way	7.39	High
Important language	2.77	Low
Important thing	4.82	Medium
Many people	7.38	High
Many reasons	2.66	Low
Many ways	3.40	Low
Native speaker	10.97	Advance
New language	5.28	Medium
New thing	1.70	Low
Only reason	6.19	Medium
Other country	2.75	Low
Other cultures	4.80	Medium
Other language	3.45	Low
Other people	2.42	Low
Other reason	3.67	Medium
Real life	9.39	High
Second language	6.77	Medium

Based on the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. On LogDice, nine colloquial expressions received medium scores, ranging from 3.67 to 6.77. Eight collocations scored low 1.70-3.45. There were also four collocations that received high scores of 7.38-9.55, whereas only one collocation scored an advanced score of 10.97.

Noun-Noun Collocations

There was only one noun-noun collocation identified in reference corpora BNC and COCA scored as advanced at 11.11.

Table 22 Degree of Fixedness for Noun + Noun collocations identified from Year3 the first phase of data collection

Collocation	LogDice Score	Level of Fixedness
Point of view	11.11	Advanced

5.2.2 The Second Set of Data Collection

a. Collocation Extracted from Year 1

Verb-Noun Collocations

Table 23 Degree of Fixedness for Verb + Noun collocations identified from Year1 the second phase of data collection

Collocation	Citations from leaners' Text	LogDice Score	Level of Fixedness
Raise Child	I think the parents should <i>raise</i> their <i>children</i> with love	6.83	Medium
Respect time	teach your child how to respect the older people, and how to <i>respect</i> the <i>time</i>	6.04	Medium
Take care	I think it's the parent's responsibility to <i>take care</i> of their children	10.08	High
Teach children	In my opinion the parents should <i>tach</i> their <i>children</i>	10.04	High
Teach everything	I think school cannot <i>teach everything</i>	3.87	Medium

Chapter 5

Respect elder	the school will tell a child how important it is to <i>respect the elders</i>	3.31	low
Have time	schools have big responsibility, sometimes the parents don't <i>have time</i>	7.12	High

The table above contains verb-noun collocations extracted from Year 1 on the second set of data collection. In addition to this, the table indicates the degree of fixedness and accessibility of collocations based on the threshold scores. All collocations identified were also found in BNC.

A total of 1 low level collocation scored at 3.31, 2 medium levels scored between 3.87- 6.83 and 3 High level scored between 7.12-10.08 on LogDice.

Adjective-Noun Collocations

Table 24 Degree of Fixedness for Adj + Noun collocations identified from Year1 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Good manner	I believe that children don't just learn <i>good manners</i> from their parents	5.72	Medium
Good member	I believe that the first steps of making a <i>good society member</i> starts	2.43	Low
Good person	help a child to be a <i>good person</i>	3.97	Medium

The table above contains adjective-noun collocations extracted from Year 1 on the second set of data collection. In addition to this, the table indicates the degree of fixedness and accessibility of collocations based on the threshold scores. All collocations identified were also found in BNC.

A total of 1 Low-level collocation scored at 2.43, 2 medium level scored between 3.97- 5.72 on LogDice.

Noun-Noun Collocations

There was only one noun-noun collocation identified in BNC and it scored high on the level of fixedness and it scored at 8.70.

Table 25 Degree of Fixedness for Noun + Noun collocations identified from Year1 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Member of society	both need to focus on children and educate them about being a <i>good member of society</i>	8.70	High

b. Collocation Extracted from Year 2

Verb-Noun Collocations

Table 26 Degree of Fixedness for Verb + Noun collocations identified from Year2 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Become member	an important part of teaching the child how to <i>become a good member of society</i>	8.21	High
Meet people	place where he could <i>meet other people</i>	8.86	High
Go to school	people should teach their children before they <i>go to school</i>	7.42	High
Have child	when the parents decided to <i>have children</i> they must learn	6.11	Medium
Have parent	Not everyone <i>has parents</i> or role models	4.88	Medium
Have part	I think parents and school <i>have equal parts</i> in raising the children	3.98	Medium
Have responsibility	Parents <i>have more responsibility</i> than the school	5.10	Medium

Chapter 5

Help a child	most effective way to <i>help</i> their <i>children</i> to become better people	7.07	Medium
Make child	parents can <i>make</i> their <i>child</i> more aware	4.71	Medium
Play a role	Family <i>plays</i> a huge <i>role</i> in this teaching together with school	11.90	Advance
Raise child	If you are incapable for <i>raising</i> a <i>child</i> , don't bring once to the world	6.83	Medium
Spend time	<i>Spend</i> their <i>time</i> in learning something useful	8.93	High
Take care	Teach him how to <i>take care</i> of himself	10.08	High
Teach a child	parents should <i>teach</i> their <i>children</i> how to be respectful and honest	10.04	High
Teach kid	it is even more important to <i>teach</i> <i>kids</i> from a young age	5.61	Medium

The table above contains verb-noun collocations extracted from Year 2 on the second set of data collection. In addition to this, the table indicates the degree of fixedness and accessibility of collocations based on the threshold scores. All collocation identified were also found in BNC.

A total 7 medium level scored between 3.98- 7.07 and 6 High level scored between 7.42-10.08 and 1 advance scored at 11.90 on LogDice.

Adjective-Noun Collocations

Table 27 Degree of Fixedness for Adj + Noun collocations identified from Year2 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Big part	parents and the family have the <i>big part</i>	7.07	Medium

Big role	I personally believe that the <i>bigger role</i> of teaching should come from the parents	3.88	Medium
First year	I think in <i>first years</i> of children lives the parents who should teach them	8.80	High
Good moral	father raised his child on <i>good morals</i>	1.79	Low
Good member	To be a <i>good member</i> of society, parents have the main role	2.43	Low
Good person	She taught me how to be a <i>good person</i>	3.97	Medium
Good place	school is a good place to make children behaviour improve	1.43	Low
Good thing	parents should do <i>good things</i> in front of them	3.75	Medium
Good way	he will be responding in a good way with them	7.39	High

Based on the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. Generally, the majority of colloquial expressions (3 combinations) scored low on LogDice, ranging from 1.43-2.43 collocations showed a medium level of fixedness 3.75-7.07 collocations also scored highly on the level of was 7.39-8.80.

Noun-Noun Collocations

There was 4 noun-noun collocation identified in BNC and it scored high on the level of fixedness.

Table 28 Degree of Fixedness for Noun + Noun collocations identified from Year2 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Member of society	make them effective <i>members of society</i>	8.70	High

Chapter 5

Point of view	how you see things from your own <i>point of view</i>	11.11	Advanced
Role model	because they want to grow up to be like their <i>role model</i>	8.37	High
School age	when the child is still under <i>school age</i> their parents are everything in their life	6.07	Medium

A total of 1 medium level collocation scored at 6.07 , two high collocations scored between 8.37- 8.70 and 1 advance scoring at 11.11 on LogDice.

c. Collocation Extracted from in Year 3

Verb-Noun Collocations:

Table 29 Degree of Fixedness for Verb + Noun collocations identified from Year3 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Acquire information	at the school they <i>acquire</i> more <i>information</i>	8.40	High
Ask permission	he learn what is true and false, <i>asking permission</i> to enter	6.18	Medium
Change behaviour	School might <i>change</i> the child's <i>behaviour</i>	5.36	Medium
Follow instructions	teaching techniques and as a result they will not <i>follow</i> their <i>instruction</i>	7.86	High
Face challenge	children will <i>face</i> many <i>challenges</i> like how to talk people	6.03	Medium
Give a child	Parents try to teach or <i>give</i> their <i>children</i> the best education	6.75	Medium
Go to school	the first five years of his life before he <i>go</i> to <i>school</i>	7.42	High
Have attitude	even if they <i>have</i> bad <i>attitude</i> they can change over time	3.26	Low

Have child	Schools <i>have</i> many <i>children</i> so they cannot focus in each one	6.11	Medium
Have friend	children when get in the school they <i>have friends</i>	4.51	Medium
Have knowledge	some parents they don't <i>have knowledge</i> about education	4.95	Medium
Have responsibility	Society <i>have</i> a <i>responsibility</i> for those children to give them good education	5.10	Medium
Have role	I think that both <i>have a role</i> or a part in teaching children	5.26	Medium
Have time	doesn't <i>have</i> enough <i>time</i> to spend it with their children	7.12	High
Help a child	they must be work together to <i>help</i> the <i>child</i>	7.07	Medium
Learn something	children maybe <i>learn something</i> bad at school	9.14	High
Make friends	build relationships with others and <i>make new friends</i>	6.19	Medium
Meet people	make them socialize and open to <i>meet new people</i>	8.86	High
Play a role	Parents <i>play</i> a crucial <i>role</i> in how they affect their children behaviour	11.90	Advance
Raise child	the way they <i>raised</i> their <i>child</i> is what makes a child good member of society	6.83	Medium
Solve problem	teach them how to <i>solve</i> their <i>problems</i> , how to respect the older people	13.08	Advance
Take care	their children learn things like how to <i>take care</i> of themselves	10.08	High
Take responsibility	parents don't <i>take</i> the whole <i>responsibility</i>	7.59	High

Chapter 5

Teach child	I think the school <i>teach</i> the <i>children</i> more than parent	10.04	High
-------------	--	-------	------

The table above shows that collocations are obtained from Year 3 learners during the second round of data collection. Also shown in the table is the threshold score for fixedness and accessibility of collocations according to the fixedness criterion. A total of 1 Low collocations 3.26, 12 medium level collocation scored between 4.51- 7.07, 9 high collocations scored between 7.12-10.08 and 2 advance scoring at 11.90 -13.08 on LogDice.

Adjective-Noun Collocations

There was only one adjective-noun collocation identified in BNC and it was not acceptable.

Table 30 Degree of Fixedness for Adj + Noun collocations identified from Year3 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Bad behaviour	some of them have <i>bad behaviour</i> so they will affect others	5.86	Medium
Big responsibility	Being a parent is a big responsibility	3.73	Medium
Big problem	child will not recognize that it is bad which is the <i>big problem</i>	2.14	Low
Big role	Teachers also have a <i>big role</i> in raising children	3.88	Medium
Elderly people	how respect adults, and <i>elderly people</i>	12.55	Advance
First place	house is the <i>first place</i> children learn from it	1.83	Low
Good manner	teach their children <i>good manners</i>	5.72	Medium
Good member	being a <i>good member</i> means to be an effective person	2.43	Low
Good person	remind them that being a <i>good person</i> is important	3.97	Medium

Good place	school might be <i>good place</i> for children to learn them the good behaviour	1.43	Low
Good thing	they should learn the <i>good things</i> they need in life	3.75	Medium
Good way	connect with people so it's the <i>best way</i> to learn by experience	7.39	High
Great impact	They can make a <i>great impact</i> on their life.	5.45	Medium
Important role	school and society played very <i>important role</i> in children's life	2.58	Low
Important thing	Learning is the <i>most important</i> thing in person life	4.82	Medium
Long time	children the take <i>long time</i> in playing video games.	11.49	Advance
Main reason	People these days thinks that parents are the <i>main reasons</i> for children to be good members	10.19	High
Moral lesson	will make the children understand their <i>moral lessons</i> well	5.01	Medium
New thing	he go to school and learn <i>new things</i>	1.70	Low
New information	schools my can learn be use or take more <i>new information</i> like about	6.42	Medium
Negative energy	need to expel his <i>negative energy</i> , need to change the routine.	7.15	High
Old people	How to respect <i>older people</i> , and how to live like a good human	2.71	Low
Young age	If they teach them from a <i>young age</i> to respect elders, follow rules	6.58	Medium

Chapter 5

Many people	I mean by dealing with so <i>many people</i> with different ages and different in their level of education	7.38	High
Many things	I think the children acquire <i>many things</i> from the three first years	3.96	Medium
Real life	how to deal in <i>real life</i> or how should you behave	9.39	High

A total of 7 Low collocations scored 1.43-2.71, 12 medium level collocation scored between 3.75-6.58, 5 high collocations scored between 7.15-10.19 and 1 advance scoring at 12.55 on LogDice.

Noun-Noun Collocations

Table 31 Degree of Fixedness for Noun + Noun collocations identified from Year3 the second phase of data collection

Collocation	Citations from learners' Text	LogDice Score	Level of Fixedness
Member of society	respect old people and respect <i>members of society</i>	8.70	High
Education system	I could never and would never blame the <i>education system</i>	8.52	High
Job opportunities	because it will help to find more <i>job opportunities</i>	8.79	High
Job interview	It could help you impress at <i>job interviews</i> and show them that you have a skill that they could use	8.10	High
Point of view	In my <i>point of view</i> , both of them are important,	11.11	Advanced
Step by step	teach them a specific things <i>step by step</i> and continuously	9.20	High
School teacher	<i>School teachers</i> encounter different kinds of students	4.67	Medium

A total of 1 medium level collocation scored 4.67, 5 high collocations scored between 8.10-9.20 and 1 advance scoring at 11.11 on LogDice.

