Sudan University of Science and Technology

College of Graduate Studies

College of Languages

Employing Corpora in Solving Collocation Learning Difficulties Encountered by University Students

A Case Study of Second Year Students, Biomedical Engineering, Bayan College of Science and Technology, Khartoum

إستخدام المخزونات اللغوية لحل مشاكل تعلم المتلازمات في اللغة الإنجليزية التي تواجه الطلاب الجامعيين

دراسة حالة طلاب المستوى الثاني بقسم الهندسة الطبية الحيوية، كلية البيان للعلوم و التكنولوجيا، الخرطوم

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Ph.D in English Language (Applied Linguistics)

Submitted by

Ibrahim Alsafi Abdaldfi

Supervised by

Dr.Abdalla Yassin Abdalla

May, 2019

Approval Page

(To be completed after the college council approval)

Name of Candidate: Ibrahim ELSafi Abduldafi Daw-ALnaeem

Thesis title: Employing Corpora in Solving Collocation Learning Difficulties Encountered by University Students

استخدام المدونات اللغوية لحل مشاكل المتلازمات ع اللغة الانجليزية الت تواجه طلاب الجامعات

Degree Examined for: Ph.D. in English (Applied Linguistics)

Approved by:

1. External Examiner

Name: Dr. Ahmed Babiker ELTahir

Signature: .... Date: 9.15/2019

2. Internal Examiner

Name: Dr. Mahmoud Ali Ahmed 9/5/2019

Signature: .... Date: 9/5/2019

3. Supervisor

Name: Dr. Abdalla Yassin Abdalla

Signature: Yassin Date: 9/5/2019

Dedication

To my beloved parents (may Allah rest their souls in peace), to my dearest wife Saadia Yaseen for her never-ending encouragement and unwavering support, to our adorable son Alsafi (my father my son as I call him), to my siblings (especially Mahir, Kreem, Khdeega and Naila), and to my mother-in-law Amna Eltayib for always being there.

Acknowledgements

Abstract

This study aims to appraise the impact of employing corpora in solving collocation learning difficulties encountered by Sudanese undergraduates. To do so, the study adopted the mixed method approach to data collection. Specifically, four tools of data collection are adopted; a pre-test and post-test, student attitude questionnaire, and teacher interview. Second year students of Biomedical Engineering at Bayan College of Science and Technology were chosen as the population. 60 students (37males and 23 females) were selected as the study sample and then based on their performance on the pre-test, they were divided into two groups; the control group in which students were taught collocations (delexical verbs) through a traditional strategy, and the experimental group in which students were taught how to use the corpus and then taught the same collocations using British National Corpus (BNC). After the experiment, results of the post-test revealed that while the performance of the two groups was improved, the results of the experimental group were significantly better than that of the control group. The questionnaire was distributed to participants of the experiment for they can judge whether the corpus is effective in learning collocations. Results of the questionnaire proved that students showed positive attitudes towards using corpora in language learning, especially collocations. Concerning the interview, five EFL teachers from different Sudanese universities were interviewed. The interview results indicated that all the subjects showed positive attitudes towards employing corpora in language teaching and learning but nearly all of them do not use corpora practically in their classrooms. The study findings indicated that corpora have the potential of raising learners' collocational competence and hence their language proficiency. The findings also showed that

using corpus-based strategies is better than using non-corpus-based strategies due to the authentic context that corpora provide. In addition, it is also indicated that EFL teachers should be trained on how to practically use corpora in their classrooms.

Based on the study findings, it is recommended that teachers should use corpus-based strategies. Furthermore, EFL learners should be trained on how to exploit corpora and use them in language learning and investigation. Based on the findings, it is suggested that, more studies need to be conducted on utilizing corpora in learning the other types of collocations in English language. In addition to that, it is suggested that, as it is hard to find an English teacher using corpora in Sudan, more study should be carried out to investigate the underlying factors.

Abstract

(Arabic Version)

يهدف هذا البحث إلي دراسة إستخدام المخزونات اللغوية لحل مشاكل تعلم متلازمات اللغة الإنجليزية التي تواجه طلاب الجامعات السودانية. من أجل ذلك، تبنت الدراسة الطريقة الكمية- الكيفية لجمع البيانات حيث تم إختيار أربع طرق لجمع البيانات و هي: إختبار ما قبل التجربة، إختبار ما بعد التجربة، الإستبيان، و المقابلة.

تم إختيار طلاب المستوى الثاني بقسم الهندسة الطبية الحيوية بكلية البيان للعلوم و التكنولوجيا. تم إختيار ٦٠ طالبا و طالبة بطريقة عشوائية كعينة للبحث. هؤلاء الطلاب تم تقسيمهم بناءً على أدائهم في إختبار ما قبل التجربة إلي مجموعتين؛ مجموعة اللاتجربة (٣٠ طالبا و طالبة) و التي من خلالها تم تدريس المتلازمات اللفظية (الأفعال ضعيفة المعنى و ملازماتها من الأسماء) للطلاب عبر إستراتيجية تقليدية و مجموعة التجربة (٣٠ طالبا و طالبة) و التي عبرها تم تدريس الطلاب كيفية إستخدام المدونات اللغوية ثم بعد ذلك تم تدريسهم المتلازمات نفسها بإستخدام المدونة القومية البريطانية. بعد إجراء التجربة، أثبتت نتائج إختبار ما بعد التجربة أنه قد تحسن أداء مجموعتي التجربة و اللاتجربة و لكن مجموعة التجربة أبلت بلاءً أفضل من مجموعة اللاتجربة.

تم توزيع الإستبيان علي مجموعة التجربة لمعرفة مدى فاعلية إستخدام المدونات في تعلم المتلازمات حيث أثبتت نتائج الإستبيان أن الطلاب أبدوا آراءً إيجابية تجاه إستخدام المدونات. أما ما يتعلق بالمقابلة، تم إجراءها مع خمس أساتذة من مختلف الجامعات السودانية. أشارت نتائج هذه المقابلة إلي أن كل المشاركين لديهم آراء إيجابية تجاه إستخدام المخزونات اللغوية في تعليم و تعلم اللغة الإنجليزية و لكن تقريباً كل المشاركين لم يستخدموا تلك المخزونات في تدريس اللغة الإنجليزية. أشارت نتائج البحث إلي أن المدونات اللغوية قادرة على رفع وعي دارسي اللغة بالمتلازمات اللفظية و من ثم إدراكهم اللغوي. أيضاً أثبتت النتائج أن إستخدام المدونات كإستراتيجية لتعليم و تعلم اللغة أفضل من إستخدام الإستراتيجيات التقليدية نسبة للسياق اللغوي الذي توفره تلك المخزونات. بالإضافة إلي ذلك، أشارت النتائج إلي ضرورة تدريب أساتذة اللغة الإنجليزية كلغة أجنبية على كيفية إستخدام المدونات اللغوية لتدريس اللغة.

بناءً على نتائج البحث، أوصى الباحث بضرورة إستخدام المدونات اللغوية كإستراتيجية لتدريس اللغة. إضافة، يجب تدريس طلاب اللغة الإنجليزية كلغة أجنبية كيفية إستغلال تلك المخزونات

اللغوية و استخدامها في التعلم و التحقيق اللغوي. بناءً على النتائج، أيضاً أوصى الباحث بضرورة أجراء مزيداً من البحث في كيفية إستخدام المدونات لتعليم متلازمات اللغة الإنجليزية الأخرى. كما أنه أوصى الباحث بضرورة إجراء مزيداً من البحث لمعرفة العوامل التي أدت إلى عدم إستخدام المخزونات لتعليم اللغة الإنجليزية في السودان.

List of Contents

List of Tables

List of Figures

Chapter One

Introduction

Chapter One

Introduction

1.0 Overview

In fact, English language has become the globe's first language and mastering it requires mastering its collocation which is considered to be a very essential area in English language learning. This is due to the idiomatic nature of English language and the interrelationship between syntax and semantics. According to Halliday (1966) "*language is organized in terms of lexico-grammatical features, rather than independent lexis or grammar*". This notion emphasizes the idiomatic nature of English language, specially the dependent relationship between vocabulary and the grammatical system. The term collocation has been brought up in the field of lexical study by Firth (1957). According to him, "*the meaning of a word should be known by the company it keeps*". Firth's view proved the fact that vocabulary items do not frequently occur by themselves, they are accompanied by surrounding words in context and they tend to occur more frequently in language units such as idioms, proverbs or fixed expressions that have both fixed form and meaning. Within the lexical approach, special attention is given to collocations and language patterns that include institutionalized utterances and expressions.

"*The notion of collocation emerged from the discipline of lexical studies, as a term defined and understood in many different ways*" (Bahns, 1993).

Similarly, Lewis (1997) maintains "*the notion of collocation is one of the favorite topics for those who follow the lexical approach, a school of thought that says language consists of grammatical lexis, not lexicalized grammar*".

Another view about the nature of collocations is provided by Lewis (1993:204) claiming that "*instead of words we consciously try to think of collocations, and to present these in expressions rather than trying to break things into ever smaller pieces, there is a conscious effort to see things in larger and more holistic ways*".

Recently, language learning and teaching have been revolutionized by a number of computer-based fields such as corpus linguistics and computational linguistics. These two disciplines have been fruitful and noticeably contributed to the development of language learning, teaching and investigation. As Zimmerman (1997) stated, specially with the introduction of work in the area of computational linguistics, corpus linguistics, and lexical approaches, a growing number of scholars (e.g. Sinclair, 1991; Nattinger & Decarrico, 1992; Lewis, 1993), representing a significant theoretical and pedagogical shift from the past with their work, pushed collocations to the center of language acquisition.

Today, it has widely been acknowledged that collocations are considered to be an important part of native speaker competence, and therefore should be integrated into second and foreign language learning and teaching (e.g., Cowie, 1992; Bahns, 1993; Wei, 1999; Lewis, 2000).

An increasing number of researchers have dealt with the problem of collocations from various aspects to provide help for learners and users of English language (e.g., Shei & Pain, 2000; Nesselhauf & Tschichol, 2002; Sun & Wang, 2003, and many others). In another context, emphasizing the significance of learning and mastering collocations, El-hadi (2011:1) stated that: "*it is important for language learners to master collocations so that their*

language sounds more natural and easily understood by native speakers". Additionally, lexicographers and linguists have also expanded the spectrum of studies on collocations.

The focus of this study is on appraising the impact of using corpora in solving the difficulties encountered by university students in learning collocations, and providing learners with different techniques for the retention of these word combinations.

According to Oxford Advanced Learner's Dictionary (2014:84) corpus is "*any systematic collection of speech or writing in a language or variety of a language. Thus, in particular, of large online collections, tagged and researchable for research purposes*".

In summary, collocations have been researched and investigated from various aspects. Nevertheless, delexical verbs (which refer to common verbs such as *make*, *do*, *take*, *get*, etc. when they are used with certain nouns they carry very little meaning of their own. In these delexical structures, most of the meaning is found in the noun, not in the verb) have received very small fortune, in spite of the fact that this form of collocation is a source of difficulty for a large number of EFL/ESL learners even to those of advanced levels, specially Arab learners of English. Thus, the current study focuses explicitly on the impact of employing corpora in solving collocation learning difficulties encountered by Sudanese university students.

1.1 Statement of the Problem

While working as EFL lecturer at a number of Sudanese universities and university colleges, the researcher has frequently noticed his students making

collocational errors in their spoken and written English language, especially delexical verbs. These verbs cause difficulties when they are used with certain nouns, because most of the meaning is found in the noun, not in the verb. For instance, in *make mistake* the focus is on *mistake* while *make* is weaken. In most cases, there is a verb which is appropriate to the delexical structure. These patterns are very common in English language and mastering them enriches one's fluency of the language. They are considered to be an area of difficulty for Sudanese EFL learners studying English language as a field of specialty and to those who study English as a university requirement. This fact is the underlying motive of conducting this study. In this regard, El-hadi (2011:68) conducted a study on assessing English collocational knowledge among Sudanese EFL university students. She stated that: "*the lower percentages are more than the higher ones, which are clear evidence that the students' performance on collocation tests is very low and collocations are a problem for the students in language learning*". It is clearly indicated by this view that collocation is a problematic part of word knowledge because basically every lexical item can to some extent build collocational relationship with other ones. Despite this difficulty, especially to Arab learners of English, few studies have researched and investigated this area. In this regard, El-hadi (2011:2) stated that: "*few studies have investigated the collocational knowledge of Arab speaking learners of English such as El-khatib, (1984); Al-Zahrani, (1998); Mahmoud, (2005). These studies showed that Arab learners of English face a great difficulty with English collocations*".

It is also assumed by this study that the difficulty of using collocations appropriately spans because teaching collocations has largely been neglected in EFL/ESL vocabulary instructions and receives little attention from teachers

4

in the classroom. However, few of the previous studies have explored the effectiveness of computer technology to the development of English language learning and teaching. Despite the fact that the contribution of this technology has been claimed to be powerful, especially the use of corpora which provide language teachers as well as learners with the necessary tools and techniques that help them understand collocations accurately. Such as KWIC (Key Word in Context), TACT (Text Analysis Computational Tool), LancsBox, etc. Taking KWIC for instance, it provides the right and left collocations of a word in different contexts of use. As in the words *great* (left collocation of the word deal) and *of* (right collocation of the word deal) in *a great deal of*. Therefore, the purpose of this study is investigating the impact of utilizing corpora in solving collocation learning difficulties faced by university students and ways of enhancing their knowledge of collocations, delexical verbs in particular.

1.2 Objectives of the Study

This study aims to investigate the impact of employing corpora in solving collocation learning difficulties encountered by Sudanese university students. It aims at increasing the awareness of English language learners of the significance of using corpora in learning English collocations by providing tools and techniques for accessing and using corpora, such as KWIC, TACT, etc.

The study aims to:

- Test the impact of using corpora as helpful and vital tools for solving collocation learning difficulties faced by English as a foreign or second language learners and ways of promoting their collocational competence so as to learn and then use the language appropriately and effectively.

- Shed light on the effectiveness of using corpora in raising students' collocational awareness and the importance of computer technology and how it facilitates understanding and learning collocations.

- Provide some insights about the attitudes of English language teachers on corpora as a new strategy to raising the awareness of English language learners as well as teachers on collocations.

1.3 Questions of the Study

The study aims to find answers to the following questions:

1- To what extent do corpora help in solving collocation learning difficulties encountered by university students?

2- What are the attitudes of English language teachers towards the utilization of corpora in solving collocation learning difficulties encountered by university students?

3- To what extent is there a significant difference between those who study through corpora and those who do not?

1.4 Hypotheses of the Study

The researcher hypothesizes that:

- Corpora help to a far extent in solving collocation learning difficulties encountered by university students.

- It is supposed that English language teachers have positive attitudes towards using corpora in solving collocation learning difficulties encountered by university students.

- It is expected that there is a significant difference between those who study through corpora and those who do not.

1.5 Significance of the Study

The findings of this study will be significant and useful to a wide range of audience such as language teachers, administrators and on top of that, they are especially significant for English language learners due to the demand for effective communication and use of English language as a lingua franca. A fact that justifies the need for more effective and life-changing approaches to English language teaching and learning such as using corpus-based techniques. Thus, educational institutions that apply the findings of this study will improve English language teaching and learning. In addition to that, the researcher can also benefit from the study because it helps him uncover critical issues concerning the learning process that many researchers were unable to explore. Hence, the impact of employing corpora in solving collocation learning difficulties, especially delexical verbs, might be approved. Besides, the study is significant due to the fact that there is a little or no studies have been conducted in the area of using corpora in learning or teaching English as a foreign language for Arab speakers in the Arab world, particularly in Sudan (context of learning and teaching). Furthermore, the study also contributes to the literature of using corpora in learning English language in general and English collocations in particular, so that future researchers and scholars could benefit from it.

1.6 Methodology of the Study

For the enhancement of the research findings, both qualitative (interview) and quantitative (tests and questionnaire) instruments are adopted by the researcher and data will be collected concurrently. So the research design is mixed. Qualitative and quantitative are not opposites as many people might

7

think, instead, they complement each other and strengthen the research findings. Emphasizing this view, Creswell (2009) defined mixed method research as an approach to inquiry that combines or associates both qualitative and quantitative forms of research. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mixing of both approaches in a study.

1.6.1 Population

The population of this study is students from Bayan College of Science and Technology, second year students of Biomedical Engineering, studying English language as a university requirement. In addition to that, the subjects of the interview are experienced English language teachers from different Sudanese universities in Khartoum State.

1.6.2 Sampling

Concerning the sampling procedure, convenience sample is selected. It is sometimes called grasp sample as we essentially grasp members of the population for our sample. It is characterized as being fast, inexpensive and easy. The sample of this study is second year students from Bayan College of Science and Technology and English language teachers from different Sudanese universities.

1.6.3 Tools

Three tools of data collection will be chosen. To begin with, a teacher-designed tests (pre-test and post-test) on collocations (delexical verbs) will be designed by the researcher so as to be suitable with the design of the research. The second instrument is an attitude questionnaire will be designed by the

researcher to evaluate the attitudes of students towards the use of corpora in English language learning. Finally, to fully measure and evaluate employing corpora in solving collocation learning difficulties encountered by university students, university academicians of EFL will also be involved in the study. To do so, an interview about utilizing corpora in English language teaching will be designed and then administered to them. After the construction of the research tools, they will be shown to experts (native speakers if possible) of the field to ensure their validity.

1.7 Limitations of the Study

In fact, the current research is limited to the scope of employing corpora in improving students' collocational competence, especially verb-noun collocations (delexical verbs). It researches, investigates and examines collocational knowledge as an area of difficulty encountered by Sudanese EFL university students. To fully evaluate this, British National Corpus (BNC) will be used to design the tests due to the fact that the corpus is an authentic source of linguistic information and it will also be employed to enlighten students (participants of the experimental group) on how to use corpora in language learning and investigation.

The second limitation is that, the participants of the study are second year students of Biomedical Engineering from Bayan College of Science and Technology, studying English language as a university requirement, Khartoum State. Whereas, the interviewees are expert EFL teachers will be chosen according to their knowledge in the field of corpus linguistics and the willingness to participate in the study. Another limitation is that, this data of this study are collected in 2018.

1.8 Summary of the Chapter

This chapter has provided description of the theoretical framework of the study with special focus on the research problem, hypotheses of the study, and the research methodology. The next chapter will be devoted to review of the previous literature conducted in the area of the study.

Chapter Two

Review of Literature

Chapter Two

Review of Literature

2.0 Overview

This chapter reviews and discusses few previous studies conducted in different places throughout the world in the area of using corpora in English language learning and teaching as well as the difficulties of learning collocations in English language. These studies are reviewed in terms of their similarity or relevance to the current study, so that a global picture of using corpora in language learning and teaching could be provided. The chapter is divided into two main parts: The conceptual framework which is composed of three sections; the concept of collocation, the concept of delexicalization, and corpus. The second part, on the other hand, reviews some previous studies.

2.1 The Theoretical Framework

This part is dedicated to the conceptual background which encompasses collocation, delexicalization, and the concept of corpus.

2.1.1 Collocation

Due to the prime significance of collocations (they are also called recurrent combinations, clusters, etc.) to the current research, this part of the chapter is devoted to the following sub-sections that briefly discuss collocation so that the concept could be clarified.

2.1.1.1 The Concept of Collocation

According to Oxford Online English Dictionary, the word 'collocation' comes from Latin *collocatio* (n-) which means in a technical sense 'to place together', or 'the action of placing things side by side or in position: the collocation of the two pieces'.

According to Oxford Concise Dictionary of Linguistics (2014), collocation is a relation within a syntactic unit between individual lexical elements; e.g. *computer* collocates with *hate* in *my computer hates me*. The term is used especially where words specifically or habitually go together: e.g. *bold* collocates with *hair* in *bold hair* or *their hair is bold*. Hence, of idioms: e.g. *blow* and *top* are part of an especial collocation in *she blew her top* (p. 64).

Collocation is also defined by Oxford Advanced Learner's Dictionary as a combination of words in a language that happens very often and more frequently than would happen by chance (2007: 281). Speaking of the nature of collocations, O'Keeffe et al. (2007) stated that collocations are not absolute or deterministic, but are probabilistic events, resulting from repeated combinations used and encountered by the speakers of any language. For instance, we say *bitterly disappointed* in preference to (but not the absolute prohibition of) *sourly disappointed*; tea is usually *strong* but cars are *powerful*, and so on (p.59).

In the same line of thought, Baker (1992) stated that the patterns of collocations are largely arbitrary and independent of meaning. For example, consider the adjectives *quick* and *fast* which mean the same thing but native speakers do not say *fast snack*s but *quick snack*s (cited in Nasr, 2011, 15).

Researchers seem not to be in agreement when it comes to defining collocation, rather; the term is governed by the discipline (context). For instance, in corpus linguistics, collocation refers to a sequence of words that co-occur more often than would be expected by chance; in phraseology, collocation is a sub-type of phraseme, and so on.

According to the researcher, collocation is a matter of statistical preference that could powerfully be demonstrated by the utilization of corpus data. Hence, the current study employs corpora in an attempt to solve collocation learning difficulties that face EFL/ESL university students.

2.1.1.2 Categorization of Collocations

Collocations have differently been classified into various categories by different scholars and researchers. The commonest categorization is Firthian's (1957) in his semantic theory. According to him, collocations are classified into lexical collocations and grammatical collocations, but they are primarily semantically (not grammatically) based (cited in Bussmann, 1998). Collocations fall into two major groups: grammatical collocations and lexical collocations (Benson et al., 1986b, p. ix). Benson et al. added, in English, as in other languages, there are many fixed, identifiable, non-idiomatic phrases and constructions. Such groups of words are called recurrent combinations, fixed combinations, or collocations. The following table explains some of the commonest language combinations which are considered to be highly required to effective communication in English and native-like language production.

Table 2.1: Definitions of idioms, transitional combinations, lexical collocations, and free combinations plus examples for each.

Classification	Examples
Idioms (= relatively frozen expressions whose meanings do not reflect the meanings of their component parts) (Benson et al., 1986a: 252-53).	-To kick the bucket (to die). -To spill the beans (to tell a secret too soon or to the wrong person).
Transitional combinations (=frozen than ordinary collocations and have a meaning close to their component parts) (Benson, 1986a: 254); transitional collocations (Benson, 1989:5).	-To foot the bill (pay the bill). -To catch one's breath (to return to normal after being busy or very active). -To carry weight (be convincing). -To see action (participate in combat). -To take one's time (act slowly).
Lexical collocations (=arbitrary recurrent word combinations) (Benson, 1986a: 253-54; 1986b: xxiv; 1989: 3).	-To reach a verdict, to commit (a) murder, to launch a missile, to withdraw an offer, to make an impression, a mistake / kind / warm / an appointment, etc.
Free combinations (=the least cohesive type of word combinations) (Benson, 1986a: 252-53; 1986b: ix,xxiv)	-To condemn /denounce / acclaim/ forget / investigate, etc. -To destroy bridge / document / factory / harbor / village, etc.

Source: Adapted from Ali (2017).

The above table shows different types of collocations. Namely, idioms, transitional combinations, lexical collocations and free combinations. These types of language clusters show varying degrees of collocatability. Like proverbs and sayings, idioms are clusters whose meanings cannot be guessed from the meaning of individual words as shown above. For example, the idiom *to kick the bucket* has nothing to do with *to die* if we try to guess the meaning of the individual words. Unlike idioms, transitional combinations are less frozen so that their meaning is closer to their constituents. On the other hand, lexical collocations have lesser degree of cohesion and higher predictability of meaning than idioms and transitional combinations. Unlike the three mentioned recurrent combinations, free combinations are the least cohesive of all and have the highest degree of predicting the meaning from their component parts.

As highlighted by Benson et al. (1986), English collocations are categorized into lexical collocations and grammatical collocations (with 7 and 8 types respectively). To begin with, lexical collocations are characterized by the fact that their meanings mostly reflect the meaning of their lexical constituents. As Benson, Benson and Ilson (1986) put it; lexical collocations are only made up of verbs, adjectives, nouns and verbs in different possible combinations. As demonstrated in the following.

(**1**) verb + noun (investigate a problem).

(**2**) adjective + noun (intolerable situation).

(**3**) noun + verb (volcanoes erupt).

(**4**) noun1 + of + noun2 (a system of life).

(**5**) adverb + adjective (easily acquired).

(**6**) verb + adverb (drive carelessly).

Grammatical collocations, on the other hand, can be defined as phrases made up of a dominant word (verb, noun, and adjective) and a preposition or grammatical structure like infinitive or clause. As shown in Benson, Benson and Ilson (1986) in their introduction to their BBI Combinatory Dictionary of English, grammatical collocations fall into the following combinations: Noun + preposition, noun + to-infinitive, noun + that-clause, preposition + noun, adjective + preposition, predicate adjective + to-infinitive, adjective + that-clause, and the English 19 verb patterns (cited in Ali, 2017).

Table 2.2: Classification of collocations plus examples.

Pattern	Example(s)
V + N/P or prepositional Phrase.	Compose music; set an alarm.
V + Adv	Walk heavily; argue heatedly.
V + N	Make decision; take place/part.
Adj + N	Strong/weak tea.
N + V	Bombs explode; alarms go off.
N1 + of + N2	A pride of lions; a bunch of keys.
Adv + Adj	Quite safe; deeply absorbed.

N + Prep	Ability in/at; kind of; changes in.
N + to + Inf	An attempt to do it; years to come.
N + that-clause	He took an oath that he would do his duty. We reach an agreement that ...; fact that.
Prep + N	On purpose; in fact.
Adj + Prep	Tired of; bored with; angry with/at.
Adj + to + Inf	Ready to go; easy to learn; likely to be.
Adj + that-clause	She was afraid that she would fail the exam; he was delighted that ...
V + Prep	I believe in ...
V + direct O + to + indirect O = V + indirect O + direct O	She sent the book to him. = She sent him the book.
V + direct O + to + indirect O (no movement for dative)	They mentioned the book to her.

V + direct O + for + indirect O =V + indirect O + direct O	She bought a shirt for her husband. =She bought her husband a shirt.
V + Prep + O	They came by train;
V + O + Prep + O	We invited them to the meeting.
V + to Inf	She continued to write.
V + bare Inf	Mary had better go.
V + V-ing	They enjoy watching TV.
V + O + to Inf	We forced them to leave.
V + O + bare Inf	She heard them leave.
V + O + V-ing	He felt his heart beating.
V + a possessive and V-ing	I cannot imagine their stealing apples.
V + that clause(rather uncommon)	The doctor suggests me that I take vitamins.
V + O + to be + C	We consider her to be well-trained.
V + O + C	She dyed her hair red.

V (+ O1) + O2	The teachers asked (the students) questions. It took/cost (us) ten minutes/cents.
V + O + adverbial	You carry yourself well (like a soldier).
V (+O) + wh-clause/wh-phrase	She asked (us) why we had come.
It + V + O + to inf	It surprised me to learn her decision.
It + V + O + that-clause	It surprised me that our offer was rejected.
V + C (Adj or N)	He was a teacher.
V + C (Adj)	The food tastes good.

Source: Based on Benson, Benson, and Ilson's (1986) categorization of collocations.

The table above exemplifies the 37 grammatical collocations which are formed by one of the lexical parts of speech (verbs, nouns or adjectives) plus a grammatical structure or a preposition. In comparison to lexical collocations, grammatical collocations are more confusing for learners, especially those with prepositions. In the context of Sudan for example, learners (native speakers of Arabic) of English find it challenging to master some grammatical collocations due to many factors

such as the interference of their native language (see section 2.2.1), lack of practice, lack of exposure, etc.

2.1.1.3 Collocational Proficiency

Despite the significance of collocational proficiency to EFL/ESL learners, attention is rarely paid to it in language teaching, language learning and language research. This notion is supported by El-Dakhs (2015) stating that, despite its prime importance for language proficiency, collocational competence is one of the most neglected areas in vocabulary studies and second language teaching/learning (p. 68). This negligence of learners' collocational proficiency has resulted in poor knowledge of collocations and consequently poor knowledge of English language. A fact that justifies the need for integrating collocations into foreign/second language research (especially corpus-based research) and curriculum development because this aspect of vocabulary knowledge is crucial to effective and more natural language learning and use. Indicating the significance of collocations to accurate language use, Nation (2001) suggested that word knowledge involves knowledge of word form, word meaning and word use. The aspect of word form refers to the spoken and written forms of words as well as their morphological composition. As for the aspect of word meaning, it entails knowledge of a number of constituents, including reference, sense, connotative meaning and sense relations. The last aspect which is word use, refers to knowing the grammatical, lexical and stylistic constraints of the use of words in the linguistic context (cited in El-Dakhs, 2015, p. 69). Although the third aspect of word knowledge (knowledge of collocates) is highly required to achieving more natural language production, it received the least attention among the three aspects.

In the past three decades, research in the field of corpus linguistics have dominated language research but scant amount of this research has been directed to English language learners' knowledge of collocations. This is considered to be the underlying motive of this study. It is believed by the current research that having solid knowledge of English vocabulary, collocations in particular, requires adopting effective strategies to English language learning and teaching such as the use of corpora because they are helpful in terms of the language data they provide.

2.1.1.4 Difficulties of Collocations to EFL/ESL Learners

It has been proved by findings of the previous studies (section 2.2 below) that collocations are problematic for learners of English language not only for lower level learners, but also for those of advanced levels. These difficulties are attributed to different factors such as the teaching syllabuses, the lack of practice and exposure, etc. This difficulty clearly appears in students' productive skills (when they speak or write). It is assumed by Wray (2002) that non-native speakers strive for achieving native-like production by trying to be idiomatic in their language production. In order to achieve this, they use different strategies such as relying on creativity and making over liberal assumptions about the collocational equivalence of semantically similar items (cited in Salih, 2016, p. 22). Assumptions made by EFL/ESL learners are most likely to be inappropriate and awkward, such as in making assumptions in producing delexical structure. For instance, we say *do housework* but not *make housework*, although the two hot verbs are considered synonymous words by most learners of English. In the German context for instance, Bahns and Eldaw (1990) found that German students' knowledge of collocations had not developed at the same rate as their knowledge of vocabulary in general (cited in Mohammad et al., 2007, p. 39). Nesselhauf (2005) for example, found in her investigation of learners' writing

development that the number of collocational errors was not different when they wrote with or without the use of a dictionary. This may suggest that either the dictionary did not provide the necessary information about the use of collocations or that learners do not seek it since they are not aware of its importance. Another important factor that has been investigated is time pressure. She found that writing with or without time pressure had no great effect on the use of collocations on learners' writing. This suggests that learners' use of collocations demonstrates lack of knowledge rather than lack of control (cited in Alharbi, 2017, p. 24). (For more on collocation difficulties to EFL/ESL learners see 2.2).

Learning vocabulary is generally difficult due to the nature of English language vocabulary. This difficulty is indicated by Nation (2001), highlighting three aspects of word knowledge; knowledge of word form, word meaning, and word use (see section 2.1.1.3). These aspects are projected in the following table.

Table 2.3: What is involved in knowing a word. R= receptive knowledge/ P= productive knowledge.

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 recognised in this word?
		P	What word parts are needed to express the 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 or types of words must we use with this one?
Use	Grammatical function	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, etc.)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Source: Adopted from Nation (2001, p. 27).

As the table above exhibits, there are nine features of word knowledge categorized under three dimensions. These dimensions are form, meaning and use. According to the present study, word use is the most challenging feature among all, especially in terms of what goes with it (its collocates). So it is suggested that corpora are the best strategy that suits the need for mastering this word knowledge for they enable the investigation of word use within context.

2.1.1.5 Underlying Approaches to Collocations

The following sub-sections discuss two prominent approaches to collocations; the frequency-based approach and the phraseological approach and the researcher's

view on collocations. The two approaches are discussed in terms of their relevance to the current study (learning and teaching English language collocations).

2.1.1.5.1 The Frequency-based Approach

The first approach to collocations is the frequency-based approach. It is related to the statistical definition of collocation in which words are collocates if, in a given sample of language, they are found together more often than their individual frequencies would predict (Jones and Sinclair 1974: 19, cited in Salih 2017, p. 17). This approach treats collocations as words that frequently co-occur and the presence of one predicts the presence of the other (s). This notion has firstly been introduced by Firth (1957) in his slogan *the meaning of a word should be known by the company it keep*s. The frequency-based approach has been applied in corpus linguistics to locate collocations (e.g. Siyanova and Schmitt, 2008 & Durrant and Schmitt, 2009). Thus, collocations have been distinguished by their frequency into frequent and infrequent collocations (Salih, 2017). Researchers who adopt the frequency-based approach to collocations have different views as to what constitutes collocations. Moon (1998) considers co-occurrences of all frequencies as collocations while Stubbs (1995) only accepts frequent co-occurrences. Others accept recurrent co-occurrences (more than once) in a given corpus as the defining criterion (Menon et al., 2012). All those researchers considered the semantic relationship between words in determining collocations while neglected the syntactic relationship between them in doing so. According to Nesselhauf (2005), in this approach, the syntactic relationship between the elements does not play a role in deciding whether they form a collocation or not (cited in Menon, et al., 2012). This study adopts the frequency-based approach as it views corpora as the best source for learning collocations for they provide the frequency and co-

occurrence of words in context and they have the potential of doing so because of their inclusion of wide authentic language data.

2.1.1.5.2 The Phraseological Approach

The second approach to collocations is the phraseological approach. Unlike the frequency-based approach which employs corpus frequency as an identification criterion, the phraseological approach uses either native speakers' intuitions (Greenbaum, 1988; Hasselgren, 1994), dictionaries of collocations (Laufer and Waldman, 2011), or a combination of both (Nesselhauf, 2003) in identifying collocations. Collocations are identified in "a scalar analysis, ranging in the form of a continuum from transparent, freely recombinable collocations at one end to unmotivated and formally invariable idioms at the other" (Barfield and Gyllstad, 2009: 6, cited in Salih, 2016, p. 17). According to Nesselhauf (2004), the phraseological approach in which collocation is seen as a type of word combination, most commonly as one that is fixed to some extent (cited in Menon et al., 2012). As previously mentioned, this approach makes use of native speakers' intuitions, dictionaries of collocations, or a combination of both in judging whether something is appropriate in the language or not. Judgments based on either intuitions or collocation dictionaries are not accurate because the former is based on what native speakers think is true, which does not really reflect the actual use of language in real context, while the latter is based on dictionaries which could also be inaccurate due to either the dictionary's inability to include large amount of data (although representative source of information has to include large amount of data) or the fact that most collocation dictionaries are also based on intuitions of their writers. Due to these factors, the frequency-based approach is believed to be more suitable and appropriate (than the phraseological approach) for learning and teaching collocations.

2.1.1.5.3 Collocations in the Current Study: A Complementary Approach

There is a connection and occasional overlap between the frequency-based approach and the phraseological approach. This overlap is shown when advocates of the former (for instance, Benson et al., 1986) consider the latter as an additional defining criterion of collocations and the vice versa. Evert (2008) stresses the close connection and the occasional overlap between the two approaches (cited in Alharbi, 2017, p. 39). As the present study views it, the best way to view the frequency-based approach and the phraseological approach is to considerer them as complementary approaches rather than opposites. So that each can make use of the other in identifying and defining collocations. As stated (in section 2.1.1.5.2) previously, it is hypothesised by the current study that the frequency-based approach is more appropriate for learning collocations, but this approach can be supplemented by the phraseological approach as native speakers and collocation dictionaries could help in judging what is appropriate and what is not in the language.

2.1.1.6 Significance of Collocations to EFL/ESL Learning and Teaching

Collocations are considered to be of a great significance to achieving vocabulary knowledge of EFL/ESL learners and consequently their communicative competence. El-Dakhs (2015) confirmed that it is important to learn collocations due to their frequency of occurrence and their facilitatory effect on language processing (p. 71). Hill (2000) estimates that collocations are found in up to 70% of everything we say, hear, read, or write (p. 53). Similarly, Lewis (2000) considers collocations as the most common and most representative of English

multi-word expressions. A language phenomenon with such high frequency of occurrence deserves considerable attention. Besides, the use of pre-fabricated chunks reduces processing effort and enhances language comprehension and production (cited in El-Dakhs, p. 71). What makes collocations challenging for EFL/ESL learners to fully acquire, is the fact that each and every word in English language has to some extent collocational relationship with other words. This fact urges language learners and teachers to fully consider developing their knowledge of collocations. These facts justify the need to fully integrate and represent these clusters of language into syllabus design, language learning, and language teaching. Similarly, O'Keeffe et al. (2007) asserted that vocabulary syllabus for the basic level is incomplete without due attention being paid to the most frequent chunks, since many of them are as frequent as or more frequent than single items which everyone would agree must be taught (p. 46). O'Keeffe, McCarthy & Carter (2007:60) highlighted that the notion of collocation therefore shifts the emphasis from the single word to pairs of words as integrated chunks of meaning and usage, and collocation has now become an accepted aspect of vocabulary description and pedagogy (e.g. Lewis 2000; McCarthy and O'Dell 2005).

More recently, Salih (2016) maintained that collocational knowledge, which is stored in chunks (Wray, 2002, 2009), is held by some to be the foundation of language learning, use, and knowledge. James (1998) confirmed that adherence to the collocational conventions of a foreign language contributes greatly to one's idiomaticity and native-likeness, and not doing so announces one's foreignness (p. 152). Emphasizing the significance of collocations to all uses of the language, Nation (2001) concludes that collocational knowledge is required to all fluent and appropriate language use (cited in El-Dakhs, 2015). It is assumed that the best way to properly acquire collocations is via utilizing corpora as they provide chunks of

language within context. So they are believed to be effective in learning and teaching collocations (see 2.1.3.7).

2.1.2 Delexicalization

Due to the prime significance of the delexical structure to a successful use of English language, this section is devoted to the concept of delexicalization, common types of delexical verbs, and the significance of delexical structure to English language learning and teaching.

2.1.2.1 The Concept of Delexicalization

Delexicalization which is also known as semantic bleaching refers to a word losing its independent meaning or function, becoming more of a functional word. For instance, *make*, *do*, *take* and *get* in *make money*, *do your best*, *take the initiative* and *get acquainted*. In these contexts, the delexical verbs become more functional than lexical because their meanings are dependent on their collocates.

In a similar context, O'Keeffe, McCarthy and Carter (2007) stated that delexical verbs embrace extremely high frequency verbs such as *do*, *make*, *take* and *get* in their collocates with nouns, prepositional phrases and particles. They added that these verbs are termed delexical because of their low lexical content and the fact that their meanings in context are conditioned by the words they co-occur with (e.g. compare *to make a mistake* with *to make progress* or *to make it*).

Pedagogically speaking, delexical verbs cannot be taught or learnt in isolation away from their collocates, a fact that makes mastering and using such verbs appropriately a difficult one, putting into consideration that these verbs are indispensable part of the daily language use. Therefore, it is hypothesized by the present research that corpora have the potential of learning and teaching delexical

verbs because of the authentic and reliable data they have, so that corpuses would contribute a solution to the difficulties that encounter EFL learners in using these verbs, especially in the Sudanese context.

2.1.2.2 Common Delexical Verbs

Delexical verbs are not so many in English language, the commonest among them are *do*, *make*, *take*, *get*, and *have*. These light verbs present some challenges to English language learners when it comes to what goes with them, especially their right noun collocates i.e. their delexical structure. Most learners of English mix *do* with *make* and *take* with *have*. It is believed that this delexical structure poses some difficulties to EFL/ESL learners because they are given small attention in the process of learning as well as teaching. Similarly, O'Keefe et al. (2007) highlighted that delexical verbs are problematic since they occasionally occur with low-frequency items and learners have difficulty understanding the meaning of the whole expression. Hence, the present study hypothesizes that corpora have the potential of learning this word combination as these verbs require representative source of language data, a source that enables learners to study them within context.

The following table presents some common delexical verbs with examples and explanations.

Table 2. 4: Some collocations of do, make, take, and have.

Do	Make	Take	Have	Notes
		take action		To *take action* means to **act positively** and **decisively** to resolve a problem. (light)
		take advantage, take advantage of	have an advantage	To *take advantage* of something or someone means to **profit** from the momentary situation of the person or thing. To *have an advantage* means to **be in a better position** than someone else.
	make amends			To *make amends* means to **compensate for a loss** or **injury**.
	make an argument		have an argument	To *make an argument* for something means to **present a logical argument** in its favour. To *have an argument* with someone means **to argue** with that person. (light)
	make arrangements			To *make arrangements* means to **organise** or **arrange**. (light)
	make an			To *make an attempt* means to

	attempt			**try**, to **attempt**. (light)
do the dishes				To *do the dishes* is to **wash dirty dishes**, **cutlery** and so on by hand.
do business	make business	take business	have business	To *do business* means to **agree a commercial contract**. To *make business* is to **create a business opportunity**. To *take business* is to **deprive a competitor** of a customer by acquiring that customer. To *have business* with someone means to **have matters to be settled** with that person.

Source: Adopted from Wiktionary, 2018 (online)

The above table projects different uses for the delexical verbs *do*, *make*, *take* and *have*. A single (hot) verb can have multiple meanings depending on the noun that follows as shown above. Such structure requires corpus data so that learners can get access to the data and see how this combination behaves and consequently develop their autonomous learning.

2.1.2.3 Significance of Delexical Structure to EFL Learning

The idea that we best know the meaning of a word, not by examining it in isolation, but by the company that it keeps, is usually ascribed to Firth (1957) who describes the ways in which meanings are often created by the associations of words rather than by individual words (Greaves & Warren, 2011, p. 212). An important type of collocation is delexical structure. It is believed to be one of the

commonest structures (clusters/collocations/combinations, etc.) in English language. This is due to the fact that although delexical *(thin, vector, empty*, *light*, etc.) verbs are not so many in English language but they are considered to be core components to the delexical structures. So enhancing this type of collocation contributes to the mastery of vocabulary and consequently contributes to the fluency of English language because this type of combination is an important part of the daily language. O'Keefe et al. (2007) put delexical verbs in one of the broad categories of basic vocabulary according to their high frequencies in some corpora. This clearly demonstrates the fact that this type of collocations (delexical structure) is significant to language learning due the high frequency of the structure they combine with.

It is also assumed that learning the language in chunks helps learners master the language as these language units help them produce natural and native-like language. In this concern, Kashiha and Chan (2014) confirmed that studying single words was important in the process of building a firm base for language learners. Learners, however, at the higher level of EFL are involved with units consisting of more than one word such as phrases, idioms, collocations or chunks. They added that collocations are currently one of the most popular topics amongst language researchers who often use language corpora as a statistical tool to check how collocations are employed in authentic language.

2.1.3 The Concept of Corpus

As defined by Oxford Concise Dictionary of Linguistics (2014), corpus is any systematic collection of speech or writing in a language or variety of a language. Thus, in particular, of large online collections, tagged and searchable for research purpose (p. 82).