5.2.3 The Third Set of Data Collection

a. Collocation Extracted from Year 1

Verb-Noun Collocations

Table 32 Degree of Fixedness for Verb + Noun collocations identified from Year1 the third phase of data collection

Collocation	LogDice Score	Corpora Co-occurrences	Acceptable Collocation	Level of Fixedness
Give access	10.19	>5	yes	High
Get information	6.34	>5	yes	Medium
Learn skill	8.47	>5	yes	High
Need skill	4.21	>5	yes	Medium
Prepare student	8.51	>5	yes	High
Take a course	6.67	>5	yes	Medium
Teach student	8.66	>5	yes	High
Teach subject	5.96	>5	yes	Medium

Listed above in the table are verb-noun collocations extracted from Year 1 learners on the third set of data collection. According to the threshold score, the table also presents the fixedness and accessibility of collocation. A total of two collocations were not identified as acceptable collocations due to not meeting the criteria for acceptable collocation.

Based on the threshold, the remaining candidate combinations were identified as collocations and listed accordingly. On LogDice, 4 collocations expressions received medium scores, ranging from 4.21-6.67. There were also four collocations that received high scores of 8.47-10.19.

Adjective-Noun Collocations

There was only one adjective-noun collocation identified in BNC.

Table 33 Degree of Fixedness for Verb + Noun collocations identified from Year1 the third phase of data collection

Collocation	LogDice Score	Corpora Co-occurrences	Acceptable Collocation	Level of Fixedness
Important role	2.58	>5	Yes	low

Noun-Noun Collocations

Table 34 Degree of Fixedness for Noun + Noun collocations identified from Year1 the third phase of data collection

Collocation	LogDice Score	Corpora Co-occurrences	Acceptable Collocation	Level of Fixedness
Work environment	4.50	>5	yes	Medium

b. Collocation Extracted from Year 2

For the third set of data collection, students produced more collocations than second set.

Verb-Noun Collocations:

Table 35 Degree of Fixedness for Verb + Noun collocations identified from Year2 the third phase of data collection

Collocation	LogDice Score	Copora Co-occurrences	Acceptable Collocation	Level of Fixedness
Develop skill	9.97	>5	yes	High
Enter university	6.91	>5	yes	Medium

Expand knowledge	6.06	>5	yes	Medium
Find a job	7.41	>5	yes	High
Get a job	7.79	>5	yes	High
Get information	6.34	>5	yes	Medium
Give access	5.99	>5	yes	Medium
Have responsibility	11.38	>5	yes	advanced
Have experience	5.57	>5	yes	Medium
Have knowledge	4.95	>5	yes	Medium
Increase knowledge	5.78	>5	yes	Medium
Improve skill	8.31	>5	yes	High
Learn skill	8.47	>5	yes	High
Help student	6.26	>5	yes	Medium
Need skill	4.21	>5	yes	Medium
Provide support	8.72	>5	yes	High
Take a class	6.67	>5	yes	Medium
Take a course	6.67	>5	yes	Medium

A total of 11 medium level collocation scored between 4.21 -6.91, 6 high collocations scored between 7.41-9.97 and 1 advance scoring at 11.38 on LogDice.

Adjective-Noun Collocations

There was three adjective-noun collocation identified in BNC.

Chapter 5

Table 36 Degree of Fixedness for adj + Noun collocations identified from Year2 the third phase of data collection

Collocation	LogDice Score	Copora Co-occurrences	Acceptable Collocation	Level of Fixedness
General information	6.41	>5	Yes	Medium
Real life	9.39	>5	yes	High
Main Function	7.58	>5	Yes	high

A total of 1 medium level collocation scored at 6.41, 2 high collocations scored between 7.58-9.39 on LogDice.

Noun-Noun Collocations

Table 37 Degree of Fixedness for Noun + Noun collocations identified from Year2 the third phase of data collection

Collocation	LogDice Score	Copora Co-occurrences	Acceptable Collocation	Level of Fixedness
Work environment	4.50	>5	yes	Medium
Student life	8.46	>5	Yes	High
Work experience	8.74	>5	Yes	High

A total of 1 medium level collocation scored at 4.50, 2 high collocations scored between 8.46-8.74 on LogDice.

c. Collocation Extracted from Year 3

Verb-Noun Collocations

Table 38 Degree of Fixedness for Verb + Noun collocations identified from Year3 the third phase of data collection

Collocation	LogDice Score	Copora Co-occurrences	Acceptable Collocation	Level of Fixedness
Develop skill	9.97	>5	yes	High
give course	3.73	>5	yes	Medium
Graduate University	6.74	>5	yes	Medium
Have knowledge	4.95	>5	yes	Medium
Help student	6.26	>5	yes	Medium
Improve skill	8.31	>5	yes	High
Manage time	6.22	>5	yes	Medium
Need skill	4.21	>5	yes	Medium
Prepare student	8.51	>5	yes	High
Take a course	6.67	>5	yes	Medium

Listed above in the table are verb-noun collocations extracted from Year 3 learners on the third set of data collection. According to the threshold score, the table also presents the fixedness and accessibility of collocation.

A total of 7 medium level collocation scored at 3.73- 6.74, and 3 high collocations scored between 8.31-9.97 on LogDice.

Adjective-Noun Collocations

There was only one adjective-noun collocation identified in BNC.

Table 39 Degree of Fixedness for Adj + Noun collocations identified from Year3 the third phase of data collection

Collocation	LogDice Score	Copora Co-occurrences	Acceptable Collocation	Level of Fixedness
-------------	---------------	-----------------------	------------------------	--------------------

Chapter 5

Main Function	7.58	>5	Yes	High
Creative person	7.83	>5	Yes	High
Good place	1.43	>5	yes	Low

A total of 1 Low level collocation scored at 1.43, 2 high collocations scored between 7.58-7.83 on LogDice.

Noun-Noun Collocations

Table 40 Degree of Fixedness for Noun + Noun collocations identified from Year3 the third phase of data collection

Collocation	LogDice Score	Copora Co-occurrences	Acceptable Collocation	Level of Fixedness
Access to knowledge	6.07	>5	Yes	Medium
Computer skill	5.43	>5	Yes	Medium
Work life	2.99	>5	Yes	Low
Point of view	11.11	>5	Yes	Advanced

A total of 1 Low 2.99, 2 medium level collocation scored at 5.43-6.07, 1 advanced collocation scored at 11.11 on LogDice.

5.3 Collocation progression

After looking into collocations count and fixedness. It is important to look into learners overall performance and compare their performance through the academic year and across years of study.

Level of Fixedness for year 1

Table 40 showed the distribution for the four levels of fixedness (advance, high, low and medium) for each phase and collocations. For the phase one, it was noted for adjective + noun

collocations that most students showed low level (n=11, 47.8%), then followed by high (n=6, 26.1%) and medium (n=6, 26.1%). For verb+noun collocations, the students tended to have medium level (n=12, 52.2%). For the phase two, less than half of students had high level (n=3, 42.9%) in verb+noun. For the phase three, students tended have either high level (n=4, 50%) or medium level (n=4,50%) in verb + noun collocations.

Table 41 Distribution for levels of fixedness (advance, high, low and medium) for each phase and collocations in year 1

a	phase	Collocations		Level of Fixedness							
				Advance		High		Low		Medium	
				N	%	N	%	N	%	N	%
	1	Collocations	Adjective + noun Collocations	0	0.0%	6	26.1%	11	47.8%	6	26.1%
			Noun + noun collocation	1	50.0%	1	50.0%	0	0.0%	0	0.0%
			Verb + noun collocations	1	4.3%	7	30.4%	3	13.0%	12	52.2%
	2	Collocations	adjective + noun collocations	0	0.0%	0	0.0%	1	33.3%	2	66.7%
			noun + noun collocations	0	0.0%	1	100.0%	0	0.0%	0	0.0%
			Verb + noun collocations	0	0.0%	3	42.9%	2	28.6%	2	28.6%
	3	Collocations	adjective + noun collocations	0	0.0%	0	0.0%	1	100.0%	0	0.0%
			noun + noun collocations	0	0.0%	0	0.0%	0	0.0%	1	100.0%
			Verb + noun collocations	0	0.0%	4	50.0%	0	0.0%	4	50.0%

a. year = 1

Level of Fixedness for year 2

Table 41 presented the distribution for the four levels of fixedness (advance, high, low and medium) for each phase and collocations in year 2. For the phase one, it was noted for adjective + noun collocations that most students showed medium level (n=15, 46.7%), then low (n=10,

Chapter 5

33.3%) and high (n=5, 16.7%). In terms of verb+noun collocations, the students seemed to show medium level (n=13, 56.6%). For the phase two, the levels of student were equally distributed between low, medium and high (n=3, 33.3%) in noun + noun collocation. For the phase three, it was found that the verb + noun collocations were the most common used, and students tended have medium level (n=11, 61.1%) in verb + noun collocations.

Table 42 Distribution for levels of fixedness (advance, high, low and medium) for each phase and collocations in year 2

a	phase	1	Collocations	Adjective + noun Collocations	Level of Fixedness (year 2)							
					Advance		High		Low		Medium	
					N	%	N	%	N	%	N	%
					1	3.3%	5	16.7%	10	33.3%	14	46.7%
					1	33.3%	2	66.7%	0	0.0%	0	0.0%
					0	0.0%	6	26.1%	4	17.4%	13	56.5%
		2	Collocations	adjective + noun collocations	0	0.0%	3	33.3%	3	33.3%	3	33.3%
					1	25.0%	2	50.0%	0	0.0%	1	25.0%
					1	6.7%	6	40.0%	0	0.0%	8	53.3%
		3	Collocations	adjective + noun collocations	0	0.0%	2	66.7%	0	0.0%	1	33.3%
					0	0.0%	2	66.7%	0	0.0%	1	33.3%
					1	5.6%	6	33.3%	0	0.0%	11	61.1%

a. year = 2

Level of Fixedness for year 3

The distribution for the four levels of fixedness (advance, high, low and medium) for each phase and collocations in year 3 was presented in Table 42. For the phase one, most students showed medium level (n=9, 40.9%), then low level (n=8, 36.4%) in adjective + noun collocations that. With respect to verb+noun collocations, the students tended to have medium level (n=8, 47.1%), followed by high level (n=6, 35.3%). For the phase two, the medium levels (n=12, 46.2%) in

adjective + noun collocations was the highest. Also, in the phase three, it was found students tended have medium level (n=7, 70%) in verb + noun collocations.

Table 43 Distribution for levels of fixedness (advance, high, low and medium) for each phase and collocations in year 3

			Level of Fixedness (year 3)							
			Advance		High		Low		Medium	
			N	%	N	%	N	%	N	%
phase1	Collocations	Adjective + noun Collocations	1	4.5%	4	18.2%	8	36.4%	9	40.9%
		Noun + noun collocation	1	100.0 %	0	0.0%	0	0.0%	0	0.0%
		Verb + noun collocations	0	0.0%	6	35.3%	3	17.6%	8	47.1%
2	Collocations	adjective + noun collocations	2	7.7%	5	19.2%	7	26.9%	12	46.2%
		noun + noun collocations	1	14.3%	5	71.4%	0	0.0%	1	14.3%
		Verb + noun collocations	2	8.3%	9	37.5%	1	4.2%	12	50.0%
3	Collocations	adjective + noun collocations	0	0.0%	2	66.7%	1	33.3%	0	0.0%
		noun + noun collocations	1	25.0%	0	0.0%	1	25.0%	2	50.0%
		Verb + noun collocations	0	0.0%	3	30.0%	0	0.0%	7	70.0%

a. year = 3

5.4 Difference in levels of fixedness between phases and years

For each phase, I looked overall level of fixedness based on using the average Log Dice score and then to classify collocations into low: 0.1-3.5, medium: 3.6-7, high:7.1-10.5 and advance: 10.6-14. I investigated whether there was not any significant difference between the three phases for each year. The one-way ANOVA test was used as the Log Dice score was normally distributed

Chapter 5

using the Shapiro test. The test of homogeneity showed that the phase had the same variance using the Levene's test. Figure 1 showed that there variation in Log Dice scores among the students, and it was noted that the scores were ranging from low to high.

For year one, it was observed that there phase tended to have slightly different mean Log Dice score (phase 1=5.68, phase 2=6.19 and phase 3=6.61) but all these means indicted that the level of fixedness was medium, Table 43. Although the average score increased as the phase increased, the ANOVA test showed no significant differences between the three phase (F=.603, p-value=.550).