Another definition is provided by O'Keeffe, McCarthy and Carter, according to them, a corpus is a collection of texts, written or spoken, which is stored on a computer. In the past, the term was more associated with a body of work. For example, all the writings of one author (2007, p.1).

Similarly, Oxford Advanced Learner's Dictionary (2014) defines corpus as a collection of written or spoken texts. From the linguistic perspective, a corpus is a large amount of language data stored on a computer for the purpose of linguistic analysis. A general definition of corpus is provided by The Expert Advisory Group on Language Engineering Standards (EAGLES), saying that "it can potentially contain any text type, including not only prose, newspapers, as well as poetry, drama, etc., but also word lists, dictionaries, etc." (cited in Meyer, 2004, p. xi). According to this definition, a collection of any language meant to represent something would constitute a corpus.

According to Bussmann (2006), corpus is a finite set of concrete linguistic utterances that serves as an empirical basis for linguistic research. The value and quality of the corpus depend largely upon the specific approach and methodology of the theoretical framework of the given study. Note, for example, the different value placed on empirical data in structuralism and in generative grammar (p, 260).

According to the researcher, corpus can be defined as a collection of wide range of linguistic data collected from a wide range of registers/contexts meant to represent a language or part of a language. This collection of data is used for language analysis and investigation. It is believed that corpus linguistics has gained popularity after the arrival of computers due to the fact that computers have the ability to store millions of texts and the ability to analyze these texts using software programs. Because of the advent of computers, these huge amounts of information

are ready for quantitative and qualitative analysis of the language. This claim is supported by Bibber, Conard and Reppen (1998) claiming that: a corpus is a principled collection of texts available for quantitative and qualitative analysis.

Recently, a corpus has come to be regarded as a body of text made available in computer-readable form for purposes of linguistic analysis (Meyer, 2004). In the same context, Crystal (2008) added that, computer corpus is a large body of machine-readable text. Increasingly, large corpora (especially of English) have been compiled since the 1980s, and are used both in the development of natural language processing software and in such applications as lexicography, speech recognition, and machine translation.

2.1.3.1 Corpus and Corpus Linguistics

As discussed (in section 2.1.3) above, corpus (the plural is corpora or corpuses) refers to any collection of language data meant to represent language or part of a language. This language data is used for linguistic analysis and investigation. The useful information we currently find in dictionaries (frequency of a word, collocations, and authentic examples of the word in context) is greatly due to the development of corpus linguistics (Badea, 2015: 305). Whereas, corpus linguistics is a branch of linguistics concerned with the scientific study of that collection of language. According to McCarthy & O'Keeffe, linguists often generally see corpus linguistics as a "method of exegesis based on detailed searches for words and phrases in multiple contexts across large amounts of text" (2010, p. 3, cited in Badea, 2010, p. 305). The fact that whether corpus linguistics is a methodology or approach is controversial issue among linguists, but in fact it is a methodology that helps in the analysis of language.

2.1.3.2 Types of Corpora

There are many different kinds of corpora available for different uses and purposes. They can contain written or spoken (transcribed) language, old texts or modern texts, and they can also contain texts of whole books, newspapers, etc. or parts of texts. Which type of corpus we are investigating depends mainly on the purpose (s) of creating that corpus. Corpora can be balanced/representative, monitor, parallel, comparable, diachronic, specialized, multimedia, etcetera. This part discusses the common different types of existing corpora with some examples so that newcomers to the field of corpus linguistics can have a quick overview of the types of corpora available. To begin with, corpora can be general or sometimes called balanced corpora. A good example of general corpora is the British National Corpus (for more on BNC see section 2.1.3.3 below), a corpus composed of a large number of language varieties. On the other hand, there is specialized corpora (composed of specific text type), designed to represent a small set of language and they are small in amount in comparison to general corpora. According to Lee (2011), specialized corpora are usually smaller in scale than general language corpora precisely because of their narrower focus. This is not a problem, however, as the greater homogeneity among texts in a specialized area confers the advantage of fewer texts being required for the corpus to be representative of that language variety (p. 114). Many corpora have been created to serve special purpose (s), a good example of that is Michigan Corpus of Academic Spoken English (MICASE). A 1.8-million words corpus of spoken (transcribed) English compiled from about 190 hours of various speech events recordings in a North American university. MICASE is designed to serve the needs of practitioners of English for Academic Purposes. One good purpose that specialized corpora serve is that they allow researchers to examine differences across national contexts. Furthermore,

another corpus type which is designed to serve the needs of learners is called learner corpus. As shown in Lee (2011), these corpora represent the general linguistic proficiency of university-level learners of English, but in terms of research writing, where the writing is done for the communication of disciplinary content or research rather than for the sake of a language proficiency class or test (p. 116). The commonest examples of learner corpora are International Corpus of Learner English (ICLE), consists of 3.7 million words of written language, mainly composed of argumentative essays from 116 graduate and undergraduate students of English language from more than twenty different language backgrounds. Another example of learner corpora is Louvain International Database of Spoken English Language (LINDSEI), contains transcripts of two types of speech by speakers from eleven different L1 backgrounds.

Some corpora cover English of early periods and have been used to study the changes in the language. Other historical corpora focus on specific historical period or a specific genre of language. Lee (2011: 113), demonstrated that there are three main collections of historical English that cover a wide span of time and genres: the diachronic part of the Helsinki Corpus of English, ARCHER (A Representative Corpus of Historical English Registers), and COHA (Corpus of Historical American English). The Helsinki Corpus (1.6 million words) covers the period from around 750 to 1700, and thus spans old English (413,300 words), middle English (608,600 words) and early modern (British) English (551,000 words). ARCHER is a multi-genre corpus (currently 1.8 million words) covering the early modern English period right up to the present (1650-1990) for both British and American English. Examples of historical corpora that focus on specific time period are; The Lambert Corpus which is composed of around one million words of literary prose covering the period from 1640 – 1740 (a period of one hundred years). Another example is the Corpus of Early English

Correspondence (CEEC), a 2.7-word corpus covering the period from 1417 to 1681.

Finally, parallel and comparable multilingual corpora are also different types of corpora. Translation (parallel) corpora contain original texts and their translations into one or more other languages, while comparable corpora contain original texts in two or more languages or varieties of language matched by criteria such as genre, time of publication, etc. (Johansson, 2007, cited in Lee, 2011). Both kinds of corpora are useful in making comparative and translation studies as well as in lexicography (in designing corpus-based dictionaries). In parallel corpora, the two components are aligned on a paragraph-to-paragraph or sentence-to-sentence basis. The English-Norwegian Parallel Corpus (ENPC) and English-Swedish Parallel Corpus (ESPC) are good examples of a parallel bidirectional corpus: each corpus has four related components, allowing for various types of comparison to be carried out (Lee, 2011: 119). On the other side, there are very few examples of comparable corpora, the commonest one is the International Corpus of English (ICE), containing around million words texts collected from different varieties of English around the world.

This part has provided a short survey of types of corpora with the intention of providing the audience with some knowledge of the categorization of corpora. The classification provided above is some of many and some corpora may find their way to come under more than one category.

2.1.3.3 Existing Corpora

There is a wide range of language corpora available, some of these corpora are freely available, some can be bought (commercial corpora), and some are not publically available.   As shown in Meyer (2004) and O'Keeffe (2007), the

following are some examples of available corpora that can be harnessed by language learners, teachers, and other users of the language.

To begin with, the most outstanding and commonly used corpus is the British National Corpus (BNC) which is created by Lancaster University, Oxford University and the British Library. It contains 100 million words of written (such as newspapers, essays, periodicals and journals, books, letters and memoranda, etc.) English that constitutes 90% of the data, as well as spoken English (such as conversations, radio shows, government meetings, discussions, etc.) that represents 10% of the corpus. The BNC was built from a wide range of sources to represent the British variety in the late 20th century. In contrast to the BNC, there is the American National Corpus (ANC) which is very small in comparison to BNC, containing only 22 million words of English, 83% written data such as books, newspapers, internet postings, magazines, travel guides and many more. And 17% for the spoken data such as lectures, phone calls, seminars, narratives, and many more. Unlike BNC, ANC was built to represent the American English, and it is annotated for lemma and parts of speech and includes genres such as tweets, emails, etc.

Secondly, there is the International Corpus of English (ICE) began in 1990 basically aiming to collect material for comparative studies of English worldwide. Each ICE consists of one million words of spoken and written English produced after 1989. To ensure compatibility among the component corpora, each team is following a common corpus design, as well as a common scheme for grammatical annotation. Examples of ICE are ICE Great Britain, ICE East Africa, ICE Ireland, ICE India, ICE Singapore, ICE New Zealand, and ICE Philippines. Similarly, there is Cambridge International Corpus which is one billion words of British and American English. The British variety includes 450 million written, 17 million

spoken including the CANCODE corpus, 20 million written academic, 30 million written business, and 1 million spoken business (CANBEC). On the other side, the American variety includes 200 million written, 22 million spoken (Cambridge Cornell Corpus of Spoken North American English), 7 million written academic, and 30 million written business. In addition to learner English that includes 19 million learners' written English (the Cambridge Learner Corpus) and 12 million error coded learner written English.

Thirdly, the Brown Corpus which is 1 million words corpus of American English texts compiled by Henry Kucera and W. Nelson Francis at Brown University in the 1960s. It consists of texts samples from 15 different text categories. The Brown Corpus of Standard American English was the first of the modern, computer-readable, general corpora. Fourthly, the Macmillan World English Corpus consists of over 220 million words of spoken and written mostly British and American English. The ratio is about 90% written and 10% spoken, and the sources include academic discourse, print and broadcast journalism, fiction, recorded conversations (including telephone calls), recorded business meetings, general non-fiction, emails, legal texts, radio documents, broadcast interviews, ELT course books, texts written by learners of English, essays and examination scripts, etc.

Fifthly, International Corpus of Learner English (ICLE), a 2-million corpus of argumentative essays of different topics written by advanced non-native speaker learners of English from different linguistic backgrounds such as Arabic, Russian, German, French, Bulgarian, Dutch, Spanish, Italian, Swedish, Finnish, etc. It was launched by Sylviane Granger at University of Louvain-la Neuve, Belgium in 1990. Then the corpus was released in 2002 in CD-ROM format accompanied by a

handbook that describes the structure and the status of English in the countries of origins of the learners.

Finally, there is Bank of English Corpus which is a representative subset of COBUILD corpus that consists of 415 million words of speech and writing, mainly British English but content of South Africa, New Zealand, Australia, North America and other Commonwealth countries is also included in the corpus. The vast majority of the texts are from written English compiled from sources such as newspapers, magazines, books and websites. Whereas, there is also a large amount of spoken data collected from informal conversations, radio broadcasts, etc. and texts are continually added. The copies of the corpus are held both at University of Birmingham (this version can be accessed for academic research) and HarperCollins Publishers.

In summary, this is few number of the available English language corpora (designed for different purposes) that can be used in language learning, teaching and investigation. They are presented in a simple way so that a clear background about existing corpora could be provided.

2.1.3.4 Tools and Techniques for Using Corpora

There is a number of basic corpus linguistic techniques that can be used on a corpus using standard software such as *Wordsmith Tools* (1999) and *Monoconc Pro* (2000) (O'Keeffe et al., 2007). The following are some of the fundamental corpus techniques that can be utilized to exploit corpora.

2.1.3.4.1 Concordancing

Concordancing is one of the most effective and commonly used corpus techniques. It is used for identifying every occurrence of a particular word or

phrase in a corpus via software. As shown in O'Keeffe, McCarthy and Carter (2007), the search word or phrase is often referred to as the *node* and concordance lines are usually presented with the node word/phrase in the centre of the line with seven or eight words presented at either side. These are known as Key-Word-In-Context displays (or KWIC concordances). In one common type of concordancing display, each occurrence of the chosen word is presented on a single line, with the word in the middle and context on each side. These displays are referred to as KWIC displays or Key Word in Context (Bibber, Conard and Reppen, 1998). From these listings, it is possible to see the different meanings of words in different contexts of language use. In addition to that, concordance lines help develop learners' autonomy as learners can surf the internet for new senses and uses of words rather than relying on ready-made sentences and expressions in textbooks.

One of the disadvantages of using KWIC displays is the large amount of search results learners can get from a corpus, but this challenge can be minimized by the teacher telling learners what exactly do they have to get from such search.

Concordance lines can provide many examples of patterns of use that is why they are applied to language classroom and are now being used in ELT materials (O'Keeffe et al., 2007). So concordance lines are believed to be effective in investigating collocations (especially delexical verbs) as they provide contexts and occurrences of words/phrases in a corpus. When making concordancing search, the search word is shown in the middle of the box surrounded by contexts (collocates) on both sides (left and right) as projected in figure 2.1 below.

Figure 2.1: Shows concordance lines for *way* from Limerick Corpus of Irish English (LCIE).
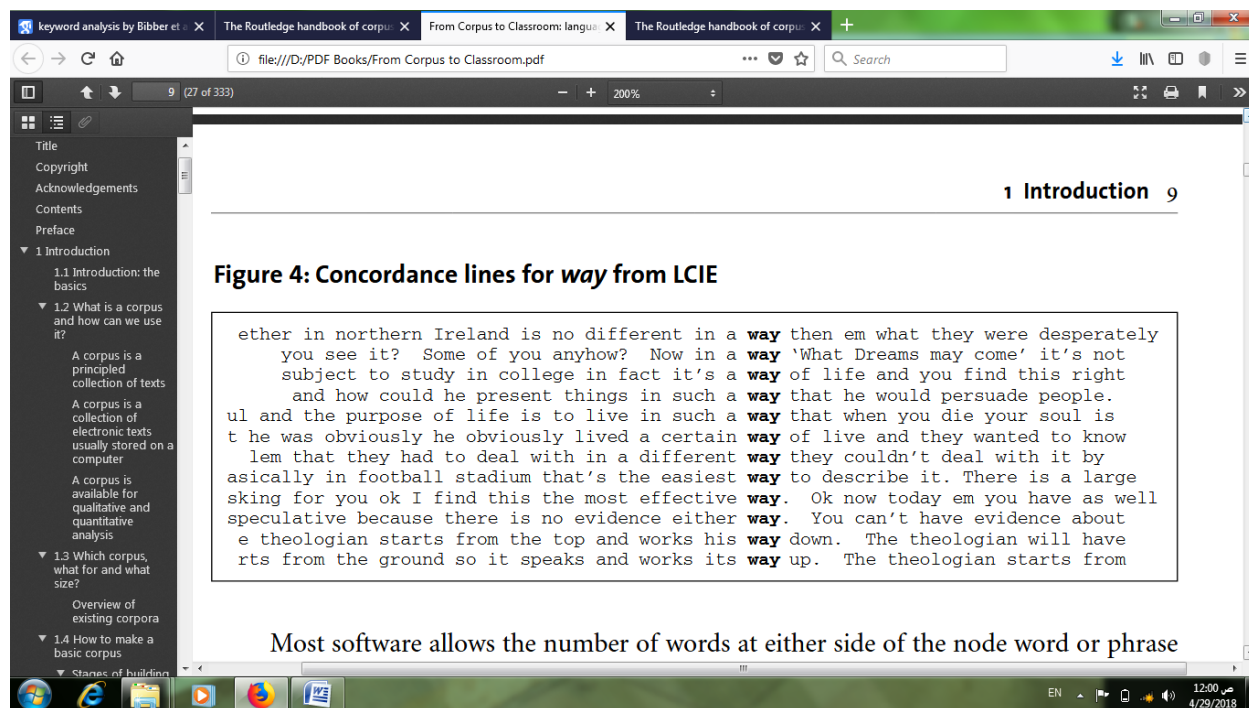
Source: From O'Keeffe et al. (2007).

2.1.3.4.2 Wordlists or Word Frequency Counts

Another commonly used corpus software is word frequency counts or wordlists. It is used to quickly and easily calculate the number of occurrences of words in a corpus. As demonstrated in O'Keeffe et al. (2007), by running a word frequency list on your corpus, you can get a rank ordering of all the words in it in order of frequency. This function facilitates enquiry across different corpora, different language varieties and different contexts of use. Wordlists can be generated in alphabetical order (as shown in table 2.2 below)

Figure 2.2: Word frequency lists generated in alphabetical order.

Table 10.2 Frequency list (alphabetical)

N	Token	Freq.	%
265	huh	14	0.33
266	hum	4	0.09
267	hundred	1	0.02
268	hundreds	2	0.05
269	I	129	3.02
270	I'd	5	0.12
271	I'm	6	0.14
272	idea	10	0.23
273	if	15	0.35
274	imagine	1	0.02

Source: From Evison (2010).

or in order of frequency (from the more frequent to the less frequent) using a computer program called TACT (Text-Analysis Computing Tool). When investigating the frequency of a word in a corpus, it is often useful to collectively consider the different forms of the word, and to distinguish between these forms, the term LEMMA is used to refer to the base form of the word (written in small capital letters) and grammatically annotated corpora are used to provide information about each individual word such as word class and so forth. Bibber et al., (1998: 30), stated that, many words in English have multiple grammatical functions. For instance, the word *deals* can either be used as a third person singular verb or a plural noun. Similarly, both deal and dealing can be used as a verb or a noun. When the words in a corpus have not been annotated for their grammatical category, the corpus is referred to as "untagged". Frequency lists based on an

43

untagged corpus are limited in usefulness, because they do not tell us which grammatical uses are common or rare. This technique is very useful in the analysis of word meanings and functions. An example of frequency counts is shown below (figure 2.3).

Figure 2.3: Word frequency lists generated in rank order (in LIBEL corpus of spoken academic English).

Table 10.1 Frequency list (rank order)

N	Token	Freq.	%
1	the	203	4.76
2	I	129	3.02
3	a	116	2.72
4	and	109	2.55
5	it	89	2.09
6	to	86	2.02
7	think	81	1.9
8	of	80	1.87
9	you	78	1.83
10	yeah	76	1.78

Source: From Evison (2010).

2.1.3.4.3 Key Word Analysis

This function allows us to identify the key words in one or more texts. Key words, as detailed by Scott (1999), are those whose frequency is unusually high in comparison with some norm. Key words are not usually the most frequent words in a text (or collection of texts), rather; they are the more 'unusually frequent' (cited in O'Keeffe et al., 2007: 12). In key word analysis, two pre-existing word lists are

compared and one of these is assumed to be a large word list which will act as a reference file or benchmark corpus. The other is the word list based on the text (s) which you want to investigate. The larger corpus will provide background data for reference comparison (O'Keeffe, et al., 2007). For example, as shown in figure 2.3 displayed above that *the* is the most frequent word in the LIBEL corpus of spoken academic English; if we select one economics lecture from this corpus and generate a word list, we can also see that *the* is again the most frequent word. However, if we compare this economics lecture word list with the larger one from the LIBEL corpus using keyword software (such as that found in Wordsmith Tools), it will tell us which words occur with unusual frequency, or 'keyness'. These words are then referred to as the key words. An example of key word analysis is projected below.

Table 2.5: Key words from an economics lecture relative to a general corpus of academic lectures.

1	tax	15	higher
2	income	16	percent
3	system (s)	17	rates
4	average	18	ordinary
5	basic	19	sixty
6	rate	20	marginal
7	supply	21	scheme
8	poor	22	labour
9	thousand	23	terms
10	impact	24	cost (s)
11	equity	25	characterized

12	under	26	workers
13	both	27	system
14	figures	28	negatives

***Source*:** From O'Keeffe et al. (2007).

2.1.3.4.4 Cluster Analysis

As shown in O'Keeffe, McCarthy and Carter (2007), the analysis of how language systematically clusters into combinations of words or 'chunks' (e.g. *I mean, this*, *that and the other*, etc.) can give insights into how we describe the vocabulary of a language (p. 13). Such kind of cluster analysis can also have implications for what vocabulary we teach and how the task of acquiring vocabulary and developing fluency can be approached by learners. The process of generating cluster lists or chunks is a corpus technique similar to making single word lists. Instead of asking the computer to rank all of the single words in the corpus in order of frequency, we can ask it to look for word combinations, for example 2-, 3-, 4-, 5-, or 6-word combinations (O'Keeffe et al., 2007). For the purpose of the explanation, the following table shows the 20 most frequent 3-word combinations from 10 million words (five million written and five million spoken) of the Cambridge International Corpus (CIC) using Wordsmith Tools.

Table 2.6: The 20 most frequent three-word chunks in 10 million words from CIC.

	Chunk	Frequency per million words		Chunk	Frequency per million words
1	I don't know	588	11	a couple of	166
2	a lot of	364	12	do you want	159
3	one of the	320	13	you have to	158
4	I don't think	248	14	be able to	157
5	it was a	240	15	a bit of	155
6	I mean I	220	16	you want to	153
7	the end of	198	17	and it was	148
8	there was a	193	18	it would be	142
9	out of the	190	19	do you know	138
10	do you think	177	20	you know what	137

Chapter 3 looks in detail at chunks in spoken and written corpora and at the peda-

Source: From O'Keeffe et al. (2007).

2.1.3.5 How to Build a Corpus

There is a wide number of corpora available in the web for use and the number is still increasing every year. So before you decide to build a corpus, make sure that there is no available corpus that meets your needs or these needs are not adequately represented by existing corpora. In this case you need to create a corpus that suits these needs. Building a corpus can be relatively easy or difficult task depending on your needs and the nature of research questions posed. According to Reppen (2011), depending on the types of research questions being addressed, the task of constructing a corpus can be a reasonably efficient and constrained task, or it can be quite a time-consuming task. Having clearly articulated question (s) is an essential first step in corpus construction since this will guide the design of the

corpus. As demonstrated in Bibber, Conard and Reppen (1998), a corpus is not simply a collection of texts. Rather; a corpus seeks to represent a language or some part of a language (p. 246). They also added that, the appropriate design for a corpus therefore depends upon what it is meant to represent. One of the most essential requirements in designing a corpus is being representative, i.e. the corpus must represent the language or variety of a language meant to be represented. For instance, in designing a corpus of Sudanese colloquial women language, data from different sources must be collected such as coffee chats (the gathering of women for coffee), phone calls, etc., so that the corpus can fully represent the colloquial language spoken by women in Sudan. There must be a match between the language being examined and the type of material being collected (Bibber 1993, cited in Reppen, 2010, p. 31). Another issue which highly needs to be considered is corpus size. The size of a corpus is also based on what needs to be represented. For example, in designing a corpus that represents works written by al-Tayeb Salih, the works written by him are not so many so that a small-size corpus can fully represent them. Unlike designing a corpus that represents the Sudanese colloquial women language as discussed earlier which is much bigger and needs to capture all the colloquial language spoken by women in Sudan.

As corpus data can be compiled from spoken or written language sources, building a corpus is also decided by which corpus (written or spoken) we need to design. To begin with, building a spoken corpus is much more daunting and time-consuming task than building written one because of many factors, prime among them is the material being collected which has to be transcribed (which requires long time) so as to be inserted in the corpus. According to O'Keeffe et al., it is useful to keep in mind that one hour of continuous every day, informal conversation yields approximately 12,000 to 15,000 words (2007, p. 5). For this

small amount of spoken data to be transcribed it requires days of hard work. This is plus ethical considerations such as getting permission from speakers and so on. On the other side, designing a written corpus is quicker and easier than the spoken one. As shown above, designing a corpus is based on what is meant to be represented. It is very important to carefully plan the construction of a corpus: what size it will be, what types of texts will be included in it, and what population to be sampled to supply the texts that will comprise the corpus (Meyer, 2004).

What makes compiling written corpora easier and faster than assembling a spoken one is the availability of data. Written data can be downloaded from the internet, can easily be typed, or can be available in electronic format so that it can easily be inserted in a corpus. Concerning database texts, any individual text in a corpus needs to be traceable to its source information (that is, who wrote it, where and when it was published, genre, number of words and so on, especially for purposes of subsequent use in relation to copyright). This information can be stored at the beginning of each file (as 'header information') or in a separate database (O'Keeffe et al., 2007, p. 8).

2.1.3.6 Corpora and Language Learning and Teaching

It is of no doubt that language learning and teaching have been revolutionized by corpus linguistics for the last three decades or so in a variety of different ways. According to O'Keeffe and McCarthy (2010), corpora, corpus-analytic tools and corpus evidence have been increasingly used in English language teaching and learning for the last two decades (see, for example, Sinclair 1987, 1991, and 2004). Language corpora provide systematic access to naturally occurring language, and corpus-linguistic methods support exploratory and discovery learning (Bernardini 2004), which encourages autonomous learning and teaching (Braun 2005, cited in

O'Keeffe and McCarthy, 2010, p. 320). Corpora have also made available softwares and techniques such as concordance lines, Data Driven Learning (DDL), and so on. In addition to that, language teaching and learning have indirectly benefited from corpora. This is clearly reflected in areas such as lexicography as dictionaries are mainly based on corpora, especially those designed by major publishers, as well as teaching materials which are either corpus-based or corpus-informed, and many other areas of language learning and teaching.

A useful and clear distinction between direct and indirect use of corpora in language learning and teaching is made by Leech (1997). These applications are broadly explained in the following sections.

Figure 2.4: Direct and indirect applications of corpora in language learning and teaching.

GIRGIN.pdf   ✕   Corpora and language teaching.pdf ✕   +

← → C ⌂   ⓘ file:///D:/My Ph.D project/PDF files/Corpora and language teaching.pdf   ••• ▼ ☆   🔍 Search   ⦚⦚\ ⬚ ● ≡

□   ↑ ↓   113 (2 of 20)   − +   180%   ⬍   ⤢ 🖨 🔖 »

the use of corpora in language learning and teaching

indirect applications: hands-on for researchers and materials writers

direct applications: hands-on for teachers and learners (data-driven learning, DDL)

effects on the teaching syllabus

effects on reference works and teaching materials

teacher-corpus interaction

learner-corpus interaction

Fig. 7.2: Applications of corpora in language teaching

EN ▲ 🏳 🔋 📶 🔊   11:46 ص
4/27/2018

Source: From Leech (1997).

50

The use of corpora in language learning and teaching has been categorized into direct and indirect as demonstrated above. The direct use of corpora involves directly using them by learners and teachers in language teaching (LT) classrooms such as the use of data-driven learning (DDL) and corpus tools and methods that assist in teaching and learning the language. The indirect use of corpora in LT classrooms involves exploiting corpora by researchers and materials writers so as to develop the learning and teaching processes and even sometimes corpora help in decision making concerning what to teach and how to teach it.

2.1.3.6.1 Direct Applications of Corpora in Language Learning and Teaching

While the indirect approach centres on the impact of corpus evidence on syllabus design or teaching materials and is concerned with corpus access by researchers and though to a lesser extent materials designers, the direct approach is more teacher- and learner-focused. Instead of having to rely on the researcher as mediator and provider of corpus-based materials, language learners and teachers get their hands on corpora and concordancers themselves and find out about language patterning and the behavior of words and phrases in an "autonomous" way (Bernardini 2002, p. 165). According to Breyer (2011), corpus linguistics has not only made indirect contributions to the field of language teaching but from an early stage, language practitioners and applied linguists have recognized the potential of employing corpus resources and tools directly in the classroom. The 1980s has witnessed the transition from powerful and expensive mainframe computers to more affordable micro computers that provided researchers and teachers with ready access to computers for the first time. It was also a time in which the debate on the role of the computer in teaching emerged and caused considerable concern among practitioners (Breyer, 2011). Since that time, corpus

linguistics has become indispensable part of the processes of teaching and learning of languages as language learners, teachers and other practitioners have constant access to corpora. A fact which is clearly shown by the use of some of the corpus-based techniques and tools such as concordance lines (see section 2.1.3.4.1) and Data-Driven Learning (DDL), a method of learner-focused activities with a teacher as facilitator of the learning process.

2.1.3.6.2 Indirect Applications of Corpora in Language Learning and Teaching

As McEnery, Xiao, & Tono (2006) claimed, even if most language teachers and learners have not heard of a corpus, they have been using the products of many corpus-based studies (cited in Girgin, 2011, p. 18). Even though most language teachers are not aware of what corpus is and how it can raise their awareness to provide meaningful input to language learners, corpus can have some crucial effects on the design of dictionaries, textbooks, course books, and grammar books, all of which are used in language classes. Language learning can be informed by linguistic descriptions that rely on the results that a corpus-based study reveals and benefits from these results to provide input for language learners (Barlow, 1996). This proved the fact that corpora have revolutionized language teaching and learning in various ways because of the real and authentic language data they have. In this concern, Sinclair (2004) added that, previous pedagogical descriptions which had not emerged from a corpus-based investigation can be evaluated in the light of 'new evidence', which is provided by corpus-based investigation (p. 271). Observations derived from corpus research have impacted on language description as evidenced in dictionaries and reference grammar as well as corpus studies have impacted on pedagogical textbooks and syllabus design. The selection of language features, the order of which to teach them in, and how to teach them has also been

influenced by research on corpus linguistics (Breyer, 2011). We can say that corpus linguistics has gone so far by not only impacting the designation of teaching materials and syllabus design but also the way these materials are to be taught. To measure to what extent corpora are effective in language teaching and learning, many studies have been conducted. For example Bibber and Reppen (2002) investigated aspects of materials development for grammar instruction by comparing English as a second language (ESL) and EFL materials with results from empirical corpus studies. They come to the conclusion that there are often sharp contrasts between the information found in grammar materials and what learners encounter in the real world of language use. They argue that a revision of grammar references based on corpus investigations of actual language use could help to improve the language learning process (cited in Breyer, 2011, p. 46). (See, for example, Mindt, 1986; Schlüter, 2002; Frazier, 2003; Römer, 2004; and Conrad, 2004).

2.1.3.7 Corpora and Collocation Learning and Teaching

Recently, language learning and teaching, especially collocations, have received great attention from scholars and researchers in the field of corpus linguistics, but scant attention has been given to the direct use of corpora in developing learners' collocational competence in the classroom setting such as using concordance lines data to assist the learning and teaching of collocations. According to Li (2017), it is rare to observe direct uses of corpora in a course to develop learners' collocational competence, probably because it presents teachers with several practical obstacles (p. 154). These obstacles could be the large amount of data learners might get when searching the corpus and the possible difficulty of the language of corpora, especially to lower level learners as well as the time teachers spend to familiarize their students with corpus use. Regardless of all these challenges of corpus use,

corpus assistance to language learners and teachers is undeniable. Such as helping students become autonomous learners by accessing corpora and practically using them, the accessibility of authentic language that learners can exploit, and many other advantages. Wu et al., (2016) added that, corpus use not only facilitates learning and writing, but also arouses learners' awareness of collocations and increases their confidence in language use (p. 20). In a similar context, Vyatkina (2016) confirmed that, paper-based DDL materials were more effective than traditional methods for teaching new collocations to students at lower-proficiency levels (cited in Li, 2017, p. 154). In addition, the researcher believes that it is easy for language learners and teachers to get access to corpora via their smart phones anywhere anytime so that the process of learning is not restricted to the classroom and learners can maintain learning as long as it is possible for them to regularly and constantly exploit corpora.

The task of integrating corpora into language curriculum is to some extent challenging. In this concern, Tribble (2015) reported that user-friendliness and no-cost access are major factors that hinder the application of corpus tools. What learners need is a tool designed specifically for language pedagogy, with a user-friendly interface that requires minimal typing and clicking and straightforward search functions that require little linguistic and metalinguistic knowledge (cited in Wu et al., 2016, p. 21). Most of the studies that employed corpora in language learning and teaching involved short-term experiments but the current one is trying to adopt corpora as part of the curriculum so that learners as well as teachers could frequently get access to them and practically use them.

2.1.3.8 Applications of Corpora

Beyond language use and description, corpora cover wide range of applications and fields such as lexicography, sociolinguistics, grammar, translation, stylistics,

forensic linguistics, and almost every field in which language is a central component by way or another benefit from corpora. In the beginning, corpora are now standard tool for lexicographers, who use multi-million word corpora to examine word frequency, patterning, and semantics in the compilation of dictionaries (O'Keeffe et al., 2007, p. 17). Rather than designing dictionaries based on intuitions, corpus-based and up to date dictionaries are now produced by many outstanding publishers such as Collins Birmingham University International Database (COBUILD) dictionary project, directed by John Sinclair at the University of Birmingham in 1980. Secondly, since corpora consist of real and natural language that includes all sorts of written and spoken language, they provide valuable insights to the area of sociolinguistics which investigates how language varies from social group to another and from geographical area to another. According to O'Keeffe and McCarthy (2010), since the early days of corpus linguistics, there has been an interest in using corpora as a means of exploring linguistic variation, to study for example differences between regional varieties of English. This was the motivation for the compilation of the first-generation written corpora, the Brown and Lancaster– Oslo/Bergen corpora, whose parallel structure and content enabled reliable comparative studies of high-frequency lexical items and grammar. In addition to that, as lexicography, grammar is greatly influenced by corpus linguistics. The valuable contribution of corpus linguistics to grammar is that it increases researchers' ability to systematically study the variation in a large collection of texts produced by far more speakers and writers, and covering a far greater number of words than could be analysed by hand (Conard, 2011, p. 228). Major grammar of English are now corpus-based, corpus-informed or corpus-oriented (for example, Quirk et al. 1985; Bibber et al. 1990; Sinclair 1990; Carter and McCarthy 2006).

Furthermore, stylistic linguists are interested in the styles of different genres of the language such as newspapers style, the style of radio broadcasts, the language used by diplomats and other genres of the language. For stylisticians, corpora are considered to be important source of information. To conclude, another area which is concerned with using language in crime investigation and law is called forensic linguistics. The influence of corpora in this area includes the analysis of documents from confessions to suicide notes, authorship identification in academic settings (e.g. issues of plagiarism), ransom notes, threat letters, readability/comprehensibility of legal language, forensic phonetics (e.g. speaker identification), police interview and interrogation data, as well as the discourse of the courtroom setting (O'Keeffe et al. 2007).

This part has dealt with the theoretical framework; specifically the concepts of collocation, delexicalization and corpus are elaborately explained. The description is dedicated to clarifying the concepts, the types/taxonomy and the relevance of each one to the current study.

2.2 Review of Previous Studies

This section deals with previous studies similar or related to the present one. It reviews studies carried out both in the Sudanese context and outside Sudan so as to provide a global image of the problem under investigation.

2.2.1 Local Previous Studies

2.2.1.1 Rabie, E. (2017) in his doctoral thesis, investigated the problems of understanding and using idiomatic expressions and collocations among Sudanese university EFL students majoring in English language. In this context, Rabie (2017) claimed that:

"Idiomatic expressions put non-native speakers in hot water both in written and oral contexts. In addition to that, material writers and language teachers find it difficult to develop effective materials and teach them effectively so that they can have a subsidiary place in syllabus" (p. 2).

The study adopted the descriptive analytical method of research. A questionnaire was designed by the researcher for English language teachers and a test on idiomatic expressions and collocations was designed for the students. The teachers' questionnaire consisted of 15 statements about the investigated problems and teachers' views about these problems as well as the importance of idiomatic expressions and collocations to foreign language learning (FLL). The questionnaire was administered to 40 teachers at different Sudanese universities with different years of experience ranging from 5 - 20. On the other hand, the test was administered to 120 students from Sudan University of Science and Technology.

Findings of the study revealed that the students have problems in understanding and using idiomatic expressions and collocations and mastering these language patterns enhances students' ability of expression and develops their receptive and productive skills. Furthermore, it is also showed that idiomatic expressions and collocations are significant to learning a foreign language.

Based on the research findings, it is recommended that more attention needs to be paid to figurative language. It is also recommended that teachers should use different techniques for teaching idiomatic expressions and collocations

Lastly, it is suggested by Rabie (2017) that more studies need to be carried out to investigate the effective ways of learning and teaching figurative language as well as further studies need to be conducted on the cruciality of idiomatic expressions and collocations to foreign language learners.

It is clearly demonstrated by the above study that Sudanese EFL learners (even those majoring in English language) find it difficult to deal with idiomatic expressions and collocations. The main factors behind these difficulties are attributed to the teaching materials and the teaching techniques adopted. It is assumed that corpora are vital tools to learning and teaching collocations and idiomatic expressions and other clusters of the language because dealing with such language units is a matter of probability and context preference rather than prohibition. This notion is demonstrated by O'keeffe, McCarthy and Carter (2007) claiming that, common verbs such as *get*, *go*, *turn* and so on display distinct preferences for what they combine with. Things *turn* or *go grey*, *brown*, *white*, people *go* (but not *turn*) *mad*, *insane*, *bold*, *blind* (p, 59). So that dealing with these verbs is a matter of quantity which corpora are very effective at. Therefore, the current study adopted corpora as an effective strategy to learning collocations.

2.2.1.2 Ali, A. (2017) in her doctoral dissertation, examined the Sudanese EFL undergraduates' knowledge and use of collocations. The study sought to explore the knowledge and use of lexical collocations and their potential effect upon EFL learners as well as their impact on learners' language proficiency at university level. Regarding this, Ali (2017) maintained that, the potential problem arises from the fact that reference to collocations though early made it in the syllabus students have been exposed to at the secondary and university level, the clarity of the concept still remains far and behind. Students had a go at the term in question at their secondary school studies without being told explicitly about the nature and possible application of it (p. 3).

Participants of the study were English language teachers and EFL students from College of Languages, Sudan University of Science and Technology. The study

adopted the descriptive and analytic method and two data collection tools were used; a completion test for students and a questionnaire for teachers.

Results of the study proved that students possess limited knowledge and awareness of English language collocations. Besides, the deviant answers of students' tests indicated learners' poor knowledge of lexical and grammatical collocations usage.

Based on the findings of the study, it is recommended that EFL teachers should raise the awareness of their students on the significance of collocations as well as they should foster their collocational competence. Furthermore, EFL teachers should provide examples of collocational differences between L1 and L2, especially those collocations which do not have equivalents in L1. Finally, it is recommended that teachers should encourage their students to use a dictionary which provides good coverage of collocations.

Once again, the above study proved the fact that Sudanese EFL learners face some difficulties in dealing with collocations, especially lexical collocations which are considered to be essential to achieving high level of proficiency in English language. Unlike the study above, it is hypothesized by the present one that these difficulties arise due to the fact that students have not been exposed to corpus-based syllabuses which represent real language used by native speakers. So it is strongly believed that unlike dictionaries as suggested by Ali (2017), using corpora in language learning and teaching has been proved to be effective because corpora are wide language data collected from different registers. These collections provide different senses and uses of words in different contexts of language use. Therefore, the present study employs corpora in solving such collocation learning difficulties encountered by EFL/ESL learners.

2.2.1.3 Nasr, O. (2011) conducted a Ph.D. study on assessing English collocational knowledge among Sudanese EFL university students. The study is carried out in an attempt to investigate the knowledge of collocations and to come out with constructive recommendations for both teachers and learners in order to improve the processes of teaching and learning the language.

Emphasizing the significance of collocations, Nasr (2011) highlighted that:

> *"Collocations are both indispensable and problematic for language learners and therefore should play an important part in second language teaching, especially to advanced learners"* (p. 3).

Subjects of the study were 150 students drawn from second, third, and fourth year students majoring in English language from Khartoum University, Faculty of Arts. A collocation test of 50 items was used to collect the data, and then the data was statistically analysed using the SPSS package.

Results of the statistical analysis indicated that there is a significant difference in students' knowledge of collocations between the three levels. It is also revealed that the adjective-noun is the easiest while verb-noun is found to be the most difficult one. The findings of the study proved that the subjects' knowledge of collocations is poor.

Based on the findings of the study and emphasizing the significance of collocations, Nasr (2011) recommended that, firstly, many studies on collocations have been conducted in the past few decades but our knowledge and understanding of collocations is still insufficient. It is teachers' role to teach students how words work in combination. When students come across collocations, teachers should concentrate on that during the teaching process. She added that, collocations should be taught within context and teachers should provide students with

supplementary reading such as literature and other disciplines materials of the language in order to gain more vocabulary of the language, so that learners can gradually get exposed to the language.

The study reviewed above confirmed that, collocations, especially verb-noun collocations cause difficulties to Sudanese English language learners even for those of advanced levels. Nasr's (2011) study is in line with the present one concerning collocations as an area of challenge to Sudanese learners of English but different strategy for tackling these difficulties (the use of corpora) and different tools of data collection are adopted by the current research.

2.2.2 External Previous Studies

2.2.2.1 Alzi'abi, S. (2017) conducted a study in Jordan entitled: Guessing verb-adverb collocations: Arab EFL learners use of electronic dictionaries. The study explored the way advanced Arab EFL learners handle verb-adverb collocations using two learner's dictionaries. Alzi'abi (2017) stated that, collocational studies have recently attracted a great deal of interest, but to date, hardly any study has tackled Arab EFL learners' competency in the use of verb-adverb collocations.

The sample of the study was 88 M.A. students majoring in English at five Jordanian universities consisting of 45 males and 41 females. Their ages range between 23 - 34. These subjects have been studying English language for more than 16 years prior to the enrolment to the trial. These were required to look up 22 verbs; 12 frequent and 10 infrequent, and then guess the adverb collocates of each verb.

The study results showed that even advanced EFL learners had considerable difficulty in providing extra adverb collocates of both frequent and infrequent

verbs. Dictionary use was effective and the subjects performed significantly better especially with infrequent verbs.

As projected above, collocations pose some challenges for EFL learners, even to those of advanced levels. According to the present research, these difficulties are attributed to a number of factors; prime among them is the teaching syllabuses students have been exposed to. It is suggested that learners performance would be much better if they are exposed to corpus-based materials. Hence, the prime aim of this research is employing corpora in tackling these difficulties.

2.2.2.2 Manafi, S. and Ghaffarof, S. (2013) in the Iranian context, examined the effect of collocational competence on translation accuracy of translation trainees. The study aimed to scrutinize the effect of productive and receptive knowledge of lexical and grammatical collocations on the accuracy of translation done by Iranian EFL learners studying translation course at university.

Manafi et al. (2013) stated that:

> *"Although it is widely acknowledged that collocations play an important role in second language acquisition, a number of previous studies have reported students' lack of collocational competence and the difficulties they encounter in learning and using collocations"* (p. 76).

Participants of the study were 60 M.A. students studying English language at Azad University, Iran. Their productive knowledge of collocations was measured by three gap-filling tests: verb-noun and adjective-noun collocation tests where the initial letter of the collocate was provided, and a verb-preposition test where the meaning of the phrasal verb was supplied. On the other hand, the participants' receptive knowledge of collocations was measured by an appropriate judgment in which subjects have to circle the number corresponding to the underlined part of a

sentence that is judged unacceptable. Concerning the translation quality of learners, a test including five paragraphs was given to them, and then the accuracy of the translated work was measured by Khanmohamad and OsanoRubic (2009) model.

Results of the study indicated that there is a significant relationship between the receptive and productive knowledge of lexical and grammatical colligations and the accuracy of translation.