Table 44 Mean Difference in levels of fixedness between phases for each year using one-way ANOVA

year	phase	N	Mean	SD	ANOVA
					F-test (p-value)
1	1	48	5.68 (medium)	2.67	.603 (.550)
	2	11	6.19 (medium)	2.66	
	3	10	6.61 (medium)	2.38	
2	1	56	5.55 (medium)	2.37	4.44(.014)
	2	28	6.49 (medium)	2.73	
	3	24	7.19 (high)	1.76	
3	1	40	5.78 (medium)	2.64	.950(.390)
	2	57	6.55 (medium)	2.85	
	3	17	6.35 (medium)	2.48	

For year 2, I noticed the average score Log Dice increase when students moved to higher phase (phase 1=5.55, phase 2=6.49, and phase 3=7.19) where phase 1 & 2 showed medium (mean: phase 1=5.55, phase 2=6.49) level while phase three had high level (mean=7.19), Table 4. The difference between the phases was statistically significant using the ANOVA (F=4.44, p-value<.014).

For year 3, the phase one had lower score (mean=5.78) than phase 2 (mean=6.55) and 3 (mean=6.35), noticed that the average scores were in medium level. The ANOVA test showed no significant difference between the three phases ($F=0.950$, $p\text{-value}=.390$).

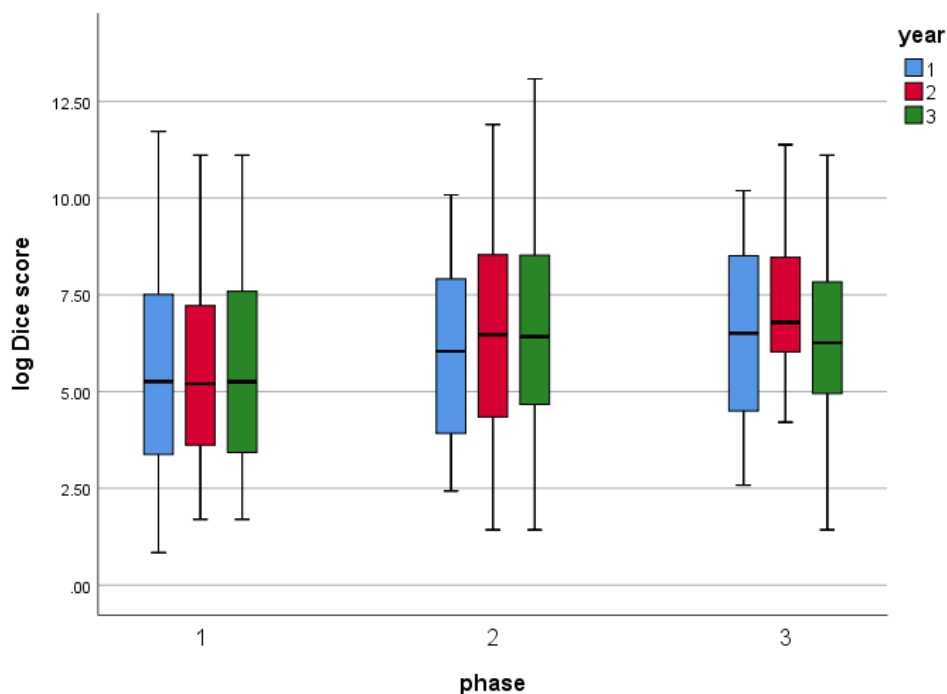


Figure 17 Distribution of Log Dice between phase and years.

Since it may be expected that the level of fixedness would be higher for higher years of study, we compared the score Log Dice between the three years, Figure 17. The average score seemed to be lower in year 1 (mean=5.89) compared to year 2 (mean=6.16) and 3 (mean=6.25), notice that all averages lied in medium level of fixedness, Table 44. The ANOVA table showed that no significant difference between the three years ($F=0.421$, $p\text{-value}=.627$)

Table 45 Mean Difference in levels of fixedness between the three years using one-way ANOVA

				ANOVA
year	N	Mean	SD	F-test (p-value)
1	69	5.89 (medium)	2.62	.421(.657)
2	108	6.16 (medium)	2.43	

3	114	6.25	2.72
		(medium)	

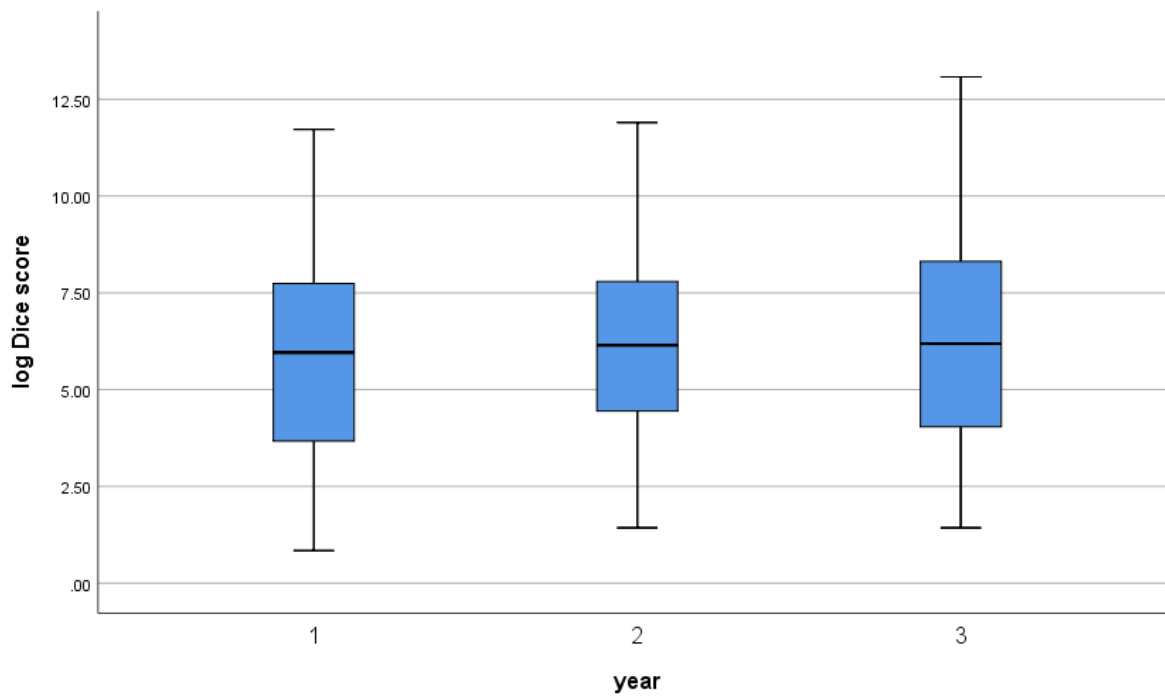


Figure 18 Distribution of Log Dice between the three years

Chapter 6 Discussion

There is an increased interest in using corpora to inform and enhance language teaching and assessment (Weigle and Goodwin, 2016). However, most of the research that has been done only quantitatively measured the performance of learners at one point in time and this does not answer questions related to language development over time (Belz, 2007, Staples *et al.*, 2013). Many researchers have urged to increase investigations on employing corpora in language testing and assessment especially to explore individual differences between different levels of language proficiency (Park, 2014).

Thus, this study aims to investigate second language learners use of collocations in essay writing. The goal is to examine second language learners' collocation production and development at different proficiency levels and stages in their learning over one academic year through corpus-based research. In order to do that, 807 written samples were considered in this study that were collected from 269 students. Students were asked to write 3 different essays on three different points of the academic year. The first sample was taken at the beginning of the year. The second sample was taken at the end of the first semester (after three months after the first sample). The third sample was taken at the end of the academic year. All samples were transcribed and uploaded to an online corpus (Sketch Engine) in order to do analysis. To the best of my knowledge, there is no previous study that has investigated learners' collocation, collocations production has never been done in such terms in Saudi Arabia.

Throughout this chapter, the results of the research are discussed in relation to three major themes derived from the three research questions that motivated the research in the first place:

- What type of collocations is frequently used among learners of three different years of study?
- What the degree of fixedness for identified collocations found in learners' learners writing at each year of study and across the academic year?
- What progress have students made in their collocation production during their studies? does the learners' first language (L1) factor into their production of collocations in English?

6.1 Proficiency level and Collocations Knowledge

It can be seen from the results of question one, which investigated how learners used collocations in their writing, that verb-noun collocations are the most frequently used collocations in student writing, with 147 examples representing 50% of the total number of collocations identified. These

Chapter 6

results are summarized on page 26 and figures 10 and 11 are on pages 26 and 27. In addition to these results, it was found that 121 adjectival-noun collocations were discovered, accounting for 41.2% of all collocations at the same time. A total of 119 instances of adjectival-noun collocations were recorded. A total of 26 examples of noun-noun collocations were found in student writing at the same time, which accounts for only 8.8% of all collocations in student writing.

Results demonstrated in table 5 page 26 and figures 10, 11 page 26-27 show that learners at all levels can consistently produce verb-noun, adjective-noun, and noun-adjective collocations. However, the study results showed Verb- noun collocations were the most frequently used collocations both in Year 1 (54.3%) and Year 2 (52.7%). Learners in year 3 produced (44.7%) of overall collocations were verb-noun collocation which means their collocations production is more diverse than the other years and they did not depend mostly on verb + noun collocations like the other years. A particular type of collocation may be easier to acquire than another, as in the case of grammatical collocations, which are easier to acquire than lexical collocations (O'Sullivan & Chambers 2006). The use of verbs in pattern formation has been found to be easier for learners than using adjectives (Todd, 2010); both (Alangari, 2019) and (Khoja, 2019) focus on the type of collocation and the year of study, and both study that verb-noun collocations are easier to acquire and produce for ESL learners.

The results in table 5 also show learners can produce adjectives with nouns, and the numbers do not differ much from verb-noun collocations. In Year 1 (40%) of overall collocation are adjective + noun collocation. In Year 2 Learners adjective + noun collocations came to (38.2%) of overall collocations. Learners in year 3 produced (44.7%) of overall collocations are adjective- noun collocations which equals the percentage of verb-noun collocations. The results indicate that as learners progress in their language studies, the variety of collocations increases and the dependency on verb-noun collocations decreases.

Compared to verb+noun collocations, adjective+noun collocations posed significant difficulty for learners when they tried to find collations through corpus consultation. In their study on adjectival-noun collocations in essays written by non-native speakers, Siyanova and Schmitt (2008) found that non-native speakers produced the same number of adjectival-noun collocations as native speakers, 25% of the adjective+noun collocations from L2 learners' essays are atypical, demonstrating that learners have difficulty producing adjective+noun collocations. According to a study by Chen (2002), Taiwanese students are more likely to make errors when using verbs and adjectives in conjunction with nouns in their L2 essays. This could be due to a number of factors, including: L1 interference, lack of L2 exposure and focus on grammar.

The findings by Alangari et al. (2019) indicate that when comparing ESL learners' use of both verb-noun and adjective-noun lexical collocations, the variations in structures impact on learners knowledge development in acquisition and use of collocations in either spoken or written language. As opposed to Chen and Baker (2014) who compared the collocational features and language proficiency, Alangari et al (2019) examined the impact of collocational features in relation to the geographic location of the ESL learners. The study found that compared to Saudi ESL learners, students in the United States, there was a significant variation in terms of collocational knowledge regardless of the collocations structure. The study further noted that among the Saudi students, there was an increased used of phrasal verb clause structure compared to the non-native speakers (Alangari et al. 2014).

When it came to noun-noun collocations they are the least type of collocations learners produced. In Year 1 only (5.7%) of overall collocation are noun + noun collocation. In Year 2 Learners noun + noun collocations came to (9.1%) of overall collocations. Learners in year 3 produced (10.5%) did better than year in noun- noun collocations.

There has been a critical examination of the correlation between learners' proficiency and their ability to use collocations in previous research. Among the earlier studies that explored this theme are Lateh *et al.* (2021); Granger and Bestgen (2014); Kamarudi *et al.* (2020); Chen *et al.* (2021); Wolter and Yamashita (2017); Ramadhan (2017), Alangari (2019) and Khoja (2019). According to Ramadhan (2017), the various features of collocations usually affects knowledge and development of second language. The study found that gender was an important factor for collocations but does not influence knowledge development. Additionally, the study indicated that both receptive and productive collocational knowledge among the ESL learners across genders did not increase at the same rate. The findings are consistent with those of Chen and Baker (2014) which found that the discourse features used in the process of learning a second language impacts on the specific features of collocations and consequently knowledge development among the ESL learners. However, unlike Ramadhan (2017), Chen and Baker (2014) argue that regardless of collocational features, proficiency levels also impacts on knowledge development and use of collocations among the ESL learners whereby the speakers of lower proficiency tend to shared more common features. As a result, the study hypothesizes that it is possible for proficiency as well as the collocational features to inform on collocation knowledge among second language learners.