From the study projected above, it could be inferred that learners' performance on translation tasks clearly projects their collocational awareness. This fact emphasizes the significance of the knowledge of collocations even to translation learners. This high requisition for collocational competence for all users of English language has placed corpus-based techniques on the top effective learning strategies for corpora help foster learners' knowledge of English language. Concerning the relationship between collocation knowledge and translation accuracy, corpora can be used to compare texts of different languages via what is called parallel corpora so that learners can identify similarities and differences between their L1 and English language.

2.2.2.3 Salih, A. F. (2016) carried out a doctoral dissertation on Saudi students entitled: a corpus-based study of academic collocations and patterns in postgraduate computer science students' writing. She highlighted that:

> *"Collocation has been considered a problematic area for L2 learners. Various studies have been conducted to investigate native speaker (NS) and non-native speakers' (NNS) use of different types of collocations (e.g., Durrant and Schmitt, 2009; Laufer and Waldman, 2011)"* (Salih, 2016, p. IV*).*

She added that, unlike native speakers, non-native speakers rely on a limited set of collocations in their academic writing.

The study aimed to achieve three main aims. Firstly, it examined the use of lexical academic collocations in NNS and NS Computer Science students' MSc dissertations and compares their uses with those written by expert writers in their writing of published research articles. Secondly, it explored the factors behind the under/overuse of 24 shared lexical collocations among corpora. Thirdly, it developed awareness-raising activities that could be used to help non-expert NNS students with collocation over/underuse problems. For the purpose of the study, a corpus of 600,000 words was compiled from 55 dissertations (26 written by NS and 29 by NNS). For comparison purposes, a reference corpus of 600,269 words was compiled from 63 research articles from prestigious high impact factor computer science academic journals. The Academic Word List (AWL) was used to develop lists of the most frequent academic words in the student corpora, whose collocations were examined.

Quantitative analysis was carried out by comparing the 100 most frequent noun and verb collocations from each of the student corpora with the reference corpus.

Results of the study revealed that both NNS and NS students overuse noun collocations compared to the expert writers in the reference corpus. They underuse only a small number of noun collocations. Surprisingly, neither NNS nor NS students significantly over/underuse verb collocations compared to the reference corpus. Mixed method approach was also adopted by the research in order to achieve the second aim of the study. First of all, the variant patterns of the 24 shared noun collocations between NNS and NS corpora were identified to determine whether over/underuse of these collocations could be explained by their

differences in the number of patterns including noun+preposition+noun and noun+adjective+noun that were rarely located in the writing of experts. Secondly, a categorization task and semi-structured interview were carried out with three computer scientists to elicit their views on the various factors likely influencing noun collocation choices by the writers across corpora.

The results proved that three main factors could explain the variation; these factors are sub-discipline, topic, and genre. Finally, a sample of awareness-raising activities was designed to achieve the third pedagogical aim. This sample was designed for the problematic over/underuse of some noun collocations. Then three types of activities were developed to raise the awareness of collocations: noticing collocations, noticing and identifying different patterns of the same collocations, and comparing and contrasting patterns between NNS students' corpora and reference corpus.

Results of the study suggested that academic collocation use in an ESP context (Computer Science) is related to other factors than students' lack of knowledge of collocations. These factors are expertness, genre variation, topic and discipline-specific collocations are proved to be important factors to be considered in ESP. Therefore, ESP teachers have to alert their students to the effect of these factors in academic collocation use in subject-specific discipline.

Based on the findings of the study, it is suggested that it would be useful in future to carry out a study where some of computer science postgraduate non-native speakers and native speakers are interviewed. Thus, their views about the over/underuse of noun collocations could be investigated and possibly more learnt about the underlying factors.

To the best of the research's knowledge, corpora are one of the best human inventions that could effectively be used to investigate a wide range of language features such as register variation and ESP as shown in Salih's study above. It is proved by the above study that non-native speakers writing is different from that of native speakers due to the fact that in contrast to native speakers, non-native speakers rely on certain collocation patterns in their use of the language. This reliance, according to the current study, is attributed to students' poor collocational knowledge which can best be improved via the use of corpora, especially in understanding collocations across registers.

It is believed by the current research that understanding register variation is difficult without corpus-based approach. In this concern, Conard, Bibber and Reppen (1998) maintained that, comprehensive register studies have three important requirements: inclusion of a large number of texts, consideration of a wide range of linguistic characteristics, and comparison across registers. These are difficult to be achieved without corpus-based techniques (p. 139).

2.2.2.4 Alharbi, R. M. (2017) in the Saudi context, conducted a Ph.D. dissertation on acquisition of lexical collocations: a corpus-assisted contrastive analysis and translation approach. The study examined the effects of three focus-on-form instructional approaches on the passive and active acquisition of non-congruent collocations: The non-corpus-assisted contrastive analysis and translation (CAT) approach, the corpus-assisted approach, and the corpus-assisted non-CAT approach.

Emphasizing the idiomatic nature of English language, Alharbi claimed that:

> "*Collocations, i.e. words that usually go together as one form of formulaic sequences, constitute an inherent problem for ESL/EFL*

learners. Researchers have submitted that non-congruent collocations, i.e. collocations that do not have corresponding L1 equivalents, are especially difficult to acquire by ESL/EFL learners" (Alharbi, 2017, p. I).

To fully assess the corpus-assisted contrastive analysis and translation approach and its learning outcomes, a control group was included (under no condition) for a baseline comparison. Thirty non-congruent collocations with the learners' L1 (Arabic) were selected for the study.

Participants of the study were 129 undergraduate EFL learners in a Saudi university. They were assigned to three experimental groups and a control group using a cluster random sampling method. With the help of bilingual English/Arabic corpus data, the corpus-assisted CAT group performed (L1/L2 and L2/L1) translation. The non-corpus CAT group was assigned text-based translation tasks and received contrastive analysis of the target collocations and their L1 translation options from the teacher. With the help of monolingual corpus data, the non-contrastive group performed multiple choice/gap filling tasks and focusing on the target items. Immediately after the intervention stage, the three groups were tested on the retention of the target collocations by two tests: active recall and passive recall. Then three weeks later, the same tests were administered to the participants. The corpus-assisted CAT group significantly outperformed the other two groups on all the tests. These results were discussed on light of the 'noticing task–induced involvement load', and 'push output' hypotheses and influence that L1 exerts on the acquisition of vocabulary. The discussion includes an evaluation of the three instructional conditions in relation to different determinants, dimensions and functions within the hypotheses.

It is proved by the above study that much of the natural language consists of chunks or formulaic sequences. Investigating these language patterns requires a

large and authentic amount of language data. According to the present study, this property is only provided by language corpora because they are collected from different contexts of language use and meant to represent a language or variety of a language.

2.2.2.5 Choorit, K. and Supakorn, P. (2014) carried out a corpus-based study of the phrasal verbs: *carry out*, *find out*, and *point out*. The study was conducted in Malaysia. It was exploratory and corpus-based aimed to investigate the grammatical patterns and the collocates of three phrasal verbs (PVs) in English language. These are; carry out, find out, and point out. According to Choorit et al. (2014), these phrasal verbs were chosen as the focus of the study for several reasons; firstly, they are the top three most prolific adverbial particles in forming phrasal verbs. Secondly, the lexical verbs *carry*, *find* and *point* are the top 20 lexical verb lemmas functioning in phrasal verb forms. Lastly, the phrasal verbs *carry out*, *find out*, and *point out* are the top 50 phrasal verb lemmas in the BNC, in COCA, and in Malaysian EFL textbooks Corpus.

Concerning the data collection, the search features of the COCA online program (at http://corpus.byu.edu) provided by Brigham Young University are used to retrieve 500 concordance lines for each target phrasal verb. Then, these concordance lines were placed in an Excel spreadsheet for analysis. Oxford Advanced Learner's Dictionary (2010) was used as the main reference for information about the definitions of the target phrasal verbs as well as grammatical information and collocates of these verbs.

Results of the study revealed that although most of the grammatical patterns found in this study seem to be in accordance with those shown in OALD (2010), certain patterns are not provided in the dictionary. The reason could be the limited

space of the dictionary and users' stylistic variation in phrasal verbs use. Regarding the collocates of the target phrasal verbs, most of them were lexical collocations such as nouns, pronouns and wh-words, followed by grammatical collocations such as prepositions.

Pedagogically speaking, the corpus-based data could help provide additional information not illustrated in dictionaries. In addition, learners can explore the language patterns and search for naturally occurring samples through the use of corpora.

Based on the findings of the study, the following recommendations were provided. To begin with, based on the fact that the study employed only one American-English corpus, namely the American National Corpus, with a query of 500 concordance lines, more interesting results in terms of grammatical patterns and collocations could be discovered if a variety of corpora is used and the search includes a greater number of concordance lines. Secondly, as the meaning of phrasal verbs was not the main objective of this study, future research could focus on their semantic features. Lastly, the researcher did not query for a separabilty feature of the target phrasal verbs. In other words, they can have one or more words inserted between the lexical verbs and the adverbial particles.

The study above has tackled one of the language patterns that poses difficulties to English language learners. These difficulties reside in the fact that the particles that constitute phrasal verbs are very frequent in the language, so that they are challenging for learners to master. In this regard, Carmen et al. (2010) highlighted that, particles are highly frequent, non-salient and polysemous words and as such they are, like other words with the same features, very difficult to learn. They apparently provide redundant information and are difficult to notice by learners

whose first languages have not trained them to pay attention to this specific linguistic element (p. 150).

Like delexical verbs (the focal point of the present research), phrasal verbs need to be learnt within context, so that it is assumed that these features can better be tackled by the utilization of corpora as they have the potential of providing natural language data.

2.2.2.6 Ashouri, S., Arjmandi, M., and Rahimi, R. (2014) investigated the impact of corpus-based collocations instruction on Iranian EFL learners' collocation learning. The study attempted to shed light on the impact of corpus-based on EFL learners' collocation learning and awareness. Emphasizing this, Ashouri et al. (2014) highlighted the following:

> *"Over the past decades, studies of EFL/ESL vocabulary acquisition have*
> *identified the significance of collocations in language learning due to the*
> *fact that collocations have been regarded as one of the major concerns of*
> *both EFL teachers and learners for many years"* (p. 470).

In the same context, they added that vocabulary was neglected after the growth of Communicative Language Teaching (CLT). Learners could communicate but misunderstanding is likely to happen because of the lack of knowledge of vocabulary, its collocations, and pragmatic competence. So a lot of researchers' attention was drawn to vocabulary teaching and its collocations. Afterward, they came to the truth that lack of knowledge of collocations can prevent learners from inferring and they also misunderstand native speakers. Wrong collocations can cause native speakers not to comprehend what the learners say.

Subjects of the study were 60 Iranian EFL learners. They were randomly chosen based on their scores in an OPT exam. There were two groups;

experimental and control group. The study examined the effects of direct corpus-based collocation instruction on EFL learners' collocation learning. For 15 sessions, the control group received a single-item vocabulary in the usual work of their class and the experimental group received lexical collocations instruction as treatment. The same test as post-test was given to the learners when the treatment accomplished. After that, a t-test and Kolmogorov-Smirnov test between the pre-test and post-test were calculated.

According to the results demonstrated by the statistical analysis program, the effectiveness of the treatment was noticeable. The study suggested that direct corpus-based collocation instruction can be a worthy alternative. It demonstrated that the learners who were in the experimental group got aware of the existence of collocations, learnt them, and used them properly. In addition, they started to find the collocations of every other word which they learnt during the term by themselves because the treatment appealed to them.

The experimental study above used corpora in learning collocations as an aspect of vocabulary knowledge that many EFL/ESL learners struggle to master. According to the current research, the use of corpus-based approaches can make the learning process much easier than in the past due to many reasons: corpora can be used to investigate the grammatical features of words as well as their semantic ones, so that collocation learning can effectively be achieved. Furthermore, as corpora are collected from different contexts, they can be used to investigate different senses and collocations of words. In a related context, Bibber et al. (1998) maintained that, lexicographic work requires the use of very large corpora because word senses and collocational patterns are often much common than grammatical patterns (p. 25). They added that, corpora can be used to show all the contexts in

which a word occurs, so that it is possible to identify the different meanings associated with a word.

2.2.2.7 Hong, Y. (2014) conducted a Ph.D dissertation on collocation errors among Chinese learners of English with reference to Chinese college students of Tongji University in China. The main aims of the study are; firstly, to describe and contrast the collocations of Chinese (Mandarin) and English. Secondly, it aimed to explain the major types of errors of English collocations committed by Chinese learners of English. Thirdly, identifying areas of difficulty encountered by Chinese learners in the use of collocations using contrastive analysis (CA) and error analysis (EA). For the purpose of achieving these goals, both English and Chinese collocations are described and compared. These collocations received little attention in the past.

The data were collected from one writing task administered to students during classroom hours. The learner corpus designed consisted of 117 pieces of written texts. In the process of data analysis, received categories about linguistic errors were critically investigated. In order to highlight the selected KWIC, Wordsmith tool with concordance lines of BNC and LC was employed. Oxford Advanced Learner's Dictionary was also employed to identify and analyze errors in order to count the number and percentages of errors in the EA, SPSS and MicroFinder were used in the study. CA involved two methods; theoretical CA which aimed at predicting potential learning difficulties by analyzing the differences between the structures of English and Chinese and the second method is applied CA which aimed at identifying and explaining actual errors committed by the students. The study identified seven categories of English collocation errors. They were errors on noun+noun/preposition phrase collocation, noun+verb/phrasal verb collocation, verb/phrasal verb+noun collocation, adjective+noun, and many others. Errors on

English subject-predicate structures were found to be the most common of all subcategories of English collocations among Chinese learners of English language, followed by English verb+noun collocation errors, etc. Intralingual source of errors found in the context such as ignorance of restrictive collocation rules of English was found to be most responsible for English collocation errors committed by Chinese learners of English.

Pedagogically, the study findings implied that CA-based methodology enabled students to locate and focus on the difficulties of learning English vocabulary to enhance teaching effectiveness. EA can be helpful in identifying all possible errors and thus overcoming them. Finally, it is suggested that learning strategies such as collocation strategy are important for learners in learning English vocabulary.

The above study employed one of the potential areas of using corpora which is the use of corpus to compare and contrast linguistic features and their frequencies in two languages (or more) subject to contrastive analysis. Such corpora are called parallel corpora and they are used in the above study to investigate similarities and differences between Chinese (first language) and English (target language) in order to identify the major collocation errors committed by Chinese learners of English. The study is similar to the current one in terms of employing corpora in the process of collocation learning as well as they both employ quantitative methods to data collection.

2.2.2.8 Akbar, A., Hashemian, M., and Alipour, S. (2013) carried out a study in Iran entitled; a corpus-based approach towards teaching collocation of synonyms. The study aimed to provide insights into the use of the corpus-based approach in L2 classes as well as comparing the effects of the corpus-based with the effects of the traditional approach in learning collocations of near-synonymous pairs.

Emphasizing the importance of collocations and the role of corpora in highlighting them, Brown maintained that:

> "*More attention can be placed on collocations by using concordances to highlight collocations to L2 learners*" (Brown, 1974, cited in Akbar et al. 2013, p. 57).

Subjects of the study were two intact groups of L2 learners. The first group named the experimental group studies with concordancing lines, while the other group called the control group studies with the traditional approach. The participants were similar in terms of their collocational competence and English language proficiency. During the study, the first group was taught through concordancing materials taken from the BNC. On the other side, the control group was taught the collocations explicitly. The instruments employed in the data collection process were a pre-test, post-test, pre-writing, and post-writing.

Through running a t-test and descriptive statistics, results of the study proved that there is a difference between the two approaches about their effect on the comprehension and production of collocations of synonyms.

The study above employed corpora in teaching near synonyms in English language. Such words are considered synonymous and identical in meaning by thesauruses and dictionaries. Whereas, in corpora, the situation is totally different where these synonyms are easily distinguished by their collocates, context of use, and register differences. In the context of corpora, considering such words as identical synonyms is misleading because synonymous words are used in completely different ways. In this regard, Biber, Conard and Reppen (1998) highlighted that synonymous words are typically used in very different ways so

that corpus-based lexicographic analysis are particularly well suited to uncover such systematic differences in the patterns of use.

The research is similar to the current one in terms of using corpora in language learning, especially collocations. According to both, corpora have the potential of providing real and authentic language data that students can exploit to improve their proficiency of English language.

2.2.2.9 Barabadi, B., and Khajavi, Y. (2016) investigated the effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. The study used teacher-prepared materials obtained from COCA. It aimed to compare data-driven learning and traditional methods of teaching vocabulary like consulting a dictionary or a grammar book.

Emphasizing the significance of vocabulary to language learning, Schmitt and Singleton claimed that today, the critical significance of vocabulary learning in developing second language learning is appreciated by all L2 teachers and researchers. No language teacher or learner contests the importance of vocabulary and the lexical dimensions in learning that language. Indeed, second language (L2) acquisition depends to a large extent on the development of a strong vocabulary (Schmitt, 2000 and Singleton, 1999, cited in Barabadi et al., 2016, p. 2).

Participants of the study were 62 CAE (Certificate for Advanced English) students. They were all Persian native speakers from three CAE-level classes at two English institutes in Mshhad, Iran. 28 subjects were females and the rest were males. Their ages range between 17 - 26. The three intact classes were randomly assigned to two treatment groups and one control group. 42 students were assigned to treatment group, and they were in two language classes in one language institute. In order to determine the level of participants in vocabulary knowledge,

Nation's (2001) Vocabulary Size Test was administered to them as a pre-test to make sure that all subjects are at small level of vocabulary knowledge. During the semester, the learners were exposed to teacher-prepared corpus-based materials from COCA such as displays lists, synonyms, keywords in context (KWIC), and collocate. In addition to that, they were asked to do similar searches of their own as homework. Based on their course book (Reading for CAE), a post-test was designed and administered to them at the end of the semester.

Results of the study proved that learners in the experimental group outperformed those of the control group. The better performance of the experimental group can be attributed to the fact that learners could take a more active role in the process of learning in which self-discovery, inductive learning and bottom-up processes were emphasized.

One of the effective corpus techniques used in language learning and teaching is the use of concordance lines. These lines display the right and left collocates of words in different contexts. According to the current study, it has been demonstrated by many scholars that concordancing is very helpful in language learning, collocations in particular which are considered to be an important aspect of vocabulary knowledge. In contrast to corpora, dictionaries do not fully provide context of language use. This is because of the fact that dictionaries are based on intuitions of their designers as well as the insufficient space of dictionaries. Considering the above mentioned facts, corpora are chosen by the current study as the best source of language data that could help in solving collocation learning difficulties encountered by EFL/ESL learners.

2.2.2.10 Girgin, U. (2011) investigated the effectiveness of using concordance lines on grammar teaching in a dissertation entitled; corpus-based activities at

lower levels of EFL proficiency: The effectiveness of using concordance lines on grammar learning. The study was conducted in Turkey. It aimed to determine whether lower level EFL learners would be able to use corpus-based activities effectively in order to learn five target grammar structures in English language. The study also explored the attitudes of those students towards using corpus-based activities in English grammar learning. Speaking of the potentiality of corpora in language learning, it is claimed by Girgin that: The use of corpora or corpus-based activities in language classes is very beneficial for EFL learners to get familiar with real and authentic language. Language teachers also benefit from corpora to increase the meaningful input that is provided to learners. It is also to the advantage of textbook writers to exploit corpora in order to gain an accurate reflection of the language actually used by native speakers and writers in natural situations rather than relying on their beliefs and intuitions while preparing materials in the field (Biber and Reppen, 2002, cited in Girgin, 2011, p. 1).

Six intact lower level EFL classes at Erciyes University, School of Foreign languages participated in the study. Both quantitative and qualitative data were involved in the study. The former was collected through the administration of three tests and attitude questionnaire. Whereas, the later was obtained through semi-structured interviews.

The statistical analysis of the test results revealed that the students were able to use corpus-based activities effectively in the learning of the target structures. Furthermore, it was found that using corpus-based activities in learning the target grammar structures produced similar results when compared to using a course book. The analysis of the attitude questionnaire revealed that the students held somewhat neutral attitudes towards using corpus-based activities in the learning of

five target structures. However, the analysis of students' interviews showed more positive attitudes towards using these sources in English grammar learning.

The above study examined the potentiality of corpora in grammar learning. This potentiality has not been exploited by teachers to teach grammar until recently. This claim is supported by Bibber et al. (1998) highlighting that little attention was recently given to studying how native speakers actually exploit the grammatical resources in their language.

The grammatical areas that traditional approaches have neglected turned out to be the strength of corpora because they cannot only be used to investigate grammar at advanced levels but they can also be exploited even by students of lower levels of language proficiency such as the morphological level (the structure of words). Conversely, language learners, especially lower level learners encounter some challenges of using corpora such as the large amount of search results but the job of the teacher is to ease this task by tailoring the results so that they can suit the needs of students.

2.2.2.11 Agnes, L. (2012) conducted a study in Thailand. It scrutinized the effectiveness of a corpus-based instruction in deepening EFL learners' knowledge of periphrastic causatives. It aimed to investigate whether a corpus-based instruction could deepen EFL learners' knowledge of periphrastic causatives: *make*, *cause*, and *let*.

47 Taiwanese undergraduates from two intact classes were selected as the subjects of the study. One class was chosen as the experimental group received three-month corpus-based instructions, while the other group which is the control group had no instruction on English causatives. The researcher as instructor conducted the instruction following a data-driven model of illustration, interaction

and induction and then took notes on students performance. After the instruction, a post-test was given to both groups. On the other hand, a questionnaire on learning effects and students' feelings towards corpus-based activities was administered to the experimental group. Three data sources were analyzed: pre and post-tests, field notes, and questionnaires.

Results of the study showed that the experimental group improved and outperformed the control group significantly in the post-test. Moreover, results of the questionnaire proved that the instruction was effective in increasing students' knowledge of the three causatives. However, the field notes revealed that learners face some difficulties in using causative verbs. It is suggested by Agnes (2012) that to better distinguish these verbs' usages, learners should attend to semantic distinctions more than syntactic structures. Instructors also need to provide clear guidance on data search and data interpretation.

Agne's study is similar to the current one in terms of the data collection instruments (mixed method research) and the learning strategy adopted which is employing corpora in improving EFL learners' knowledge of vocabulary. As demonstrated by the above study, corpora have been proved to be an effective strategy to English language learning and teaching because of their inclusion of large and authentic language data. These data are believed to be effective in investigating words and their grammatical environments which provide a baseline for understanding collocations of words. It is claimed by this study that understanding the use of individual words (such as periphrastic causatives, delexical verbs, etc.) cannot be achieved without their context of use which can best be provided by corpus-based activities. Consequently, corpora are suggested by the current research as a new strategy to language teaching and learning.

In summary, this part has reviewed and discussed the previous studies similar/relevant to students' difficulties of collocations and the use of corpora in language learning and teaching. These studies are discussed with regard to their designs adopted, the instruments of data collection employed and the findings obtained. Some of these studies have dealt with learners' problems with collocation learning and use, while others have investigated the use of corpora in language learning and teaching. Most of these studies have come to the fact that corpora are effective strategy to language learning in comparison to the other strategies but none of them have examined the use of corpora in learning delexical verbs. This has come to be the strength of this study.

2.2.3 Summary of the Chapter

This chapter has reviewed two main parts; the theoretical framework and review of the literature. On the first hand, the conceptual background has reviewed and discussed the core concepts of the area under investigation and their relevance to learning/teaching collocations of English language as a challenging area to both EFL and ESL learners. On the other hand, the second part of the chapter has dealt with review of the literature, presenting previous studies conducted in both the local and the international level in an attempt to fully evaluate learners' knowledge of collocations and the role of corpora in promoting that aspect of vocabulary knowledge. While viewing these previous studies, the researcher has observed that little attention is directed to the role of corpora in raising the awareness of collocations and scant or no attention is given to delexical structure (the area under investigation). Unlike the previous studies projected above, this study is trying to integrate corpora into English language curriculum so that they can constantly be accessed by learners/teachers and use them in learning/teaching.

Chapter Three

Research Methodology

Chapter Three

Research Methodology

3.0 Overview

This chapter presents and describes the methodology used in the research. It provides detailed description of the design of the study. In addition to that, it presents the research population, the sampling procedure, the sample, and the instrumentation of the study adopted for the data collection process.

3.1 The Methodology

For the purpose of enhancing the study findings, quantitative (two tests and a questionnaire) and qualitative (interview) methods are adopted by the present research. So the research design is mixed as it incorporates the two methods. As Creswell (2009) claimed, mixed method research is an approach to inquiry that combines or associates both qualitative and quantitative forms of research. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mixing of both approaches in a study (p. 2). According to Creswell (2009:205), many different terms are used for this approach, such as integrating, synthesis, qualitative and quantitative methods, multimethod, and mixed methodology, but that recent writings used the term mixed methods. Qualitative and quantitative instruments are not opposites as many people might think; rather, they complement each other and strengthen the research findings. This notion is supported by Creswell (2012), stating that the uses of both quantitative and qualitative methods, in combination, provide a better understanding of the research problem and question than either method by itself (p. 535).

3.2 Study Population and Sample of the Experiment

This part provides detailed description of the study population, the procedure of selecting the study sample, and sample of the research.

3.2.1 The Population

The population of this study is second year students of Biomedical Engineering at Bayan College of Science and Technology, studying English as a university requirement. These students are the subjects of the tests as well as the questionnaire. In addition to that, the participants of the interview are experienced English language teachers from different universities in Khartoum State. They are interviewed so as to gain their views on the use of corpora in learning and teaching English language, specifically collocations.

3.2.2 The Sampling Procedure

A number of sampling techniques are available and which one to choose largely depends on the nature of the research problem. Kothari (1990) maintained that there are different types of sample designs based on two factors; the representation bases and the element selection technique. On the representation basis, the sample may be probability sampling or it may be non-probability sampling. Probability sampling is based on the concept of random selection. Whereas, non-probability sampling is non-random sampling. Concerning the element selection technique, the sample may either be restricted or unrestricted. When each sample element is drawn individually from the population at large, then the sample so drawn is known as 'unrestricted sample'. Whereas, all other forms of sampling are covered under the term 'restricted sampling' (p, 58).

The following chart demonstrates the sampling designs as explained above.

Figure 3.1: Basic sampling designs.
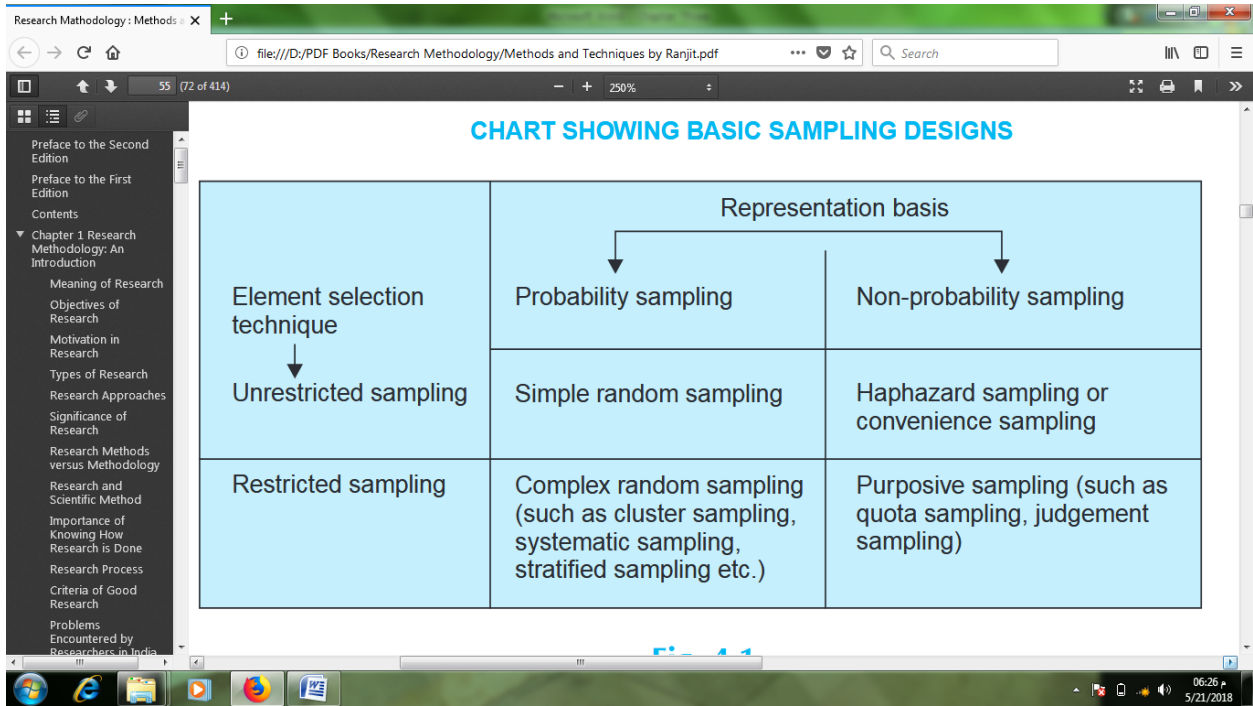
Source: Kothari (1990, p. 59).

As shown in the above chart, sampling designs are based on two factors; element selection technique and representation basis. Regarding element selection technique, it is either restricted or unrestricted. With regard to representation basis, it is either probability or non-probability sampling. The former is also called random sampling. In this sampling design, all items of the population have equal chance of being included in the sample. On the other hand, the latter refers to the fact that items for the sample are deliberately selected by the researcher to be representative of the whole population.

As highlighted in Kothari (1990:58), characteristics of a good sample design can be summarized in the following:

- Sample design must result in a truly representative sample.

- Sample design must be such which results in a small sampling error.

- Sample design must be viable in the context of funds available for the research study.

- Sample design must be such so that systematic bias can be controlled in a better way.

- Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

In the present study, convenience sample is adopted. It is defined as a statistical method of drawing representative data by selecting people because of the ease of their volunteering or selecting units because of their availability or easy access. Convenience sample is a group of subjects selected because of their availability (McMillan, 1996:91). McMillan (1996) confirmed that although we need to be very wary of convenience samples, often this is the only type of sampling possible, and the primary purpose of the research may not be to generalize but to better understand relationships that may exist. This type of sampling is formed when we select elements from a population on the basis of what elements are easy to obtain. Convenience sampling is also called grasp sample as we essentially grasp members of the population for our sample. This sample is characterized as being easy, fast, and inexpensive.

3.2.3 The Sample

McMillan (1996) defined sample as the group of elements, or a single element, from which data are obtained. Although the phrase "the sample included" is used to indicate the characteristics of the people or events in the sample, the nature of

the sampling procedure is usually described by one or more adjectives, such as random sampling or stratified random sampling (p, 86).

Pandey & Pandey (2015:45) identified seven characteristics of a good sample as follows:

- A good sample is the true representative of the population corresponding to its properties.
- The population is known as aggregate of certain properties and sample is called sub-aggregate of the universe.
- A good sample is free from bias; the sample does not permit prejudices, the learning and pre-conception, imaginations of the investigator to influence its choice.
- A good sample is an objective one; it refers to objectivity in selecting procedure or absence of subjective elements from the situation.
- A good sample maintains accuracy. It yields an accurate estimates or statistics and does not involve errors.
- A good sample is comprehensive in nature. This feature is closely linked to representativeness.
- A good sample has the practicability for research.

Concerning the present study, 60 second year students of Biomedical Engineering (Bachelor) from Bayan College of Science and Technology are selected to be the sample of the study.

3.3 The Instrumentation

According to Singh (2006), a great variety of research tools are of many kinds and employ distinctive ways of describing and qualifying the data. Each tool is

particularly appropriate for certain sources of data yielding information of the kind and in the form that would be most effectively used (p.191). Creswell (2009) highlighted that different types of research tools can be used to collect data such as questionnaires, tests, documents, audiovisual materials, interviews, observations, etcetera. Pandey, et al. (2015) maintained that tools may vary in complexity, interpretation, design and administration and each tool is suitable for the collection of certain type of information (p. 57).

Based on the study questions and hypotheses, the current research adopts the mixed method strategy (qualitative and quantitative) for the purpose of enhancing the research findings, specifically two tests, questionnaire, and interview are chosen for the data collection.

3.3.1 The Tests

The present study adopts pre-test and post-test for the collection of data. These tests are usually found in quasi-experimental studies where the participants are studied before and after the experimental manipulation.

According to Cohen, Manion and Morrison (2005:334), the pre-test and post-test must adhere to several guidelines:

- The pre-test may have questions which differ in form or wording from the post-test, though the two tests must test the same content, i.e. they will be alternate forms of a test for the same groups.
- The pre-test must be the same for the control and experimental groups.
- The post-test must be the same for both groups.
- Care must be taken in the construction of a post-test to avoid making the test easier to complete by one group than another.

- The level of difficulty must be the same in both tests.

Concerning the current research, pre-test and post-test are adopted as one of the tools of primary data collection. Because of the unavailability of already-existing tests on collocations (delexical verbs) in the previous studies, the researcher comes up with self-designed tests suitable with the present study design. The two tests are the pre-test and post-test. The pre-test aims to examine students' knowledge and use of verb-noun collocations prior to the experiment, specifically the delexical verbs *make*, *do*, *have*, and *take* as commonly used light verbs in English language. Both tests are composed of two main parts; the first part is about participant's personal information. The second part on the other hand, is about the delexical verbs and it is divided into two sections; the first one deals with fifteen questions about *make* and *do*. Whereas, the second one consists of fifteen questions about the uses of *have* and *take*. In each section, the students are given two alternatives (verbs) to choose the appropriate one. The level of difficulty is the same for both the pre-test and post-test. Then the two tests were shown to university professors and doctors from different universities in Khartoum State. Namely, Sudan University of Science and Technology, Omdurman Islamic University and Blue Nile University to ensure their validity.

Before distributing the tests to the participants, 60 students of Biomedical Engineering are selected to be the sample of the study. To make sure that they are at the same level of language, a pre-test was distributed to them. Then they were divided into two groups; the control group and the experimental group. The former was taught delexical verbs following non-corpus-based approach (traditional approach) to language teaching, while the latter was firstly enlighten by the researcher on what are corpora and how to practically use them in learning English language, especially collocations. Then they were taught the same delexical verbs

using British National Corpus through the techniques *collocates, KWIC* (keyword in context), *list of concordances*, and *compare* to compare *make* with *do* and *take* with *have*. The students were also enlightened on how to use corpora in learning and investigating other linguistic features. After that, the post-test was distributed to them.

Table 3.1: The distribution of participants of the pre-test and post-test and their gender.

No	Gender	Number	Percentage
1	Male	37	62%
2	Female	23	38%
	Total	60	100%

The table above demonstrates the distribution of the experimental group and control group and their gender. As shown in the table, 62% are males while 38% are females. This is due to the fact that the number of males admitted to the Department of Biomedical Engineering is more than their female counterparts.

Table 3.2: The experimental group participants and their distribution according to gender.

No	Gender	Number	Percentage
1	Male	17	57%
2	Female	13	43%
	Total	30	100%

The projected table above shows the distribution of the experimental group and their sex. As shown, male students constitute 57% of the participants. This is because male students admitted at the department are more than female students.

3.3.2 The Student Attitude Questionnaire

According to Singh (2006), a questionnaire is a form which is prepared and distributed for the purpose of securing responses. Generally, these questions are factual and designed for securing information about certain conditions or practices of which recipient is presumed to have knowledge (p. 191). Singh (2006) also highlighted that the questionnaire is considered to be the most flexible of tools and possesses a unique advantage over others in collecting both qualitative and quantitative information. Critics speak of it as the lazy man's way of gaining information, because it is comparatively easy to plan and administer a questionnaire. As a matter of fact, preparation of a good questionnaire takes a great deal of time, ingenuity and hard work (2006, p. 92). Speaking of the popularity of questionnaires, Kothari (1990) stated that questionnaires are being adopted by private individuals, research workers, private and public organisations and even by governments (p, 100).

The following are characteristics of a good questionnaire as shown in Pandey & Pandey (2015:59):

- It deals with an important or significant topic.
- Its significance is carefully stated on the questionnaire itself or on its covering letter.
- It seeks only that data which cannot be obtained from the resources like books, reports and records.
- It is as short as possible, only long enough to get the essential data.

- It is attractive in appearance, neatly arranged and clearly duplicated or printed.

- Directions are clear and complete, important terms are clarified.

- The questions are objective, with no clues, hints or suggestions.

- Questions are presented in an order from simple to complex.

- Double negatives, adverbs and descriptive adjectives are avoided.

- Double barreled questions or putting two questions in one question are also avoided.

- The questions carry adequate number of alternatives.

- It is easy to tabulate, summarize and interpret.

Concerning the questionnaire of the present study, it has been designed by the researcher based on Likert's 5-point scale and five alternatives (strongly agree, agree, neutral, disagree, and strongly disagree) were provided for the respondents to choose. It consists of 30 statements about the difficulties that face students in learning collocations, students' motivation towards learning these word combinations, and the role of corpora in learning these language clusters.

With regard to the questionnaire respondents, it was distributed to the same respondents of the experimental group (30 students from Bayan College of Science and Technology) with the aim of obtaining their views and responses on the difficulties of learning collocations, students' motivation towards learning these collocations, and the role of corpora as a new strategy to learning these language patterns. These students are selected because they had undergone the research experiment (learning collocations through corpora) so that they can judge whether corpora are effective in learning the language. Thus, the researcher can fully

measure and evaluate the effectiveness and usefulness of using corpora in collocation learning.

Table 3.3: Participants of student attitude questionnaire and their gender distribution.

No	Gender	Number	Percentage
1	Male	17	57%
2	Female	13	43%
	Total	30	100%

The above table demonstrates the distribution of student's attitude questionnaire participants and their sex.

3.3.3 The Teacher Interview

According to Cohen, et al. (2005) the interview is an interchange of views between two or more people on a topic of mutual interest, sees the centrality of human interaction for knowledge production, and emphasizes the social situatedness of research data (p. 267). The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method can be used through personal interviews and, if possible, through telephone interviews (Kothari, 1990:97). The interview could either be oral or sometimes written, or either be face to face or by phone depending on the access and availability of respondents.

The following are characteristics of interviews provided by Pandey & Pandey (2015:60):

- The interviewer can probe into casual factors, determine attitudes, and discover the origin of problem.

- It is appropriate to deal with young children and illiterate persons.
- It can make cross questioning possible.
- It helps the investigator to gain an impression of the person concerned.
- It can deal with delicate, confidential and even intimate topics.
- It has flexibility.
- Sincerity, frankness, truthfulness and insight of the interviewee can be better judged through cross questioning.
- It gives no chance for respondent to modify his earlier answer.
- It is applicable in survey method, but it is also applicable in historical, experimental, case studies and clinical studies.

Concerning the study interview, it is designed to collect data from experienced EFL teachers at university level with the intention of getting their views on the use of corpora in learning and teaching English language, especially collocations. With regard to the respondents, the interview is administered to English language teachers from different Sudanese universities in Khartoum State. Namely, Karari University, Elrazi University, Sudan International University, Comboni College of Science and Technology and Hayatt University College.

The interview (Appendix 4) consists of two parts; the first part is devoted to personal information of participants, such as name of participant (optional), name of educational institution where they work, academic qualifications, and years of experience plus professional qualifications. On the other hand, the second part is dedicated to the interview questions. Five questions were designed for the interview; the first question aims at defining what corpus is. The second question deals with the views of participants on the helpfulness of corpora as a necessary source of information for teaching English language, especially collocations. The

third question centres on how often the interviewees use corpora in teaching collocations in English language. In addition to that, the fourth question revolves around comparing corpora to other vocabulary learning strategies. Finally, the fifth question is devoted to using corpora for other purposes besides learning and teaching collocations in English language.

Table 3.4: The frequency distribution of teacher's interview respondents and their places of work.

No	University	Number	Percentage
1	Karari University	1	20%
2	Elrazi University	1	20%
3	Sudan International University	1	20%
4	Comboni College of Science and Technology	1	20%
5	Hayatt University College	1	20%
	Total	5	100%

The above table displays the frequency distribution of teacher's interview respondents and their places of work.

Table 3.5: The frequency distribution of teacher's interview respondents and their academic qualifications.

No	Academic Qualification	Number	Percentage
1	M.A.	4	80%
2	Ph.D	1	20%
	Total	5	100%

The table above illustrates the frequency distribution of teacher's interview respondents and their academic qualifications.

Table 3.6: The frequency distribution of teacher's interview respondents according to their years of experience.

No	Years of Experience	Number	Percentage
1	1 – 5	2	40%
2	6 – 10	1	20%
3	11 – 15	1	20%
4	More than 15	1	20%
	Total	5	100%

The table above demonstrates the frequency distribution of participants of teacher's interview and their years of experience.

3.4 Validity and Reliability
3.4.1 Validity of the Study Instruments

According to Neville (2007), validity refers to the extent to which the data accurately measures what they were intended to measure (p. 26). Validity highlights the need to eliminate or minimize the effects of extraneous influences, variables, and explanations that might detract from a study's ultimate findings (Marczyk, DeMatteo and Festinger, 2005:158). Therefore, validity is crucial to every research as it aims to increase the accuracy and usefulness of the findings by controlling these findings from extraneous variables that could distract the study from properly achieving its goals. Cohen, et al. (2005) precisely and concisely viewed validity as, if a piece of research is invalid then it is worthless (p. 105). As

Cohen, Manion & Morrison (2005) maintained, more recently, validity has taken many forms. For example, in qualitative data validity might be addressed through the honesty, depth, richness and scope of the data achieved, the participants approached, the extent of triangulation and the disinterestedness or objectivity of the researcher. In quantitative data, validity might be improved through careful sampling, appropriate instrumentation and appropriate statistical treatments of the data (p. 105). Therefore, this study follows the parameters of both qualitative and quantitative research in achieving its validity.

To ensure the validity of the present research tools (the questionnaire, the two tests, and the interview), they were all shown to EFL experts, including the main supervisor of the study and five experienced Ph.D holders in the field of applied linguistics (the study area). These referees are from Sudan University of Science and Technology, Omdurman Islamic University and Blue Nile University. Some of them have recommended doing some amendments on the tools while others have agreed on their validity, suitability and relevance to the study.

3.4.2 Reliability of the Study Instruments

Cohen, et al. (2005) stated that reliability is essentially a synonym for consistency and replicability over time, over instruments and over groups of respondents (p. 117). Thus, reliability means generalizability and replicability.

To statistically ensure the reliability of the study tools (the two tests and the questionnaire), it is calculated using Cronbach's Alpha measurement which is a statistical measure in the Statistics Package for Social Sciences (SPSS). The following tables show the results of the three tools (student's attitude questionnaire, student's pre-test and post-test):

Table 3.7: The Reliability Coefficient of Student's Attitude Questionnaire

Reliability Statistics

Cronbach's Alpha	No of Items
0.88	30

Table 3.8: The Reliability Coefficient of Student's Pre-test

Reliability Statistics

Cronbach's Alpha	No of Items
0.84	30

Table 3.9: The Reliability Coefficient of Student's Post-test

Reliability Statistics

Cronbach's Alpha	No of Items
0.89	30

To statistically measure the reliability of the student attitude questionnaire and the two tests (the pre-test and post-test), the SPSS is used. From the SPSS main menu, *analyse* is chosen then *scale* and then *reliability analysis* is selected. The result is the above tables.