Khoja's (2019) study did not find any significant difference in collocations use among the pre-intermediate and intermediate ESL learners argued by Granger and Bestgen (2014) and Lateh et al. (2021). However, the study found that Saudi ESL learners overall produced a high number of

Chapter 6

acceptable collocations in written second language. Lastly, unlike Khoja (2019) who did not find any correlation between proficiency and collocations knowledge, the findings by Chen et al. (2021) indicated that among the ESL learners, as their proficiency levels of the language increases, other exclusivity features including their knowledge of collocations in spoken and written language also increases.

Based on my research findings, students are able to produce acceptable verb-noun, adjective-noun, and noun collocations in written texts on three levels of fixedness, despite the fact that collocations are not explicitly taught in English lessons. This is likely due to the fact that students are able to combine what they have learned and acquired in the classroom with the skills they have acquired outside of the classroom. By combining these two sources of knowledge, students are able to develop the ability to produce collocations independently. Chang, Y. (2018); Pellicer-Sánchez, A. (2017); Vilkaitė, R. (2017); Webb, S., Newton, J., & Chang, A. (2013), studies support the claim that students can acquire collocations independently, even if they are not explicitly taught in English lessons. According to these studies, students are able to learn collocations through exposure to language, both inside and outside the classroom. As students combine their knowledge and skills from various sources, they are able to develop the ability to produce collocations independently.

Lu and Dienes (2020) explore the differences between conscious and unconscious strategies of collocational knowledge acquisition. The study findings indicate that unconscious structural knowledge focuses on the acquisition of statistical irregularities. Nevertheless, the study suggests that ESL learners can acquire collocational knowledge both consciously and unconsciously. Regardless of the method of collocational knowledge acquisition, the level of proficiency was found to be a key variable that affects the process irrespective of whether the collocations are actively taught or learnt subconsciously (Lu and Dienes 2020). The findings are not consistent with the argument presented by Hsu (2002) who examined whether conscious acquisition of collocations through actively teaching ESL learners lexical collocations was possible. In general, Hsu (2002) found that actively teaching lexical collocations to ESL learners was integral to improvement of collocational knowledge both in spoken and written discourses. However, consciously acquiring collocational knowledge was further found to have a positive impact on proficiency. The study implies that when teaching a second language, collocational knowledge should also be taught as this enhanced proficiency and general language use among the ESL learners. The argument is also supported by Gries (2018) noting the need for alternative strategies for teaching languages. The study suggests that the current methods commonly used in teaching languages fail to incorporate aspects of extra-linguistic including collocations knowledge and proper use in both spoken and written discourses.

According to Alsakran (2011), environment is an integral factor to teaching second language and collocational use among ESL learners. The study noted that a comparison of Arabic ESL learners in Saudi and the United States showed significant variation in terms of the optimal strategies for teaching collocations and overall outcomes. Unlike Lu and Dienes (2020) and Hsu (2002), L2 proficiency among the learners was not a factor of consideration in this case as the study only focused on proficient second language users. The study argues that Arabic students learning second language in California had a better acquisition of the three types of collocations including the verb-preposition, verb-noun and adjective-noun collocations both in their spoken and written second language. On the other hand, Alsakran (2011) argued that Arabic students studying L2 locally had a poor grasp of the three types of the collocations due to the lack of exposure to an environment whereby the language is not used regularly. In addition to Alsakran's (2011) hypothesis on the role of environmental exposure.

The findings by Wolter and Yamashita (2017) compared proficiency and knowledge of collocations use between native and non-native speakers and found that compared to native speakers of English language, the non-native speakers had the ability to process congruent collocations much faster. However, they may not be as accurate or fluent as native speakers. Non-native speakers may not be as familiar with these collocations, and they may make mistakes when using them. For example, they might say "to sprint a marathon" instead of "to run a marathon". This is because they are not aware of the fact that the verb "to sprint" is only used to describe short, fast races. However, with practice, non-native speakers can learn to process congruent collocations correctly. There are a number of resources available to help them with this, such as dictionaries, collocation lists, and online exercises.

Additionally, the study suggested that the congruency in findings could also be explained by order of acquisition and learners' age. Learners who are exposed to congruent collocations early on in their language learning process are more likely to process them correctly than learners who are exposed to them later. This is because the early exposure gives the learners a chance to learn the collocations as a unit, rather than as two separate words. Also, younger learners are more likely to process congruent collocations correctly than older learners. This is because younger learners' brains are more plastic, which means that they are better able to learn new patterns. The findings are consistent with those of both Lateh et al. (2021) and Granger and Bestgen (2014). However, Wolter and Yamashita (2017) further argue against the popular belief that processing formulaic sequences among non-native speakers or less proficient speakers of a language differs with that of more proficient or native speakers. Nevertheless, the study shows that frequency, congruency and proficiency in general have impact on collocational processing. The findings are consistent

Chapter 6

with those of Kamarudi et al. (2020) noting that the ESL learners' knowledge of collocations increased as their language proficiency increased.

6.2 Collocation's fixedness

The term collocation fixedness describes a decomposition phrase or construction that frequently merges its components. Among the characteristics of fixed collocations is that the constituents occur repeatedly together. Fixedness of collocations can be identified as per native-speaker corpus such as the British National Corpus (BNC) and Corpus of Contemporary American English (COCA), which helps in examining the production of distinct types of collocations by learners. Sadeghi (2010) argued in order to learn L2, learners need to be knowledgeable about different types of collocations, which are fixed or restricted in different ways in different situations. By conversing with a wide range of conventional fixed collocations, L2 learners expand their language proficiency, thereby showing evidence of language acquisition.

Table 41 in page 113 summarizes year 1 learners collocations production that was detailed in tables (14-16) pages (87-90), tables (23-25) in pages (97-99) and in tables (32-34) pages (107-108) respectively in terms of the fixedness.

For adjective + noun collocations learners in phase one mostly used low fixed (n=11, 47.8%), which was followed by high (n=6, 26.1%) and medium (n=6, 26.1%) levels of fixedness.

In terms of verb-noun collocations, students tended to have a medium level of proficiency (n=12, 52.2%). The results of phase two revealed learners had a high level of fixedness (n=3, 42.9%). For the phase three, students tended to have either high level (n=4, 50%) or medium level (n=4, 50%) of verb + noun collocations. This showed an increase of the quality of collocation produced as they advanced in their studies.

For noun- noun collocation as the table shows there was no significant change in the number or degree of fixedness of the collocation produced.

For year 2 table 42 in page 114 summarizes the collocations production of the year 2 learners that was detailed in tables (17-19) pages (90- 93), and in tables (26-28) pages (99- 102), and in tables (35-37) pages (108-110) respectively according to the fixedness of the collocations.

According to the results from phase one of this study, the majority of students showed a medium level of ability for adjective + noun collocations students collocations of medium and high level of fixedness had increase from the first and second phases where it scored (33.3%) for medium level

collocations and (66.7%) for high level collocations. This result is similar to with Khojas(2019)who found most of pre-intermediate adjective- noun collocations had a medium level of fixedness.

It for verb-noun collocations were the most common used, and students tended have medium level (n=11, 61.1%) in verb + noun collocations. This result also agree with Khojas(2019) results where she found most pre-intermediate learners used collocations of medium level of fixedness

There were no significant differences in student levels between low, medium, and high in the second phase of this study when it came to noun + noun collocations most of the collocations produced had a high level of fixedness. This also agrees with khoja(2019) in that most of these noun-noun type collocations were identified as being high level collocation.

A summary of the collocation production of Year 3 learners is given in table 43 page(115) along with tables (20-22) pages (94-97), tables (29-31) pages (102-106), and tables (38-40) pages (111-112) respectively, which correspond to the degree of collocation fixation of the learners.

For adjective – noun collocations most collocations had medium level of fixedness in the first phase (40.9%) and second phase (46.2%) in the third phase most of the collocation had a high level of fixedness (66.7%). Also, Khoja (2019) found most of students collocation are medium level of fixedness. However and since this study looked even further into learners production of adjective -noun collocations has improved.

Noun -noun produced by year 3 learners had mostly medium Level of fixedness this agrees with Khoja(2019) who found most of noun-noun collocation at the intermediate level to be of medium degree.

The majority of Verb- noun scored at medium thus the number of fixed collocation has improved throughout the dear and there was no registered low fixed collocation disappear. These findings agree with khojas 2019 who found intermediate learners mostly produced medium level collocations

Phase and year differences in fixedness levels

Table 44 page 116 that ANOVA results demonstrate Year one showed slightly different mean Log Dice scores (phase 1=5.68, phase 2=6.19, and phase 3=6.61), but all means indicated that the level of fixedness was medium, Table 44. ANOVA revealed no significant differences between the three phases, despite the average score increasing as phase increased (F=.603, p=.550).

For Year 2, I observed that the average log dice score increased as students moved up to higher phases (phase 1 was 5.55, phase 2 was 6.49, and phase 3 was 7.19), as shown in page 116 Table

Chapter 6

44. Phase 1 & 2 had medium levels (mean 5.55, phase 2 was 6.49), while phase 3 had high levels (mean=7.19). The difference between the phases was statistically significant using the ANOVA ($F=4.44$, $p\text{-value}<.014$).

Regarding Year 3, phase one had a lower average score (mean=5.78) than phases 2 (mean=6.55) and 3 (mean=6.35), indicating a medium level of achievement. There was no significant difference between the three phases based on the ANOVA test ($F=0.950$, $p\text{-value}=.391$).

According to Lewis (1997), the ability to use language creatively and innovatively increases with the acquisition of fixed or semi-fixed phrases. This is because these phrases are often used together in natural-sounding language, and they can help L2 learners to achieve fluency and naturalness in their speech and writing. There is some evidence to support the idea that collocation fixedness plays a positive role in language development according to the results of the study by Mirsalari (2019). Mirsalari (2019) found some evidence to support this idea. The authors used Pearson correlation and paired samples t-tests to analyze data on the development patterns of lexical collocations among Iranian learners. They found that there was a positive correlation between collocation fixedness and language development. In other words, learners who were more familiar with fixed collocations were also more likely to be fluent and natural in their language use. Mirsalari (2019) concluded, collocations are beneficial not only in oral production, but also in written production among L2 learners as they assist them in avoiding the use of improper phrases or prolonged sentences when they are expressing their ideas in the written form. Clearly, fixed or semi-fixed collocations assist in language acquisition, and therefore should be included in L2 language instruction.

The findings on the positive influence of collocations on proficiency are consistent with the argument by Mirsalari (2019) that having a store of collocation that are fixed or semi-fixed helps L2 learners to improve their language proficiency. Similarly, Siyanova-Chanturia (2015) is in support of the statement, further arguing that unlike advanced learners, beginners or those learners with no previous L2 knowledge, have much smaller stores of single words or even phrases, and as such, they lack proficiency in the language. Beginner learners make use of noun+ adjective combinations as they increase their proficiency, with the increase being reflected in both quality and quantity of the resulting combinations (of fixed and semi-fixed collocations), for instance, "*tanned skin*". The researchers suggest that about five months might be adequate for novice learners to *gradually begin* to come up with or construct more idiomatic output that is almost similar to that of native speakers (Siyanova-Chanturia, 2015). Other scholars such as Laufer and Waldman (2011) and Huat (2012) have also explored the link, but go a step further by investigating the formation of verb-noun collocations among the three proficiency levels.

However, Huat (2012) had crucial findings as they related to qualitative development of these collocations by Malaysian EFL learners. After analysis of the beginner, intermediate and upper intermediate (advanced) levels, Huat (2012) concluded that there a sequence of development with regard to how the learners produced the verb-noun collocations, with the two groups in the higher levels producing collocations that were the same in pattern and frequency. In particular, beginners used verbs with non-specific meanings, i.e., *pick some flowers* while intermediate and advanced level learners produced collocations with particularly specific meaning, for instance, *pluck some flowers* (Huat, 2012). Therefore, it can be concluded that beginners used general collocations whereas advanced learners could use fixed collocations, and this shows the development in proficiency among the levels.

6.3 First Language Transfer

Table 16 of the findings in page () showed the degree of fixedness of collocations extracted from Year 2 on the first set of data collocation. We can see the influence of L1.

There is one notable collocation that shows the influence of L1 on L2 production and that is “Love Language”. In the English language, the collocation “Love Language” refers to the way in which a partner prefers to express and receive love from one another. There are various ways in which we can exhibit love, including acts of service, gifts, and words of affirmation. Today, love is widely used to refer to all of these ways and is used to describe all of them. An example found in COCA “So take my wife, as example. Herher primary **love language** is all about acts of service. My primary love language is all about words.” However, In their writings, students used the collocation “love language” to express how they feel about learning a language. For example, one student wrote, “Also many people learning the language because they **love this language** or loved person speak this language” Learners here used a literate Arabic translation because in Arabic the word "حب" “Houb” can be used to express how someone feels positively feels about something. Thus “like” and “love” are commonly expressed using the same word in Arabic. Lacking this knowledge and as learners are still learning and relaying on Arabic to borrow and translate here the collocation “love language has been used differently as a result of L1 influence.

Year 2 first set

Another instance In learners’ text the students used combinations “good job” was used to express the positive impact of language learning. In one sentence a student wrote: “*If you want to have a good job you must speak two languages*”, another student wrote “*I hope I can graduate with high skills and work in a good job that it fits me*”. In both examples and the rest of occurrences *good job* was used to express how learning language can influence the job opportunities one can have.

Chapter 6

Although the meaning found in BNC is similar to those of learners, the sentence constructions were different. For example: *"My job is very good for meeting people"* in another: *"Maidstone's offer of a full-time job wasn't good enough to persuade"* in all examples the meaning refers to grade the condition of the job they work in. However, the combination in BNC did not appear close to each other and adjective *good* did not proceed noun *job* as found in learners writing. One explanation is because the combination *good job* is mostly used as compliment to a job that was done well. The other combination *good way* was used in learners writing to express great methodology of doing something. For example, one student wrote: *"In my opinion, reading could be a good way to learn foreign language."* Another student also wrote: *"In my opinion, acquire new language is one of the best way to make yourself more confidence."* In the BNC however the similarity of meaning only appeared one time and it was in a spoken context: *"I think as a love story and as a novel of academic life it's in many ways very good indeed."* One explanation for student overuse of this combination is because of L1 transfer. In Arabic language, the collocation "طريقة جيدة" pronounced as "tariqah jayida" is an equivalent translation of the combination "good way" and is used commonly in Arabic to express ones opinion of how one technique is valuable.

*"Mother Language". Learners used this collocation when they want to talk about their native language. One learner writes "You have to know a new language but you have not to forget your **mother language** and used it in your daily life."* One explanation for the use of *"mother language"* instead of more native like collocations such as, *"mother tongue"*, *"native language"* or *"first language"* is that is a literal translation of Arabic collocation اللغة الأم that literary translate to *"mother language"*. It is the common collocation used in Arabic to talk about one's native language. Interestingly, one student used both collocation in their writing: *"In my opinion because in now times learning other or different languages from your mother tongue is necessary. For example, to work in a company you must have 2 different languages beside your mother language."*

Year 3 first set

In some of the students writings we can see the influence of L1 in the use of the collocation *"have language"*. In colloquial Arabic it is normal to use the combination *"have language"* to express knowing or speaking a language. It can be seen in some of the learners writings. For example *"You must have another foreign language because that helped you in travelling"*. Another student wrote *"when you have one language is not enough because sometime can't deal with another people"* In both examples the students are highlighting the importance on knowing or speaking a foreign language and because they translate it from Arabic they misused the collocation.

Studies of collocational performances have unearthed that L2 students prefer to use the available choice principle than the restricted one, using fewer collocations than their English counterparts. Besides the inefficient use, L2 learners tend to underuse or overuse some collocations. A study on the verb-noun collocations generated by both native and L2 learners of English discovered a few numbers of verb-noun collocations generated by the non-native English speakers compared to the original English speaker counterparts (Men, 2015). Similarly, other investigations found that L2 learners notably used a few intensifying adverbs that end with L1 in terms of tokens and types. Regarding the use of lexical forms, non-native speakers also recorded a lower use than the original speaker peers, whose compositions demonstrated otherwise.

The failure to use broader native-sequel expressions frequently leads to the lack of diversity in composition, making the non-native writing sound odd or foreign (Men, 2015). The lack of diversified collocation use is also featured by the underuse and overuse of some collocations. In the spoken generation of routines, sequence, and formulaic by L2 learners, studies found highly notable underuse of certain vagueness and the overuse of some. Regarding the writing performance, it was noted that L2 students generated a much lower proportion of idiomatic phrasal verbs than their native peers. They also tend to underuse more idiomatic collocations and overuse open collocations when using adjective-intensifier combinations in their writings. The forms of over utilised collocations are always associated with the linguistic combinations in the students' native dialect.

The small tokens of precast that advanced learners have in their reserves and the degree to which they are utilised and over utilised. The main reason for the underuse and overuse situation that arises in Non-native speakers' collocation is that students often hold on to fixed expressions and phrases that they feel confident in using. The fixed expressions and phrases tend to become their safe bets or islands of reliability. Thus, their heavy reliance on familiar collocations results in the overuse and avoidance of the ones they are not sure to use results to underuse. These behaviours of L2 students are hardly surprising because underuse and overuse of collocation are unavoidable situations in the interlanguage development process, as is also the case with the use of lexis and grammatical structures.

The collocation balance between native speakers and L2 learners will eventually show low diversified uses with the latter because achieving native-like competency obviously cannot meet the same level as native speakers (Men, 2015). Which is the main criticism faced by contrastive interlanguage analysis, to the level that there is a tendency of oversimplifying generalisation of students' underuse and overuse when their dialect is directly compared with native peers. To put it in other ways, underuse and overuse is hardly a particular issue of collocation. What is more

Chapter 6

crucial in non-native collocation research is to study the forms of misuse and discover the fundamental challenges encountered with collocation education.

Past studies on non-native collocation; (Nesselhauf,2003); (Laufer and Waldman, 2011) noted a considerable proportion of unfitting collocation use. An investigation on the verb-noun collocation in a collection of compositions by advanced L2 learners of English discovered that almost a third of the collocation were misused. It was concluded that these students had notable challenges in choosing the right verbs in verb-noun collocations. Even though non-native speakers generated a considerable percentage of collocation mistakes, previous investigations suggested that not all types of collocation pose similar challenges to L2 students. They faced significant problems in generating verb-noun collocations compared to other types of collocations such as adjective+ noun. They tend to acquire adjective-noun collocation easily and earlier, while the verb-noun collocations are hard and acquired later.

However, as they continue to master and advance their L2 understanding, they also improve their use of collocation. L2 learners increased their accuracy on adjective-noun collocations as they became more and more competent in their second language. This is confirmed by the massive gap between advanced learners who have better control of adjective-noun collocations and beginners who have much less command of the same. Indicating that as they continue to develop their competency, their adjective-noun collocation understanding advances. Despite their better performance on adjective-noun collocations, the phenomena generally pose unique learning challenges beyond doubts even for competent L2 students. Studies have concluded that the learning challenges face by this group include the arbitrary limitations in word combinations.

It has been proved; (Nesselhauf, 2003); (Koç, 2006); (Li and Schmitt, 2010), that the combinations of collocations with the average extent of limitation are more prone to mistakes compared to more limited combinations among L2 students. More so, they tend to learn the more restricted collocations as a whole while they use the less restricted ones smartly. The relative individual collocation infrequency input has been noted as a severe challenge to L2 Collocation learning. Studies have argued that collocations are less often than the words that constitute them; thus, students mostly lack adequate exposure to them. Familiarisation with collocations is good when studying a second language and L2 collocation; thus, continuous input aids in their learning. A significant amount of collocation input contributes to collocation studies since dialect input is vital in studying other L2 elements, yet it is insufficient.

Non-native speakers are not attentive to collocational links between words even when they face collocations. Contrary to the acquisition of collocations by native speakers, their non-native peers are affected by their mother tongue in both collocation generation and learning. This influence is

a notable aspect often identified as the connection to the frequent erroneous collocation generation in second language collocation studies.

A learner's interlanguage is differentiated from the mature dialect of a proficient speaker by the evidence that the former demonstrates aspects suggesting the incomplete mastery of the code. Their language is characterised by linguistically incorrect and contextually unfitting expressions and forms. Both forms of deviations are seen as errors when connected with the lack of proficiency in the dialect. Besides pragmatic and linguistic deviations, an interlanguage may illustrate some pragmatically and linguistically correct forms yet still feel strange or unnatural. While the language education and learning field are abundant in research of foreign language learners' pragmatic and linguistic errors, research on the strangeness of linguistic expressions and forms is stagnating.

This study found that Saudi ESL students are able to produce a high percentage of acceptable collocations, even without explicit instruction. The study also found that higher level learners are more likely to produce idiomatic collocations than lower-level learners. This is likely because higher level learners have a better understanding of the meaning and use of collocations.

The study also found that the most noticeable difference between the levels was in the production of verb-noun collocations. This is likely because verb-noun collocations are more frequent than other types of collocations.

The findings of this study suggest that Saudi ESL students are able to acquire collocations through implicit learning. However, explicit instruction may be helpful for learners who want to improve their ability to produce idiomatic collocations.

Chapter 7 Conclusion

The results indicated that university students in Saudi Arabia are able to produce acceptable collocations, the verb-noun collocations are the most commonly used collocations in students writing, followed by adjective-noun collocations. When examining the production of collocations according to the type of collocation and the year of study, it is evident that year one, year two, and year three produced more verb-noun collocations than any other type of collocations. It is then followed by adjective-noun collocations. Noun-noun collocations came last and scored only.

According to my research, university students are able to produce acceptable collocations on three levels of fixedness even though collocations are not explicitly taught in their English classes. This result is particularly noteworthy since it is consistent across three years of study. This suggests that these students have been learning how to produce collocations all by themselves by combining the knowledge and skills they have acquired, both inside and outside the classroom, into collocations, which in turn suggests that they have been learning how to produce collocations independently.

7.1 Implications of the study

Collocational teaching practice can be improved by analyzing corpus-based research to reveal the most frequent collocations and those that contain words that are highly frequent in collocations. As a result, it can also help learners learn more about the patterns in collocations and how they are used.

In a classroom setting, learners can be given tasks to accomplish or can be discussed about their production, whether it is correct or incorrect. It can be helpful to show learners samples of original native speaker output, such as those available in corpora, to further encourage correct collocational production. A teacher can use corpora for evaluating collocations produced by learners in a classroom, such as that used in this study, to evaluate collocations produced by their students. By comparing their learners' output to native speakers, they can show them examples of naturally occurring L2 languages.

Teachers and researchers may be able to benefit from this study in terms of understanding learners' misused collocations. Learning collocations involves much more than simply compiling a large number of collocations that are correctly matched; learners must also understand how and when to use collocations.

Teachers need to conduct an analysis of learner needs in order to prioritize certain collocations to teach to their students. Furthermore, this study confirmed that the L1/L2 difference plays a significant role in collocation acquisition. When referring to English collocations that have equivalents in Arabic, it is important to emphasize the techniques of phrase-by-phrase or word-by-word translation in order to achieve the effects of positive transfers.

Frequent exposure to collocations seems to facilitate the development of collocational knowledge. Students can be encouraged to keep a lexical notebook as well as engage in extracurricular reading. Consequently, they will have a better chance of encountering both old and new collocations in a range of meaningful contexts. To increase the use of the notebook in the classroom, students can organize a peer review session in which they read and compare the collocations they have recorded. Also, students can benefit from other external resources such as reference books (e.g. dictionaries) or websites when learning about lexical collocations from their classmates.

7.2 Limitations of the Study

This study was restricted to only three types of collocations and thus does not capture the participants' overall knowledge of collocations. Examining more types and considering other aspects of collocations, such as high/low frequency, may yield different results.

Although this study revealed interesting findings about Saudi EFL students' knowledge of lexical collocations, these findings cannot be generalized to the wider population. However, they do suggest a trend that should be examined by future studies.

7.3 Suggestions for future research

Future research should also address methodological problems involved in capturing and assessing internal learner processes of learning collocations. A variety of different methodologies from controlled laboratory studies to exploratory case studies should be carried out to examine the development of L2 collocational knowledge from different perspectives.

Learners' use of collocations needs to be examined further in a variety of different contexts. The relationship between collocational competence and overall language proficiency needs to be further investigated in terms of different languages. More research is also necessary regarding other types of collocations, other types of language contexts (spoken vs. written, casual vs. professionals, academic vs. daily), and

Chapter 7

other types of language learners (e.g. bilingual speakers and multilingual speakers).

The need for more research on collocations is evident in English which is specific to a specific field, such as engineering, business, and computer. Many questions remain unanswered regarding how explicitly teaching collocations in the classroom will affect second language learners' ability to acquire collocational knowledge and improve their language proficiency.

Appendix A

CONSENT FORM

Study title: Assessment of Collocation Progression in Second Language Learners' Writing at Different Levels of Proficiency: a Corpus-Based Study

Researcher name: AlShahad Aderehim

Staff/Student number: 27475093

ERGO reference number: 30197

Please initial the box(es) if you agree with the statement(s):

I have read and understood the information sheet (insert date /version no. of participant information sheet) and have had the

I agree to take part in this research project and agree for my data

I understand my participation is voluntary and I may withdraw at

Data Protection

I understand that information collected about me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous.