It is worth mentioning that the higher the alpha, the higher the reliability. The tables above showed high alpha, (0.85), (0.84), and (0.89) which reflect the tools' high level of internal consistency, so that the instruments are ready to be administered for data collection.

Concerning the interview reliability, as the reliability of qualitative instruments cannot be measured statistically, three experts in applied linguistics (Ph.D holders) were interviewed to ensure the interview reliability. This has resulted in reducing bias on the side of the interviewer as well as the content of the interview.

3.5 The Procedure

The tools of primary data collection of this dissertation are two tests, an interview and a questionnaire. Concerning the tests (Appendices 1 and 2), 30 questions about the use of delexical verbs (make, do, take and have) were designed by the researcher as an area of difficulty to EFL learners. 60 students (37 males and 23 females) were selected as the participants of the study. They were divided into two groups; the experimental group (17 males and 13 females) and the control group (20 males and 10 females). After teaching the two groups, the experimental group (using corpora) and the control group (following traditional strategy), another test (post-test) on collocations was administered to them.

The researcher adopted mixed method approach to data collection, a procedure in which the researcher mixes both qualitative and quantitative data so as to provide a comprehensive analysis of the problem under investigation.

Concerning the student attitude questionnaire (Appendix 3), the questionnaire was administered to the experimental group subjects (17 males and 13 females) for

they have already taken the experiment (learning collocations through corpora) so that they can evaluate the effectiveness of corpora in learning collocations.

Regarding the interview (Appendix 4), after being piloted by the main supervisor of the study and other five experienced professors (Ph.D holders), five interviewees were carefully selected based on their knowledge of corpus linguistics, their experience, and the willingness to volunteer. Then they were interviewed about the use and impact of corpora in teaching English language, especially collocations. The interviewees were English as a foreign language academicians teaching at different universities in Khartoum State. Finally, the interviews were transcribed, coded into themes and then analyzed qualitatively by the researcher so as to obtain valid results.

For the purpose of collecting the study data, permission from the secretary of academic affairs and the head of Biomedical Engineering Department was taken after being enlightened about the purpose of the study and the time required to collect the data. It took about 10 days to get the permission. Immediately after the approval, students were enlightened about the study and the need for their participation in the data collection process. About two days after ensuring students' participation in the study, the pre-test was administered to 60 students at Room 12 and they were directed by the researcher on how to do it. The test was invigilated by the researcher, and it took about 30 minutes to finish. After completing the test, the papers were collected and marked by the researcher.

To fully measure the effectiveness of corpora, respondents of the pre-test were divided into two groups; the control group and the experimental group (30 students for each). For about two weeks, the two groups were taught collocations (delexical verbs) differently. Concerning the control group, they were taught using non-

corpus-based strategy (chalk and board). After that, the post-test was administered to them. On the other side, the experimental group were firstly taught and enlightened about what corpus is and how to practically use it in language learning and investigation. Learners were introduced to British National Corpus user's interface, *List*, *Chart*, *Collocates*, *Compare* and *KWIC*. For the purpose of investigating the delexical verbs *make*, *do*, *have* and *take*, more emphasis was placed on *Collocates* to investigate the right noun collocates of each delexical verb and comparing the results learners have reached for each pair of the verbs (*make* with *do* and *have* with *take*). In addition, using *KWIC* is also emphasized as it provides the context in which each delexical verb occurs so that learners can enhance their knowledge of the four delexical verbs. For making best use of the corpus and to avoid being confused by the large amount of search results, learners were told to exactly limit their search results and purposes.

Figure 3.2: The British National Corpus User's Interface
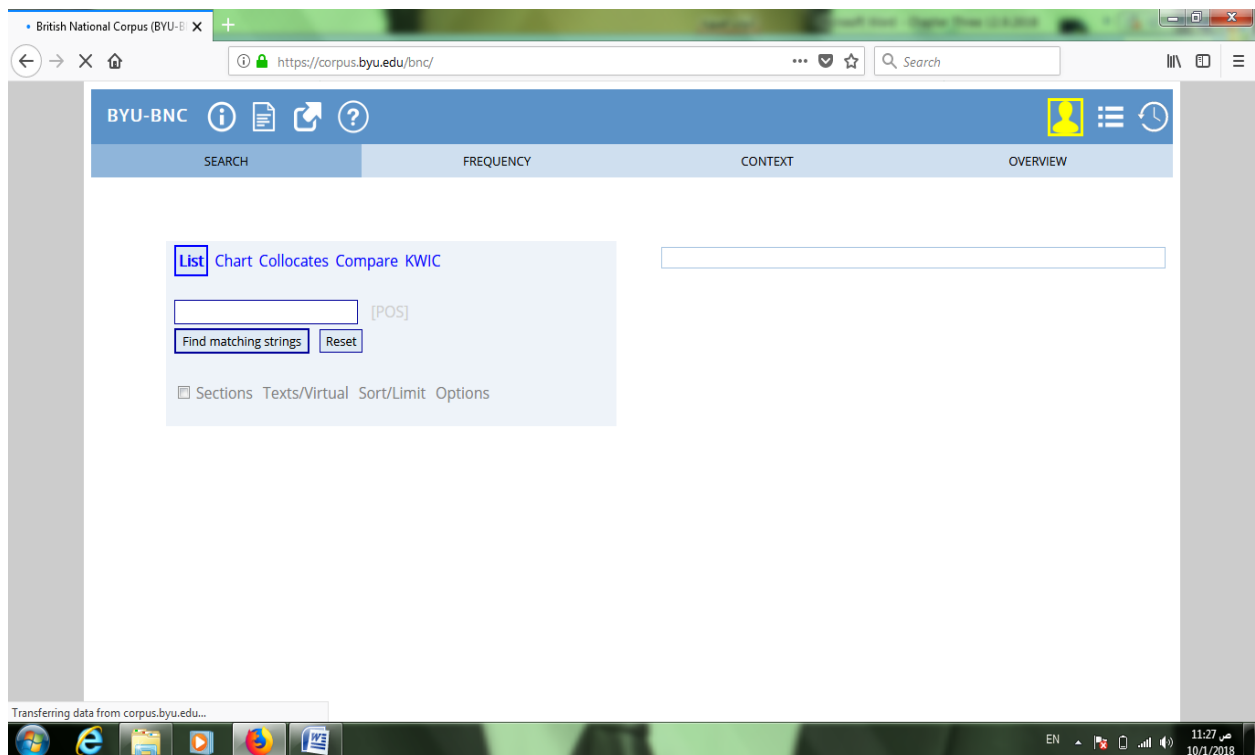
After ensuring students' readiness and ability to use the corpus, they were taught the delexical verbs and their collocates using British National Corpus (BNC) in Room B. It took about two weeks to enlighten them on what corpora are and how to use them as well as to teach them the four verbs. Immediately after teaching the experimental group, the post-test on the same delexical verbs was administered to them. It took place at Room 12. It also took about 30 minutes and it was also invigilated by the researcher. Then the papers were collected and marked by the researcher to be ready for the statistical analysis.

Regarding the studen6 attitude questionnaire, directly after finishing the post-test, the experimental group was asked to respond to the questionnaire as they know what corpora are and how to use them in language learning and investigation. They were enlightened about the significance of their participation in responding to the questionnaire. Then the questionnaire was distributed to them. They were directed by the researcher on what they have to do and the importance of being honest and transparent while responding to the questionnaire statements. While the students were responding to the questionnaire, the researcher was present to clarify any unfamiliar concept and to make sure that the students are honestly and accurately responding to the questionnaire statements. Some respondents of the experimental group were not present when the questionnaire was administered. These absentees were contacted and the questionnaire was sent to them through WhatsApp application to be done and submitted later on. Three days later, all the questionnaire papers were collected.

Concerning the interview of the study, after being piloted by the experts including the main supervisor of the study, five English language teachers were contacted and asked to participate in the study. These respondents teach English language at different universities and university colleges in Khartoum State. They

were selected according to their knowledge in corpus linguistics and volunteering to participate in the study. It took about three weeks to interview all the subjects because of the fact that they teach at different institutions in different places in Khartoum City and the fact that they are all busy.

3.6 Summary of the Chapter

This chapter has described the design of the study, the population, the sampling procedure, the sample, and the instruments employed in the process of data collection so as to achieve the research goals. The following chapter will be devoted to analysis of data, results and discussion.

Chapter Four

Data Analysis, Results and Discussion

Chapter Four

Data Analysis, Results and Discussion

4.0 Overview

This chapter presents data analysis and results of the two tests (pre-test and post-test), the student attitude questionnaire, and the teacher interview. The results of this research will be discussed in comparison to the previous studies and their findings presented in chapter two. In addition, based on the study findings, the research hypotheses will be verified.

4.1 Analysis of the Pre-test and Post-test

To obtain accurate results from the pre-test and post-test, the Statistical Packages of Social Sciences (SPSS) program is used and the following table is provided.

Table 4.1 One-Sample Test

	Test Value = 0					
	T	Df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Pre-test	36.222	59	.000	18.80000	17.7614	19.8386
Post-test	44.521	59	.000	20.73333	19.8015	21.6652

The above table shows the performance of the two groups (the experimental group and the control group). As demonstrated, there is a significant statistical difference between the two groups which means that the experimental group performed significantly better than the control group (sig less than .5). This result proves the fact that corpora are effective strategy in learning collocations in English language.

4.2 Analysis of the Student Attitude Questionnaire

For the purpose of getting accurate results from the analysis of the student attitude questionnaire, the Statistical Packages of Social Sciences (SPSS) is employed and the following figures are obtained.

Figure 4.1: Responses of participants on collocations and developing learners' receptive knowledge.

Learning the language through collocations easily develops students' receptive knowledge.
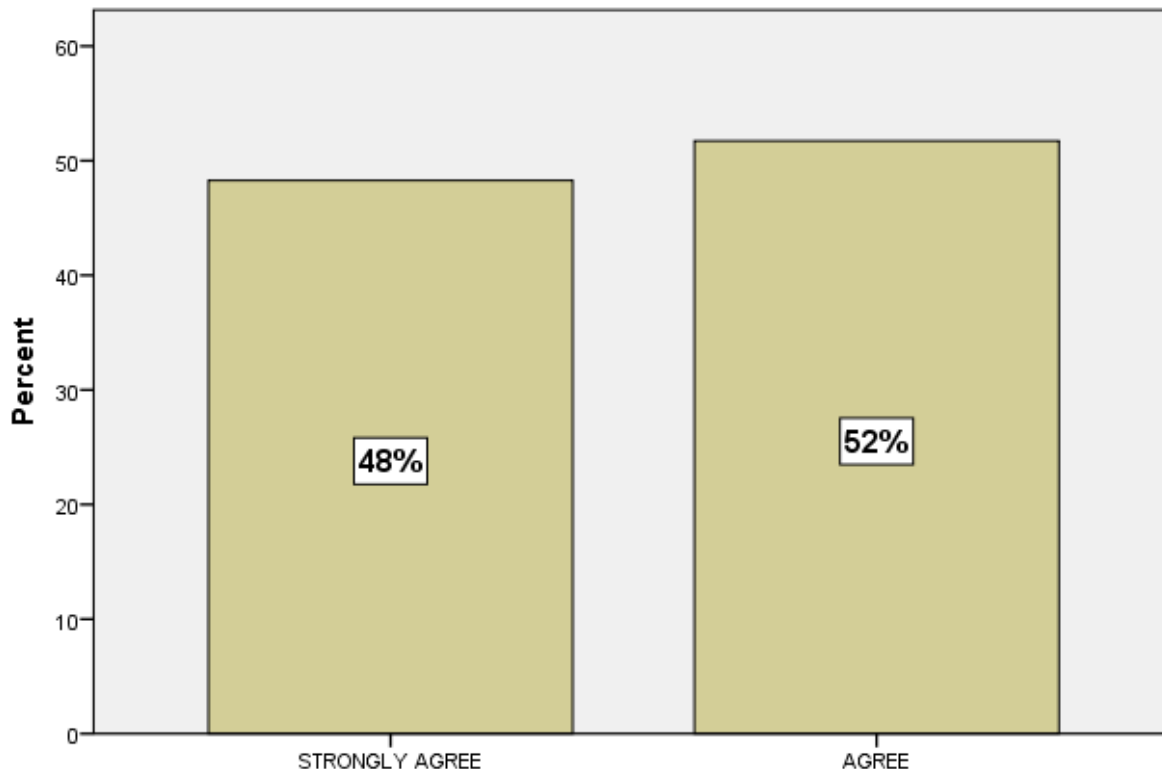
Figure 4.1 shows responses of participants on learning the language through collocations and developing students' receptive knowledge. As demonstrated, all the respondents (100%) agreed that the language is to be learnt via collocations so that students' receptive knowledge (their listening and speaking skills) can be developed. Indeed, learning the language in chunks reduces processing efforts and hence leads to accurate and native-like language production. In addition, it makes learning the language interesting.

Figure 4.2: Participants' responses on collocations and developing learners' productive knowledge.

Learning the language via collocations effectively develops students' productive knowledge.
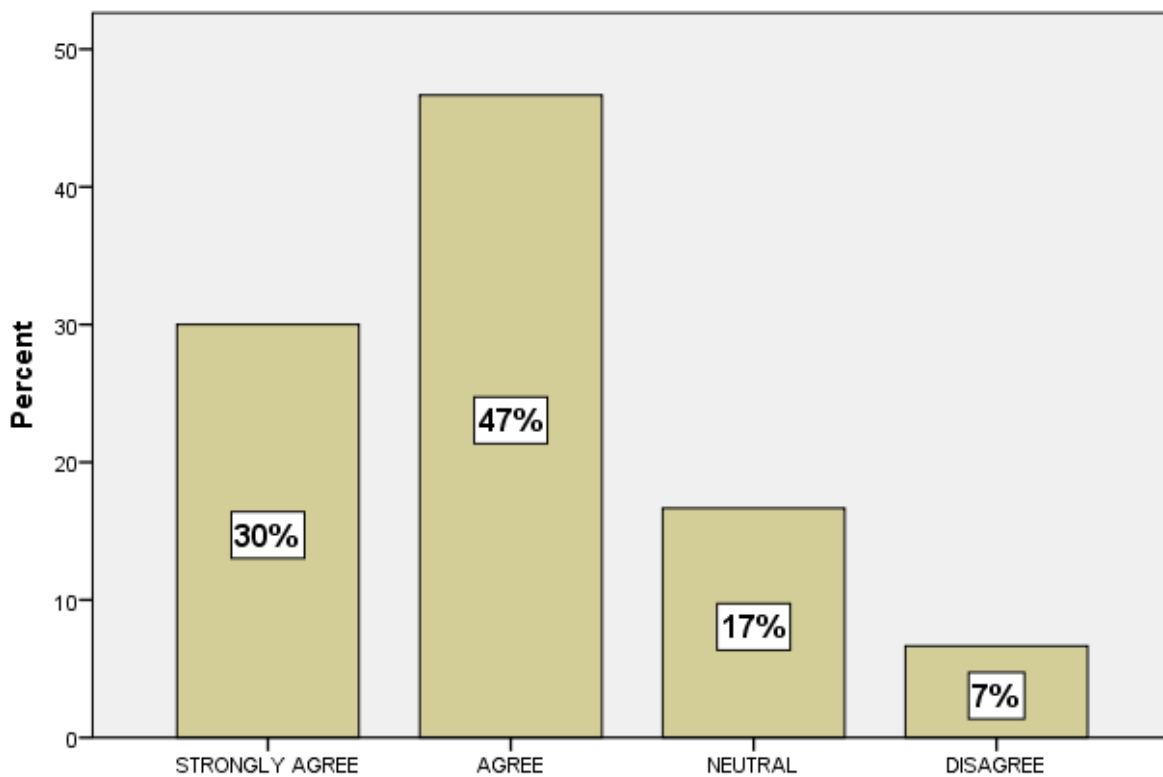
Figure 4.2 displays responses of subjects on developing students' productive knowledge through collocation learning. As shown, the vast majority of respondents (77%) agreed that learning the language through collocations

effectively develops students' productive knowledge i.e. their speaking and writing skills, while only 7% disagreed. It is worth notifying that the rest of participants (17%) reacted with *neutral*. It is strongly proved as shown in the figure that learning the language through collocations effectively develops learners' productive knowledge rather than learning the language through individual words. As it has been proved in figure 4.1 that learning the language through collocations develops learners' receptive knowledge, it is obvious that their productive knowledge develops as well.

Figure 4.3: The inappropriate use of collocations and EFL learners' native language.

The inappropriate use of collocations is attributed to the influence of EFL learners' native language.
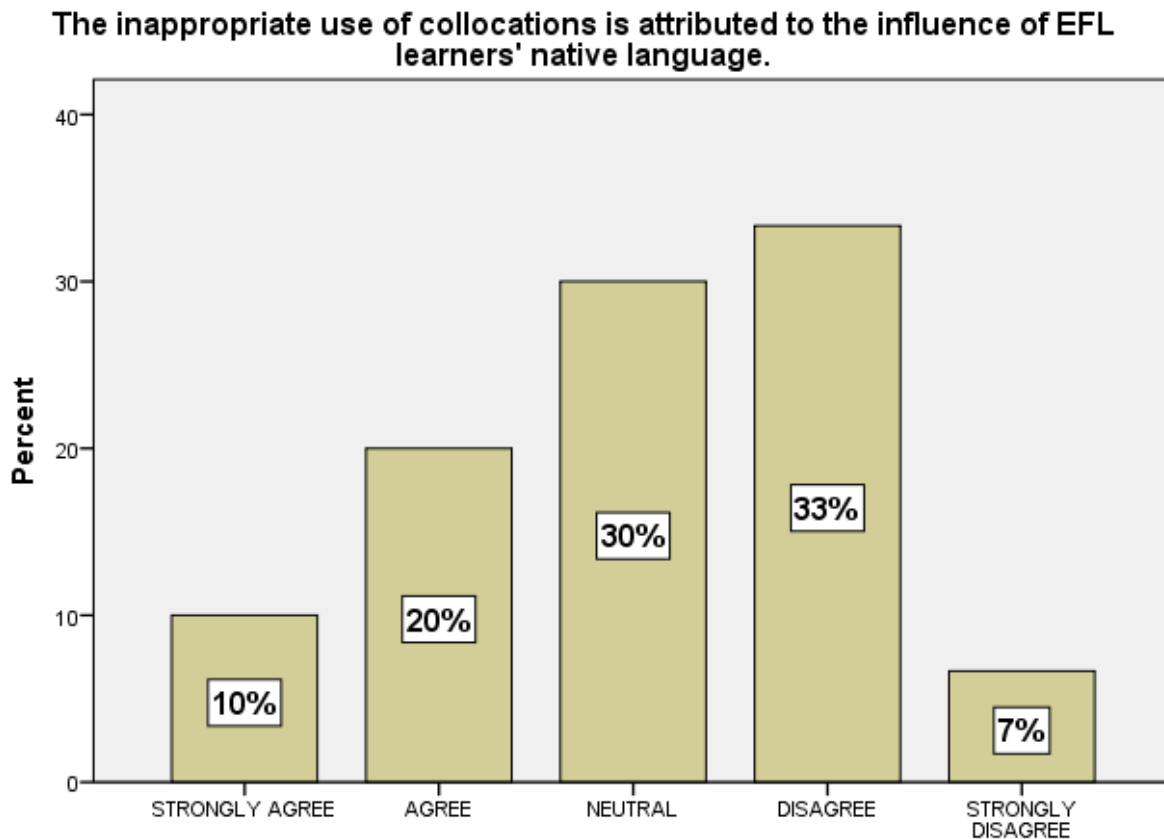
Figure 4.3 presents the inappropriate use of collocations and the influence of learners' native language. Concerning whether the native language of learners

influences the inappropriate use of collocations, *disagree* received the highest score (33%) which justifies that the native language of learners does not directly influence the appropriate use of the target language collocations (English language). Followed by *neutral* (30%) which suggests that learners are not certain whether learners' native language influences the inappropriate use of collocations. On the other side, the other three options recorded 20% and less. In the researcher's own calculation, one of the major factors behind the inappropriate use of collocations is the mother tongue interference. This happens because learners unconsciously transfer the rules of the mother tongue to the target language. As the case of Sudanese undergraduates in transferring the Arabic rules to English language when dealing with collocations, especially delexical verbs. For example, many of them mix *make* with *do* and *have* with *take* in the delexical structure.