Name of participant (print name).....

Signature of participant.....

Date.....

Participant Information Sheet

Study Title: Assessment of Collocation Progression in Second Language Learners' Writing at Different Levels of Proficiency: a Corpus-Based Study

Researcher: AlShahad Adereihim

Ethics number: 30197

Please read this information carefully before deciding to take part in this research. If you are happy to participate, you will be asked to sign a consent form.

What is the research about?

Hi ! My name is AlShahad I am a PhD student at the Modern Language department at the University of Southampton. This is an English writing skill test that I would like you to take part in for my PhD degree project.

Why have I been chosen?

Because I need university level participants who are studying English as a second language learners to take my test.

What will happen to me if I take part?

You will have to take an English writing skill test on agreed time and place.

Are there any benefits in my taking part?

You taking a part in this study will not only help me but you are going to help the field of English language learning and teaching. ESL learners, teachers, researchers will all benefit from this study.

Are there any risks involved?

There are no risks.

Will my participation be confidential?

According to University of Southampton ethical policies, all personal data will be confidential at all times and will never be declared to any third party.

What happens if I change my mind?

You have the right to withdraw from participating in the study at any time, but it would be appreciated if the researcher could be informed as soon as possible so that they can find a suitable replacement.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you should contact: Research Integrity and Governance Team on rgoinfo@soton.ac.uk

Where can I get more information?

Should you require any further information regarding this study, please contact me:

AlShahad AlDereihim: aaa4g14@soton.ac.uk

My research supervisor: Dr. Ying Zheng, Ying.Zheng@soton.ac.uk

Writing Task 1

You should spend about 40 minutes on this task.

Write an essay about the following topic:

Some people say that the only reason for learning a foreign language is in order to travel or to work in a foreign country. Others say that these are not the only reasons why someone should learn a foreign language.

Discuss both these views and give your own opinion.

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Writing Task 2

You should spend about 40 minutes on this task.

Write an essay about the following topic:

Some people think that universities should provide graduates with the knowledge and skills needed in the workplace. Others think that the true function of a university should be to give access to knowledge for its own sake, regardless of whether the course is useful to an employer.

What, in your opinion, should be the main function of a university?

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Writing Task 3

You should spend about 40 minutes on this task.

Write an essay about the following topic:

Some people think that parents should teach children how to be good members of society. Others, however, believe that school is the place to learn this

Discuss both these views and give your own opinion.

Give reasons for your answer and include any relevant examples from your own knowledge or experience.

Write at least 250 words.

Appendix B

Table 46 Test of Normality

type		Tests of Normality						
		Kolmogorov-Smirnov ^a				Shapiro-Wilk		
		year	c	df	Sig.	Statistic	df	Sig.
Adjective + noun Collocat	relative	1.00	.291	28	.000	.582	28	.000
	frequency	2.00	.314	41	.000	.480	41	.000
		3.00	.312	51	.000	.490	51	.000
Noun + noun collocation	relative	1.00	.292	4	.	.796	4	.095
	frequency	2.00	.351	10	.001	.527	10	.000
		3.00	.351	12	.000	.664	12	.000
Verb + noun collocations	relative	1.00	.284	38	.000	.576	38	.000
	frequency	2.00	.373	58	.000	.297	58	.000
		3.00	.429	51	.000	.239	51	.000

a. Lilliefors Significance Correction

Table 47 Test of Normality

type		Tests of Normality						
		Kolmogorov-Smirnov ^a				Shapiro-Wilk		
		year	c	df	Sig.	Statistic	df	Sig.
Adjective + noun Collocat	logY	1.00	.176	28	.026	.849	28	.001
	2	2.00	.147	41	.026	.884	41	.001
		3.00	.146	51	.008	.869	51	.000
Noun + noun collocation	logY	1.00	.297	4	.	.827	4	.160
	2	2.00	.279	10	.027	.741	10	.003
		3.00	.183	12	.200*	.898	12	.149
Verb + noun collocations	logY	1.00	.126	38	.132	.900	38	.003
	2	2.00	.172	58	.000	.814	58	.000
		3.00	.157	51	.003	.813	51	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 48 Test of Normality

type		Tests of Normality						
		Kolmogorov-Smirnov ^a				Shapiro-Wilk		
		year	c	df	Sig.	c	df	Sig.
Adjective + noun Collocat	log Dice score	1.00	.104	28	.200*	.971	28	.619
		2.00	.083	42	.200*	.974	42	.459
		3.00	.078	51	.200*	.980	51	.521
Noun + noun collocation	log Dice score	1.00	.413	4	.	.698	4	.011
		2.00	.291	10	.016	.700	10	.001
		3.00	.194	12	.200*	.873	12	.072
Verb + noun collocations	log Dice score	1.00	.132	38	.095	.951	38	.094
		2.00	.080	58	.200*	.979	58	.397
		3.00	.083	51	.200*	.973	51	.289

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Table 49 Spearman's correlation

type				Correlations		
				year	relative frequency	
Adjective + noun Collocat	Spearman's rho	year	Correlation Coefficient	1.000	-.218*	
			Sig. (2-tailed)	.	.017	
			N	121	120	
	relative frequency	Correlation Coefficient	-.218*	1.000		
		Sig. (2-tailed)	.017	.		
		N	120	120		
Noun + noun collocation	Spearman's rho	year	Correlation Coefficient	1.000	-.201	
			Sig. (2-tailed)	.	.326	
			N	26	26	
	relative frequency	Correlation Coefficient	-.201	1.000		
		Sig. (2-tailed)	.326	.		
		N	26	26		
Verb + noun collocations	Spearman's rho	year	Correlation Coefficient	1.000	-.245**	
			Sig. (2-tailed)	.	.003	
			N	147	147	
	relative frequency	Correlation Coefficient	-.245**	1.000		
		Sig. (2-tailed)	.003	.		
		N	147	147		

*. Correlation is significant at the 0.05 level (2-tailed).

** . Correlation is significant at the 0.01 level (2-tailed).

Bibliography

- Ackermann, K., & Chen, Y.-H. (2013). Developing the Academic Collocation List (ACL)—A corpus-driven and expert-judged approach. *Journal of English for academic purposes*, 12(4), 235-247.
- Akeel, E. S. (2016). The Role of Native and Learner Corpora in Vocabulary Test Design. *English Language Teaching*, 9(7), 10-15.
- AlAjaji, E. A. (2016). *Self-representation in academic writing: a corpus-based exploratory study of the College of Nursing students' academic writing* [University of Birmingham].
- Alangari, M. A. (2019). *A corpus-based study of verb-noun collocations and verb complementation clause structures in the writing of advanced Saudi learners of English* [University of Reading].
- Alderson, C. (1996). Do corpora have a role in language assessment. *Using corpora for language research*, 248-259.
- Alghizzi, T. M. (2012). The role of English writing instruction methodologies on the types of written mistakes/errors Saudi EFL pre-university students can identify in their writings (Unpublished Master's thesis). University College Cork, Cork, Ireland.
- Alharbi, R. M. S. (2017). *Acquisition of lexical collocations: A corpus-assisted contrastive analysis and translation approach* [Newcastle University].
- Alijani, S., Hafezieh, M. S., & Ghorbani, M. (2020). The reliability and validity of a collocational knowledge examination: A pilot study with East Asian English learners. *International Journal of Language Testing Research and Evaluation*, 6(1), 63-82.
- Al-Khairy, M. A. (2013). Saudi English-Major Undergraduates' Academic Writing Problems: A Taif University Perspective. *English Language Teaching*, 6(6), 1-12.
- Almukhaizeem, Y. S. (2013). Investigating students' ability in handling problems in writing mechanics at King Saud University. *International Journal of Linguistics*, 5(6), 22.
- Alqahtani, M. (2015). The importance of vocabulary in language learning and how to be taught. *International journal of teaching and education*, 3(3), 21-34.
- Alsakran, R. A. (2011). *The productive and receptive knowledge of collocations by advanced Arabic-speaking ESL/EFL learners* [Colorado State University].
- Alshammari, S. R. (2016). Improving Saudi English Learners' second-Language Acquisition In Argumentative Writing Through Self-Regulated Strategy Development.
- 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.
- Alzahrani, A. (2021). The Effects of Two Association Measures on L2 Collocation Processing. *International Journal of English Linguistics*, 11(5).

- Anderson, R. C., & Freebody, P. (1979). Vocabulary Knowledge. Technical Report No. 136.
- Anthony, L. (2014). *AntConc*. In Waseda University.
<http://www.laurenceanthony.net/software/antconc/>
- Appel, R. (2020). A Contrastive Interlanguage Analysis of Linking Adverbials in EFL Writing: Identifying L1 Related Differences. *Learner Corpus Studies in Asia and the World*, 4, 24-40.
- Archibald, A. (2004). Writing in a second language. *The Higher Education Academy Subject Centre for Languages, Linguistics and Area Studies*. Retrieved July, 20, 2016.
- Ariffin, A., & Abdi, B. M. (2020). The relationship between collocation competence and writing skills of EFL learners. *AJELP: Asian Journal of English Language and Pedagogy*, 8(1), 41-52.
- Arnold, I. V., & English, W. M. (1986). High School.
- Aston, G. (1997). Small and large corpora in language learning. PALC,
- Barfield, A. (2012). Lexical collocations. *The encyclopedia of applied linguistics*.
- Barfield, A. W. (2006). *An exploration of second language collocation knowledge and development*. Swansea University (United Kingdom).
- Barker, F. (2006). Corpora and language assessment: Trends and prospects. *Research notes*, 26, 2-4.
- Barker, F. (2010). How can corpora be used in language testing. *The Routledge handbook of corpus linguistics*, 633-645.
- Barker, F. (2013). Corpus-Based Testing. *The encyclopedia of applied linguistics*.
- Barker, F. (2014). Using corpora to design assessment. *The companion to language assessment*.
- Belz, J. A. (2007). The role of computer mediation in the instruction and development of L2 pragmatic competence. *Annual Review of Applied Linguistics*, 27, 45-75.
- Benson, M., Benson, E., & Ilson, R. (1997). *The BBI dictionary of English word combinations*. John Benjamins Pub. Co.
- Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, 8(4), 243-257.
- Biber, D. (2009). A corpus-driven approach to formulaic language in English. *International Journal of Corpus Linguistics*, 14(3), 275-311.
- Boulton, A. (2012). Corpus consultation for ESP: A review of empirical research. *Corpus-informed research and learning in ESP: Issues and applications*, 261-291.
- Brezina, V. (2018). Statistics in corpus linguistics: A practical guide. In: Cambridge University Press.
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173.
- Cadierno, T., & Eskildsen, S. W. (2015). Usage-based perspectives on second language learning. In: De Gruyter Mouton Berlin, Germany.

Bibliography

- Callies, M. (2015). Using Learner Corpora in Language Testing and Assessment: Current Practice and Future Challenges. In E. Castello, K. Ackerley, & F. Coccetta (Eds.), *Studies in Learner Corpus Linguistics: Research and Applications for Foreign Language Teaching and Assessment* (pp. 358). Peter Lang.
- Callies, M., & Götz, S. (2015). Learner corpora in language testing and assessment: Prospects and challenges. *Callies & Götz (eds)*, 1-9.
- Carlsen, C. (2012). Proficiency level—A fuzzy variable in computer learner corpora. *Applied Linguistics*, 33(2), 161-183.
- Carter, R., & McCarthy, M. (2006). *Cambridge grammar of English: A comprehensive guide*. Cambridge University Press.
- Chan, T. P., & Liou, H. C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations. *Computer assisted language learning*, 18(3), 231-251.
- Chang, Y. (2018). Features of Lexical Collocations in L2 Writing: A Case of Korean Adult Learners of English. *English Teaching*, 73(2), 3-36.
- Charles, M., Pecorari, D., & Hunston, S. (2009). *Academic writing. [electronic resource] : at the interface of corpus and discourse* [Computer File]. London ; New York : Continuum, c2009. <http://search.ebscohost.com/login.aspx?direct=true&db=cat02326a&AN=usl.1421167&site=eds-live>
- Chen, A. C.-H. (2021). Acquisition of L2 Collocation Competence: A Corpus Analysis of Exclusivity, Directionality, Dispersion and Novel Usage. *Taiwan Journal of TESOL*, 18(1), 29-61.
- Chen, W. (2019). Profiling collocations in EFL writing of Chinese tertiary learners. *RELC Journal*, 50(1), 53-70.
- Chen, Y.-H., & Baker, P. (2016). Investigating criterial discourse features across second language development: Lexical bundles in rated learner essays, CEFR B1, B2 and C1. *Applied Linguistics*, 37(6), 849-880.
- COCA. (2017). *Corpus of Contemporary American English* <http://corpus.byu.edu/coca/>
- Cotos, E. (2014). Enhancing writing pedagogy with learner corpus data. *ReCALL*, 26(2), 202-224.
- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of quantitative linguistics*, 17(2), 94-100.
- Cowie, A. P. (2002). *English dictionaries for foreign learners: A history*. Oxford University Press on Demand.
- Coxhead, A. (2000). A new academic word list. *TESOL quarterly*, 34(2), 213-238.
- Crossley, S. A., Salsbury, T., & McNamara, D. S. (2015). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics*, 36(5), 570-590.
- Cumming, A. (2010). Theories, frameworks, and heuristics: Some reflections on inquiry and second language writing. *Practicing theory in second language writing*, 19-47.

- Demir, C. (2017). Lexical collocations in English: A comparative study of native and non-native scholars of English. *Journal of Language and Linguistic Studies*, 13(1), 75-87.
- Demirel, E. T., & Kazazoğlu, S. (2015). The comparison of collocation use by Turkish and Asian learners of English: the case of TCSE corpus and icnale corpus. *Procedia-Social and Behavioral Sciences*, 174, 2278-2284.
- Díaz-Negrillo, A., & Thompson, P. (2013). Learner corpora: Looking towards the future. *Automatic treatment and analysis of learner corpus data*, 9-30.
- Ding, J. (2018). A Review of Related Literature on Collocation: Firthianism and Beyond. *Linguistic Prefabrication*, 17-34.
- Division, C. o. E. C. f. C. C.-o. E. C. M. L. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes*, 43, 49-61.
- Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding development and proficiency in writing: quantitative corpus linguistic approaches*. Cambridge University Press.
- Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming.
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching*, 47(2), 157-177.
- El-Dakhs, S., Abdel, D., Amroun, F. P., & Charlot-Muhammad, M. (2018). What works better for collocation learning-Explicit instruction or incidental learning? A case study of Arab female undergraduate learners of English. *Electronic Journal of Foreign Language Teaching*, 15(1), 39-54.
- Ellis, N. (2012). Frequency-based accounts of second language acquisition. *The Routledge handbook of second language acquisition*, 193-210.
- Ellis, N. C. (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.
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. *The handbook of second language acquisition*, 14, 63.
- Ellis, N. C. (2008). The periphery and the heart of language. *Phraseology: An interdisciplinary perspective*, 1-13.
- English Profile. (2015). *Grammar Spotlight*. <http://englishprofile.org/english-grammar-profile/grammar-spotlight>

Bibliography

- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text & Talk*, 20(1), 29-62.
- Evert, S. (2008). Corpora and collocations. *Corpus linguistics. An international handbook*, 2, 1212-1248.
- Eyckmans, J. (2009). Toward an assessment of learners' receptive and productive syntagmatic knowledge. In *Researching collocations in another language* (pp. 139-152). Springer.
- Ezza, E. S. (2010). Arab EFL learners' writing dilemma at tertiary level. *English Language Teaching*, 3(4).
- Farooqui, A. S. (2016). *A corpus-based study of academic-collocation use and patterns in postgraduate Computer Science students' writing* University of Essex].
- Fernández, B. G., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *ITL-International Journal of Applied Linguistics*, 166(1), 94-126.
- Fioravanti, I., Senaldi, M. S. G., Lenci, A., & Siyanova-Chanturia, A. (2021). Lexical fixedness and compositionality in L1 speakers' and L2 learners' intuitions about word combinations: Evidence from Italian. *Second Language Research*, 37(2), 291-322.
- Firth, J. R. (1957). The technique of semantics. *Papers in Linguistics 1934-1951*. In: Oxford University Press, London.
- Flowerdew, J. (2014). *Academic discourse*. Routledge.
- Flowerdew, L. (2004). The argument for using English specialized corpora to understand academic and professional language. *Discourse in the professions: Perspectives from corpus linguistics*, 11-33.
- Foster, I., Ghani, R., Jarmin, R. S., Kreuter, F., & Lane, J. (2020). *Big data and social science: data science methods and tools for research and practice*. CRC Press.
- Friedberg, C., Mitchell, A., & Brook, E. (2017). Understanding academic language and its connection to school success. *The Education Digest*, 82(6), 58.
- Fulcher, G. (2013). *Practical language testing*. Routledge.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, 67(S1), 155-179.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.
- Gitsaki, C. (1996). The development of ESL collocational knowledge.
- Gitsaki, C. (1998). Second language acquisition theories: Overview and evaluation. *Journal of communication and international studies*, 4(2), 89-98.

- Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, 19(2), 115-135.
- González, A. O., & Ramos, M. A. (2013). A Comparative Study of Collocations in a Native Corpus and a Learner Corpus of Spanish [Article]. *Procedia - Social and Behavioral Sciences*, 95, 563-570. <https://doi.org/10.1016/j.sbspro.2013.10.683>
- Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and lexical phrases. *Phraseology: Theory, analysis and applications*, 145-160.
- Granger, S. (2004). Computer learner corpus research: current status and future prospects. *Language and Computers*, 52(1), 123-145.
- Granger, S. (2012). How to use foreign and second language learner corpora. *Research methods in second language acquisition: A practical guide*, 7-29.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24.
- Granger, S. (2018). *Formulaic sequences in learner corpora: Collocations and lexical bundles* [Chapter]. Taylor and Francis. <https://doi.org/10.4324/9781315206615>
- Granger, S., & Bestgen, Y. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study. *International Review of Applied Linguistics in Language Teaching*, 52(3), 229-252.
- Gries, S. T. (2018). On over-and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies*, 1(2), 276-308.
- Gries, S. T., & Ellis, N. C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65(S1), 228-255.
- Guo, X., Zheng, L., Zhu, L., Yang, Z., Chen, C., Zhang, L., Ma, W., & Dienes, Z. (2011). Acquisition of conscious and unconscious knowledge of semantic prosody. *Consciousness and cognition*, 20(2), 417-425. <https://doi.org/10.1016/j.concog.2010.06.015>
- Halliday, M. A. (1966). Some notes on 'deep' grammar. *Journal of Linguistics*, 2(1), 57-67.
- Halliday, M. A. (2004). *An introduction to functional grammar*. revised by CMIM Matthiessen. London: Arnold.
- Hammad, E. A. (2016). Palestinian university students' problems with EFL essay writing in an instructional setting. *Teaching EFL writing in the 21st century Arab world: Realities and challenges*, 99-124.
- Haugen, E. (2018). *The Norwegian Language in America, a Study in Bilingual Behavior, Volume 2*. In *The Norwegian Language in America, a Study in Bilingual Behavior, Volume 2*. University of Pennsylvania Press.
- Haugen, E. I. (1969). *The Norwegian language in America: A study in bilingual behavior* (Vol. 1). Indiana University Press.

Bibliography

- Hawkey, R., & Barker, F. (2004). Developing a common scale for the assessment of writing. *Assessing Writing*, 9(2), 122-159.
- Henriksen, B. (2013). Research on L2 learners' collocational competence and development—a progress report. C. Bardel, C. Lindqvist, & B. Laufer (Eds.) *L*, 2, 29-56.
- Hill, J., Lewis, M., Kozłowska, C. D., & Dzierzanowska, H. (1997). *Dictionary of selected collocations*. Language Teaching Publications.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Psychology Press.
- Hoey, M. (2012). *Lexical priming: A new theory of words and language*.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics*, 19(1), 24-44.
- Howarth, P. A. (2013). Phraseology in English academic writing. In *Phraseology in English Academic Writing*. Max Niemeyer Verlag.
- Hsu, J.-Y., & Chiu, C. Y. (2008). Lexical collocations and their relation to speaking proficiency of college EFL learners in Taiwan. *Asian EFL Journal*, 10(1), 181-204.
- Hsu, J.-y. T. (2002). *Development of collocational proficiency in a workshop on English for general business purposes for Taiwanese college students*. Indiana University of Pennsylvania.
- Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System*, 53, 13-23.
- Huang, L.-S. (2001). Knowledge of English Collocations: An Analysis of Taiwanese EFL Learners.
- Huat, C. M. (2012). Learner Corpora and Second Language Acquisition. In K. Hyland, C. M. Huat, M. Handford, & S. Hunston (Eds.), *Corpus Applications in Applied Linguistics* (pp. 191-207). Continuum International Publishing Group. <https://search.ebscohost.com/login.aspx?direct=true&db=edszmh&AN=2017060078&site=eds-live>
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Hyland, K., & Tse, P. (2007). Is there an “academic vocabulary”? *TESOL quarterly*, 41(2), 235-253.
- Jesmani, M., Jafarpour, B., Bellout, M. C., & Foss, B. (2020). A reduced random sampling strategy for fast robust well placement optimization. *Journal of Petroleum Science and Engineering*, 184, 106414.
- Jones, S., & Sinclair, J. (1974). English lexical collocations. *Cahiers de lexicologie*, 24(1), 15-61.
- Jones, S., & Sinclair, J. M. (2012). English lexical collocations-A study in computational linguistics. *Cahiers de lexicologie*, 1974(24), 17-63.
- Joyce, P. (2018). L2 vocabulary learning and testing: The use of L1 translation versus L2 definition. *The Language Learning Journal*, 46(3), 217-227.

- Jun, L., & Dienes, Z. (2021). Do second language learners always know what they know about L2 collocations? Exploring the acquisition of conscious and unconscious structural knowledge of L2 collocations.
- Kamarudin, R., Abdullah, S., & Aziz, R. A. (2020). Examining ESL learners' knowledge of collocations. *International Journal of Applied Linguistics and English Literature*, 9(1), 1-6.
- Kaplan, R. B. (1966). Cultural thought patterns in intercultural education. *Language Learning*, 16(1), 1-20.
- Karim, K., & Nassaji, H. (2013). First Language Transfer in Second Language Writing: An Examination of Current Research. *Iranian Journal of Language Teaching Research*, 1(1), 117-134.
- Khoja, H. Y. Y. (2019). *Exploring the Use of Collocation in the Writing of Foundation-Year Students at King Abdulaziz University* [University of Leeds].
- Kilgarriff, A. (2009). Simple maths for keywords. Proceedings of the Corpus Linguistics Conference. Liverpool, UK,
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7-36.
- Kjellmer, G. (1984). Some thoughts on collocational distinctiveness. J. Aarts & W. Meijs (eds.) *Corpus linguistics. Recent developments in the use of computer corpora in English language research*. In: Amsterdam: Rodopi.
- Klassen, R. (2002). Writing in early adolescence: A review of the role of self-efficacy beliefs. *Educational psychology review*, 14(2), 173-203.
- Klotz, M. (2003). Oxford collocations dictionary for students of English. In: Oxford University Press.
- Koban Koc, D. (2021). Frequency and development of collocations in Turkish EFL learners' essays. *University of South Florida M3 Center Publishing*, 3(2021), 25.
- Koç, G. (2006). Developing collocational awareness. PQDT-Global.
- Kolesnikova, O. (2016). Survey of word co-occurrence measures for collocation detection. *Computación y Sistemas*, 20(3), 327-344.
- Koosha, M., & Jafarpour, A. A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL journal*, 8(4), 192-209.
- Krashen, S. D., & Terrell, T. (1983). Natural approach. In: Pergamon New York.
- Kučera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. In: Dartmouth.
- Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, 34(4), 513-535.

Bibliography

- Kyle, K., Crossley, S., & Verspoor, M. (2021). Measuring longitudinal writing development using indices of syntactic complexity and sophistication. *Studies in Second Language Acquisition*, 43(4), 781-812.
- Lam, S. L., & Crosthwaite, P. (2018). Appraisal resources in L1 and L2 argumentative essays: A contrastive learner corpus-informed study of evaluative stance. *Journal of Corpora and Discourse Studies*, 1(1), 8-35.
- Lateh, N. H. M., Shamsudin, S., Raof, A. H. A., Mohamed, A. F., Mahmud, N., NS, M. N., & Fatihah, N. (2021). Learners' collocation use in writing: Do proficiency levels matter? *Indonesian Journal of Applied Linguistics*.
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29(4), 694-716.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647-672.
- Le Bruyn, B., & Paquot, M. (2021). *Learner corpus research meets second language acquisition*. Cambridge University Press.
- Lee, C. Y., & Liou, H. C. (2003). A study of using web concordancing for English vocabulary learning in a Taiwanese high school context. *English Teaching & Learning*, 27(3), 35-56.
- Leech, G. (1991). The State of the Art in Corpus Linguistics. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 8-29). Longman.
<http://search.ebscohost.com/login.aspx?direct=true&db=edszh&AN=1991013130&site=eds-live>
- Leech, G. (1997). Teaching and language corpora: A convergence. *Teaching and language corpora*, Lewis, M. (1997). *Lexical approach to language teaching*. Cambridge University Press.
- Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. *Perspectives on formulaic language: Acquisition and communication*, 22-46.
- Li, P., Cadierno, T., & Eskildsen, S. W. (2014). Tracing an L2 Learner's Motion Constructions over Time: A Usage-Based Classroom Investigation. *The Modern Language Journal*, 98(2), 612-628. <https://search.ebscohost.com/login.aspx?direct=true&db=edszh&AN=2014651737&site=eds-live>.
- Lin, Y. P. (2002). The effects of collocation instruction on English vocabulary development of junior high school students in Taiwan. National Kaohsiung Normal University, Kaohsiung, Taiwan.
- Lorenz, G. R. (1999). Adjective Intensification-Learner's Versus Native Speakers: A Corpus Study of Argumentative Writing. In: Rodopi.

- Lozano, C., & Mendikoetxea, A. (2013). Learner corpora and second language acquisition. *Automatic treatment and analysis of learner corpus data*, 59, 65-100.
- Lüdeling, A., & Kytö, M. (2008). Corpus linguistics: An international handbook. In: Walter de Gruyter GmbH.
- Macis, M., & Schmitt, N. (2017). Not just 'small potatoes': Knowledge of the idiomatic meanings of collocations. *Language Teaching Research*, 21(3), 321-340.
- Malamatidou, S. (2017). *Corpus triangulation: Combining data and methods in corpus-based translation studies*. Routledge.
- Manchón, R. (2012). *L2 writing development: Multiple perspectives* (Vol. 6). Walter de Gruyter.
- Marcus Callies Bremen, S. G. G. (2013). *(Learner) Corpora and their Application in Language Testing and Assessment*
- Martyńska, M. (2004). Do English language learners know collocations? *Investigationes linguisticae*, 11, 1-12.
- McEnery, T. (2019). Corpus linguistics. In: Edinburgh University Press.
- McEnery, T., & Wilson, A. (2003). Corpus linguistics. *The Oxford handbook of computational linguistics*, 448-463.
- McEnery, T., & Xiao, R. (2011). What corpora can offer in language teaching and learning. *Handbook of research in second language teaching and learning*, 2, 364-380.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.
- Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics*, 26(1), 32-47.
- Men, H. (2015). Vocabulary increase and collocation learning : a corpus-based cross-sectional study of Chinese EFL learners. In: Birmingham City University.
- Men, H. (2017). *Vocabulary increase and collocation learning: A corpus-based cross-sectional study of Chinese learners of English*. Springer.
- Mirsalari, S. A., Bazvand, A. D., & Khoram, A. (2019). An investigation into the developmental patterns of lexical collocation among Iranian EFL learners. *Applied Linguistics Research Journal*, 3(2), 48-69.
- Mohamed, A. H., & Omer, M. R. (1999). Syntax as a marker of rhetorical organization in written texts: Arabic and English. *IRAL, International Review of Applied Linguistics in Language Teaching*, 37(4), 291.
- Mukherjee, J. (2005). The native speaker is alive and kicking: Linguistic and language-pedagogical perspectives. *Anglistik*, 16(2), 7-23.
- Mukherjee, J. (2009). The grammar of conversation in advanced spoken learner English. *Corpora and language teaching*, 203-230.

Bibliography

- Mustafa, F. (2015). Using corpora to design a reliable test instrument for English proficiency assessment. Proceedings of the TEFLIN International Conference: Teaching and assessing L2 learners in the 21st century,
- Nation, I. Paul 2001. Learning Vocabulary in Another Language. *Cambridge: CUP*.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223-242.
- Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing.
- Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, 21(3), 298-320.
- Nizonkiza, D. (2012). Quantifying controlled productive knowledge of collocations across proficiency and word frequency levels. *Studies in Second Language Learning and Teaching*, 2(1), 67-92.
- Obeidat, A. M., & Mahadi, T. S. T. (2019). The translation of Arabic religious cultural collocations in literary texts into English: An application of domestication and foreignization translation strategies. *International Journal of Humanities, Philosophy and Language*, 2(6), 155-165.
- Obukadeta, P. (2019). *Collocations in a Learner English Corpus: analysis of Yoruba-speaking Nigerian English learners' use of collocations* [Kingston University].
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
- Oppenheimer, D., Zaromb, F., Pomerantz, J. R., Williams, J. C., & Park, Y. S. (2017). Improvement of writing skills during college: A multi-year cross-sectional and longitudinal study of undergraduate writing performance. *Assessing Writing*, 32, 12-27.
- Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English. *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins, 67-83.
- Paquot, M. (2017). L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL learner writing. *Second Language Research*, 33(1), 13-32.
- Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, 35(1), 121-145. <https://doi.org/10.1177/0267658317694221>
- Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, 32, 130-149.
- Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, 3(1), 61-94.
- Park, K. (2014). Corpora and language assessment: The state of the art. *Language Assessment Quarterly*, 11(1), 27-44.
- Parkes, J. (2007). Reliability as argument. *Educational measurement: issues and practice*, 26(4), 2-10.

- Paulsen, G. (2018). Manner and adverb: Fuzzy categorial boundaries in collocations. *Eesti Rakenduslingvistika Ühingu aastaraamat*(14), 117-135.
- Pavičić Takač, V. (2013). How word choice matters: An analysis of adjective-noun collocations in a corpus of learner essays. *Jezikoslovlje*, 14(2-3), 385-402.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, 191, 225.
- Pellicer-Sánchez, A. (2017). Learning L2 collocations incidentally from reading. *Language Teaching Research*, 21(3), 381-402.
- Polio, C. (2017). Second language writing development: A research agenda. *Language Teaching*, 50(2), 261-275.
- Polio, C., & Yoon, H.-J. (2021). Exploring multi-word combinations as measures of linguistic accuracy in second language writing. *Learner corpus research meets second language acquisition*, 96-121.
- Qin, J. (2014). Use of formulaic bundles by non-native English graduate writers and published authors in applied linguistics. *System*, 42, 220-231.
- Rakedzon, T., & Baram-Tsabari, A. (2017). To make a long story short: A rubric for assessing graduate students' academic and popular science writing skills. *Assessing Writing*, 32, 28-42.
- Ramadhan, J. M. (2017). The Influence of Features of Collocations on the Collocational Knowledge and Development of Kurdish High School Students: A Longitudinal Study.
- Rebuschat, P., & Williams, J. (2012). *Statistical learning and language acquisition*. [electronic resource] [Computer File]. Berlin ; Boston : De Gruyter Mouton, 2012.
<http://search.ebscohost.com/login.aspx?direct=true&db=cat02326a&AN=usl.1606445&site=eds-live>
<http://site.ebrary.com/lib/soton/Doc?id=10634504>
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225.
- Römer, U., O'Donnell, M. B., & Ellis, N. C. (2015). Using COBUILD grammar patterns for a large-scale analysis of verb-argument constructions. *Corpora, grammar and discourse: In honour of Susan Hunston*, 73, 43.
- Sadeghi, Z. (2010). The Importance of Collocation in Vocabulary Teaching and Learning. *Translation Journal* . <https://translationjournal.net/journal/52collocation.htm>
- Scott, M., & Press, O. U. (2008). Wordsmith Tools: version 4.0: single-user licence. In.
- Service, E. T. (2006). The official guide to the new TOEFL IBT:[find out all about the new TOEFL internet-based test; get over 500 real TOEFL questions and essay topics; learn the test-maker's strategies for success]. In: McGraw-Hill.

Bibliography

- Service, E. T. (2008). *Official Guide to the New TOEFL iBT*. McGraw-Hill Publishing.
<http://www.mylibrary.com?ID=179180>
- seyed Ali, M., Ali Darabi, B., & Alireza, K. (2019). An Investigation into the Developmental Patterns of Lexical Collocation among Iranian EFL Learners [article]. *Applied Linguistics Research Journal*, 3(2), 48-69. <https://doi.org/10.14744/alrj.2019.02486>
- Shin, D., & Nation, P. (2007). Beyond single words: The most frequent collocations in spoken English. *ELT journal*, 62(4), 339-348.
- Shin, D., & Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT journal*, 62(4), 339-348.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
- Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review*, 64(3), 429-458.
- Siyanova-Chanturia, A. (2015). Collocation in beginner learner writing: A longitudinal study. *System*, 53, 148-160.
- Siyanova-Chanturia, A., & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, 36(5), 549-569.
- Smadja, F. (2021). Macrocoding the lexicon with co-occurrence knowledge. In *Lexical acquisition: Exploiting on-line resources to build a lexicon* (pp. 165-189). Psychology Press.
- Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and EAP writing development: Lexical bundles in the TOEFL iBT writing section. *Journal of English for academic purposes*, 12(3), 214-225.
- Stuart, W. (2008). RECEPTIVE AND PRODUCTIVE VOCABULARY SIZES OF L2 LEARNERS. *Studies in Second Language Acquisition*, 30(1), 79-95. <https://doi.org/10.1017/S0272263108080042>
- Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of language*, 2(1), 23-55.
- Tai, S., & Chen, H.-J. (2015). Are Teachers Test-Oriented? A Comparative Corpus-Based Analysis of the English Entrance Exam and Junior High School English Textbooks. *Research-publishing.net*.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing.
- Tong, Y. T. (2004). A Study of the Collocation Errors in the Oral and Written Production of the College Students in Taiwan. Unpublished M.A. thesis, National Taiwan Normal University, Taiwan.
- Toth, G. (2007). *Linguistic interference and first-language attrition: German and Hungarian in the San Francisco Bay area* (Vol. 59). Peter Lang.

- Tsang, W.K. & Wong, M. (2002). Conversational English: an interactive, collaborative and reflective approach. In: Richards, J.C. & Rendandya, W. (Eds.) *Methodology in Language Teaching*, New York: Cambridge University Press, 212-224.
- Tyler, A. (2010). Usage-based approaches to language and their applications to second language learning. *Annual Review of Applied Linguistics*, 30, 270-291.
- Vasiljevic, Z. (2014). Teaching collocations in a second language: Why, what and how. *Elta Journal*, 2(2), 48-73.
- Vilkaitė, L. (2017). Incidental acquisition of collocations in L2: Effects of adjacency and prior vocabulary knowledge. *ITL-International Journal of Applied Linguistics*, 168(2), 248-277.
- Walker, C. P. (2011). A Corpus-Based Study of the Linguistic Features and Processes Which Influence the Way Collocations Are Formed: Some Implications for the Learning of Collocations [research-article]. *TESOL quarterly*, 45(2), 291-312.
<https://search.ebscohost.com/login.aspx?direct=true&db=edsjrs&AN=edsjrs.41307632&site=eds-live>
- Wang, Y., & Shaw, P. (2008). Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME journal*, 32, 201-232.
- Wang, A., & Pellicer-Sánchez, A. (2022). Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning*, 72(3), 765-805.
- Wardle, E., & Roozen, K. (2012). Addressing the complexity of writing development: Toward an ecological model of assessment. *Assessing Writing*, 17(2), 106-119.
- Webb, S., & Kagimoto, E. (2009). The effects of vocabulary learning on collocation and meaning. *TESOL quarterly*, 43(1), 55-77.
- Webb, S., & Kagimoto, E. (2011). Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics*, 32(3), 259-276.
- Webb, S., Newton, J., & Chang, A. (2013). Incidental learning of collocation. *Language Learning*, 63(1), 91-120.
- Weigle, S. C., & Goodwin, S. (2016). Applications of Corpus Linguistics in Language Assessment. *Contemporary Second Language Assessment: Contemporary Applied Linguistics*, 4, 209.
- Wijitsopon, R. (2017). Collocations and local textual functions of quantifiers in learner english essays [Article]. *Linguistic Research*, 34(1), 1-49.
<https://doi.org/10.17250/khisli.34.1.201703.001>
- 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.
- 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.

Bibliography

<http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ1018131&site=eds-live>
<http://dx.doi.org/10.1017/S0272263113000107>

- Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, 40(2), 395-416.
- Wongkhan, P., & Thienthong, A. (2021). EFL Learners' Acquisition of Academic Collocation and Synonymy: Does their Academic Experience Matter? *RELC Journal*, 52(3), 523-538.
- Wray, A. (2002). Formulaic language and the lexicon.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication*, 20(1), 1-28.
- Wynne, M. (2005). *Developing linguistic corpora: A guide to good practice* (Vol. 92). Oxbow Books Oxford.
- Wynne, M., Sinclair, J., Leech, G., Burnard, L., McEnery, A., Xiao, R., & Thompson, P. (2000). Developing linguistic corpora. *OTA Guides*.
- Zakharov, V. (2017). Evaluation and combining association measures for collocation extraction. *Proceedings of the International Conference IMS-2017*,
- Zhang, Lei. (2020). The Significance of Collocational Knowledge for Learners of English and Teaching Implications. 10.2991/assehr.k.201127.048.