Figure 4.4: Undergraduates' collocation learning difficulty and how they (collocations) are taught at schools.
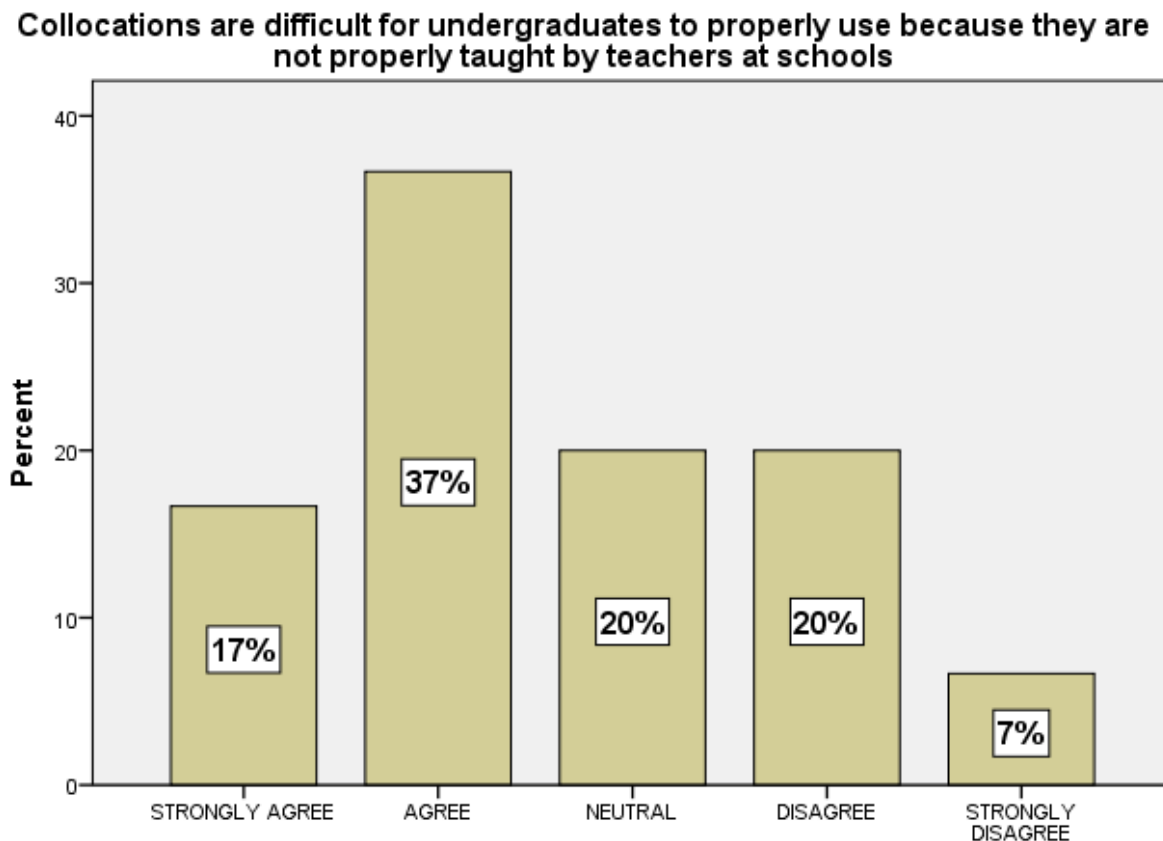
As figure 4.4 shows, *strongly agree* and *agree* have recorded a percentage of 54% but *disagree* and *strongly disagree* have scored a percentage of 27% which clearly proves the fact that collocations are difficult for undergraduates to properly use because they are not properly taught by teachers at schools. This suggests that collocational knowledge is accumulative knowledge. In other words, collocations should be given enough attention and should properly be taught from the early stages of language learning i.e. the primary and the secondary schools. Not doing so results in poor collocational knowledge and hence poor language proficiency.

Figure 4.5: Undergraduates' collocation difficulty and how they (collocations) are taught at university.
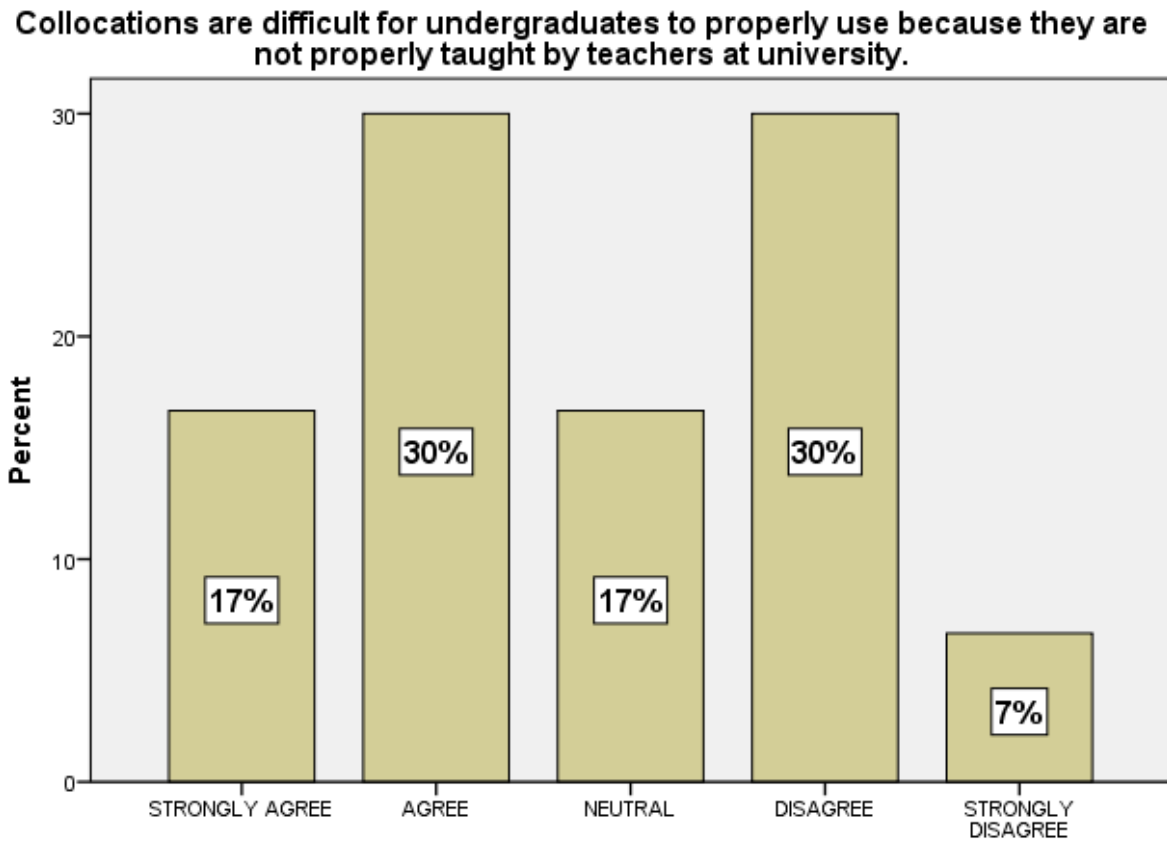
Concerning collocations difficulty and the way they are taught at university, *strongly agree* and *agree* have recorded 47% while *disagree* and *strongly disagree* have scored 37%. To be noted that 17% of the respondents remained neutral. This is obviously suggests that collocations are not properly taught at university. This could be attributed to many factors such as the teaching strategies adopted, the syllabuses students have been exposed to, teacher's motivation, and many other factors related to the teaching and learning process.

Figure 4.6: Participants' responses on collocations and the attention given to them in the classroom.

Collocations are confusing for students because they are not given enough attention in the classroom.
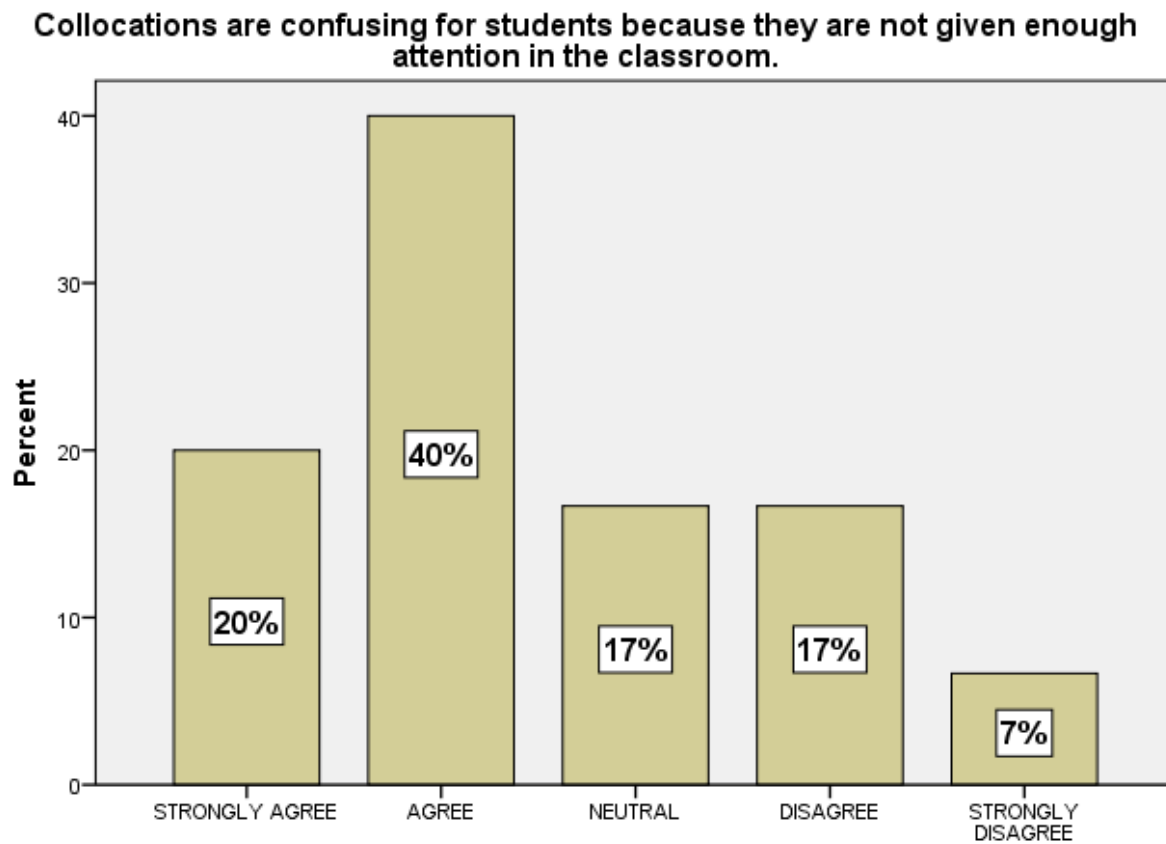
Concerning collocations and the attention given to them in the classroom, it is clearly proved in figure 4.6 above that 60% of the respondents agreed that collocations are not given enough attention in the classroom, while 24% of them disagreed. The rest of subjects (17%) remained in the middle. As clearly demonstrated by the figure above that collocations are not actually given enough attention in the classroom despite the fact that collocational competence is essential for achieving high competency in the language. In addition, even if collocations are given little attention, it is given to their width rather than their depth. In other words, the focus is given to the quantity of collocations rather than their quality.

Figure 4.7: Responses of participants on understanding collocations and the role of learners' L1 and English language.

Understanding collocations properly requires solid cultural and linguistic background of learners' L1 (Arabic) and English language.
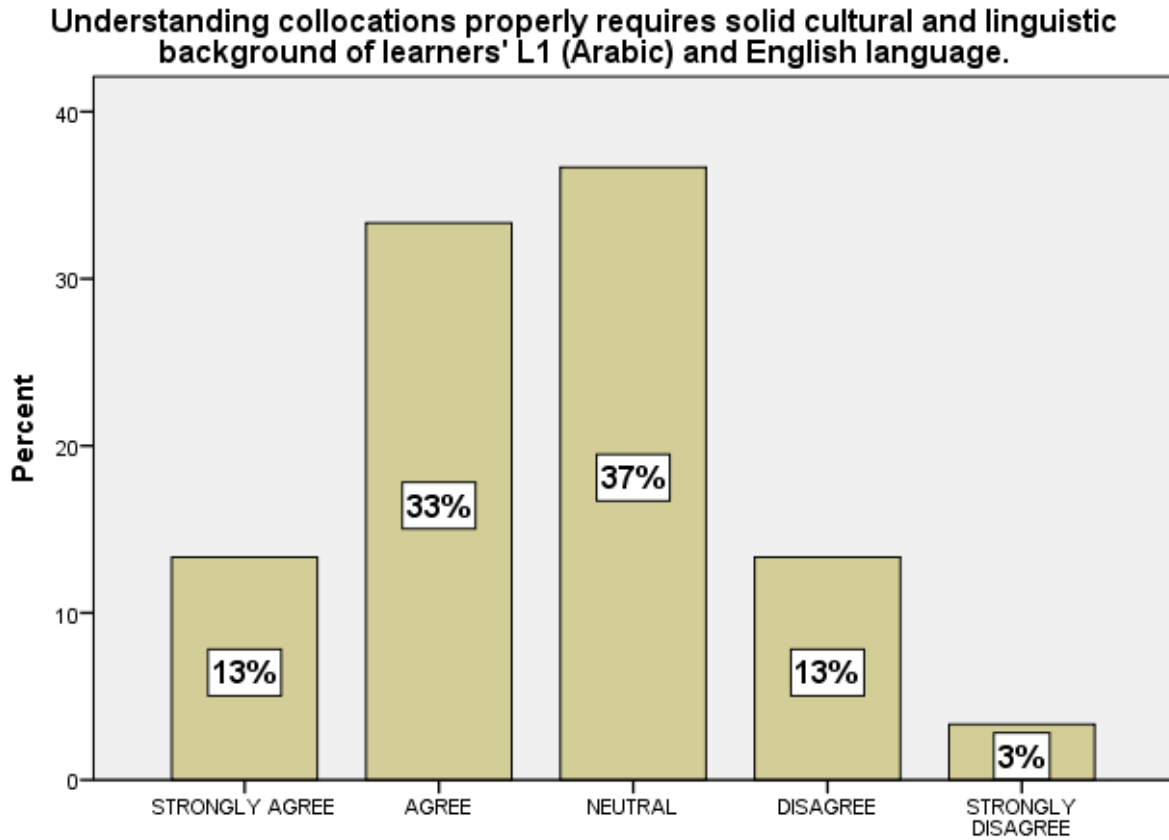
It is of no doubt that understanding collocations properly requires solid linguistic and cultural background of both learners' native language and English language. This claim is proved by figure 4.7 above by the fact that 46% of the participants responded with *strongly agree* and *agree*, while 16% responded with *disagree* and *strongly disagree*. It is worth mentioning that the other 37% of respondents chose to stand in the middle. This suggests that knowing the learners' first language (Arabic) and English language and their cultural background is beneficial for contrastive purposes. In other words, the similarities between the native language of learners (Arabic) and English language ease the process of learning the latter as well as the dissimilarities as they make learners aware of the

differences between the two languages and hence make the process of learning English language easy.

Figure 4.8: Using collocations and the role of learners' L1 and English language.

To properly use collocations, learners should have cultural and linguistic background of both their L1 (Arabic) and English language.
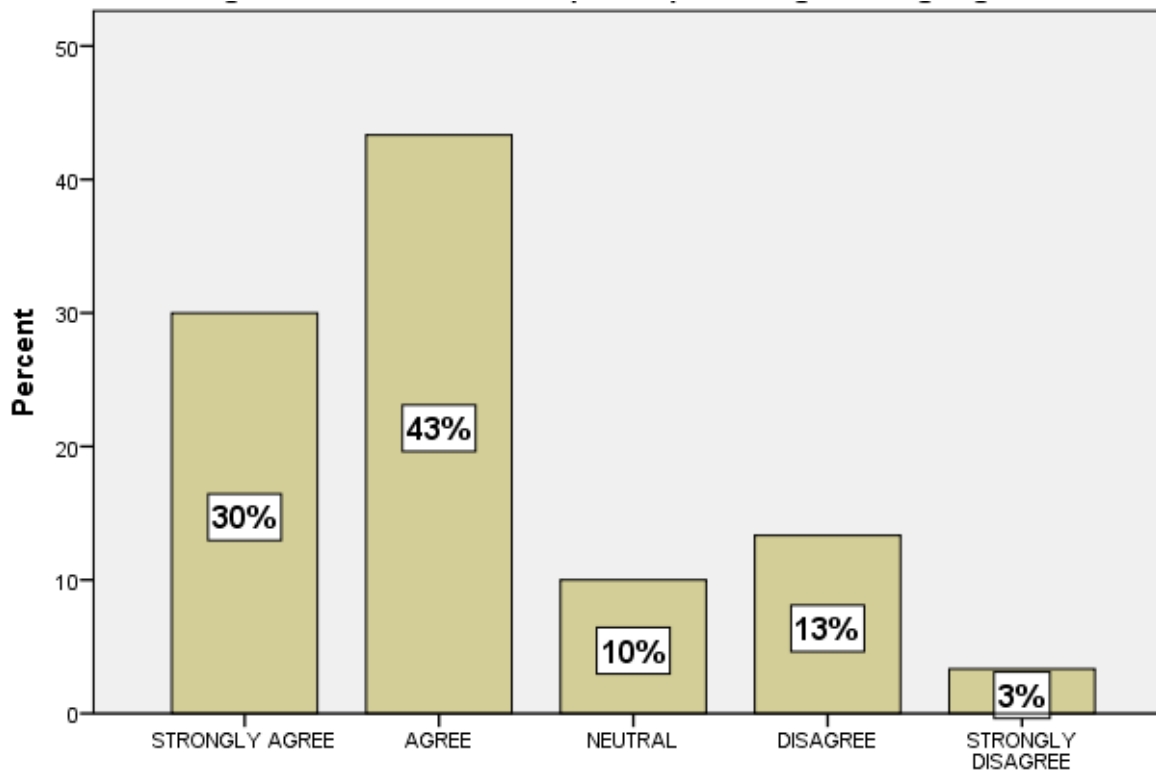
As understanding collocations properly requires linguistic and cultural background of both learners' first language and English language as discussed in figure 4.7, it is also essential that for learners to effectively use collocations they have to have linguistic and cultural knowledge of English language besides their native language. This is clearly demonstrated by the fact that *strongly agree* and *agree* have recorded a slightly high percentage (73%) while *disagree* and *strongly disagree* received 16%. To be noted, *neutral* received only10%. This justifies that knowing the similarities and differences between the native language (Arabic) and

the target language (English) makes learning the language easy as explained in figure 4.7.

Figure 4.9: Participants' responses on lack of practice and its effect on learners' collocational competence.

Learners' lack of practice is one of the factors affecting their collocational competence.
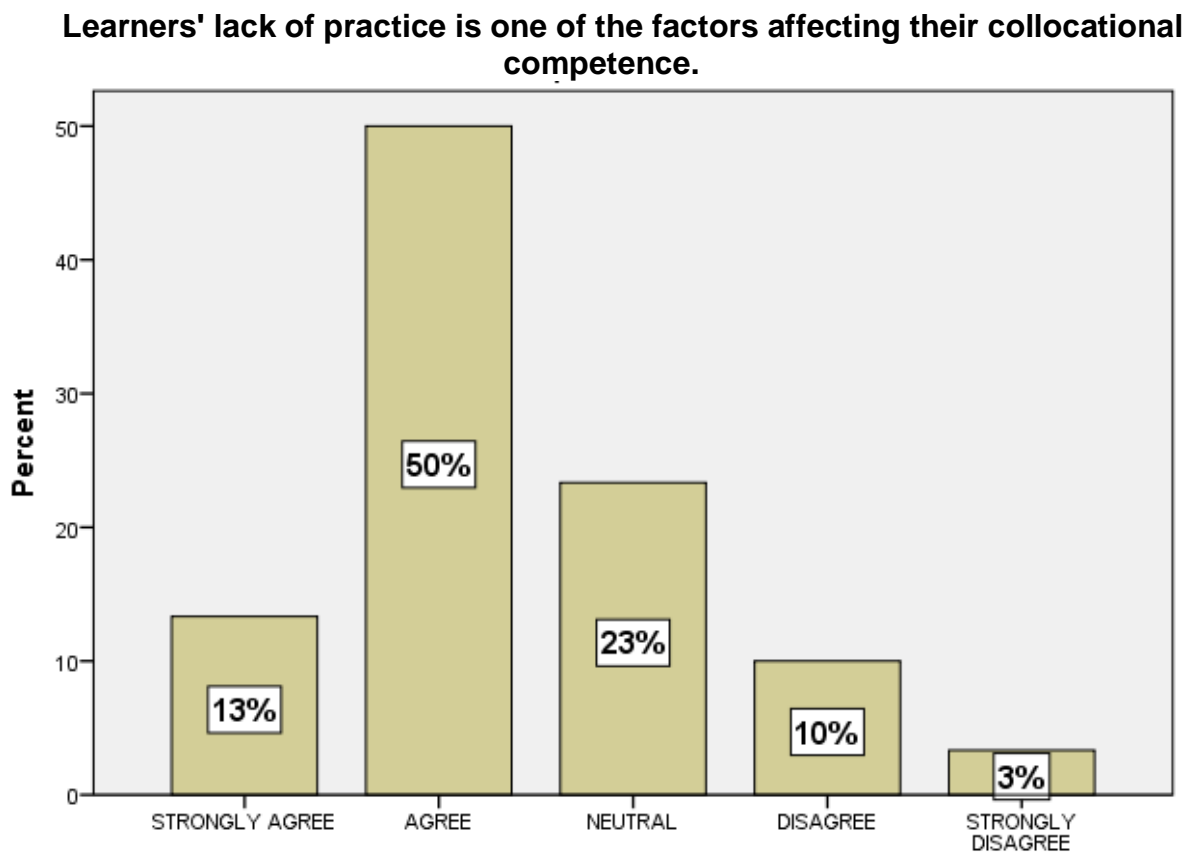
One of the most essential factors for learning any language is practice. If there is no practice there is no successful learning process. This fact is clearly demonstrated by the high percentage scored by *strongly agree* and *agree* (63%). This demonstrates that 63% of the participants are aware of the significance of practice for developing learners' collocational competence and consequently their competence in the language. On the other side, only 13% of the participants chose *disagree* and *strongly disagree* while the remaining 23% of them were neutral.

Figure 4.10: Responses of participants on being aware of collocations and their proper use.

Some undergraduates are fully aware of the significance of collocations, so they properly learn them.
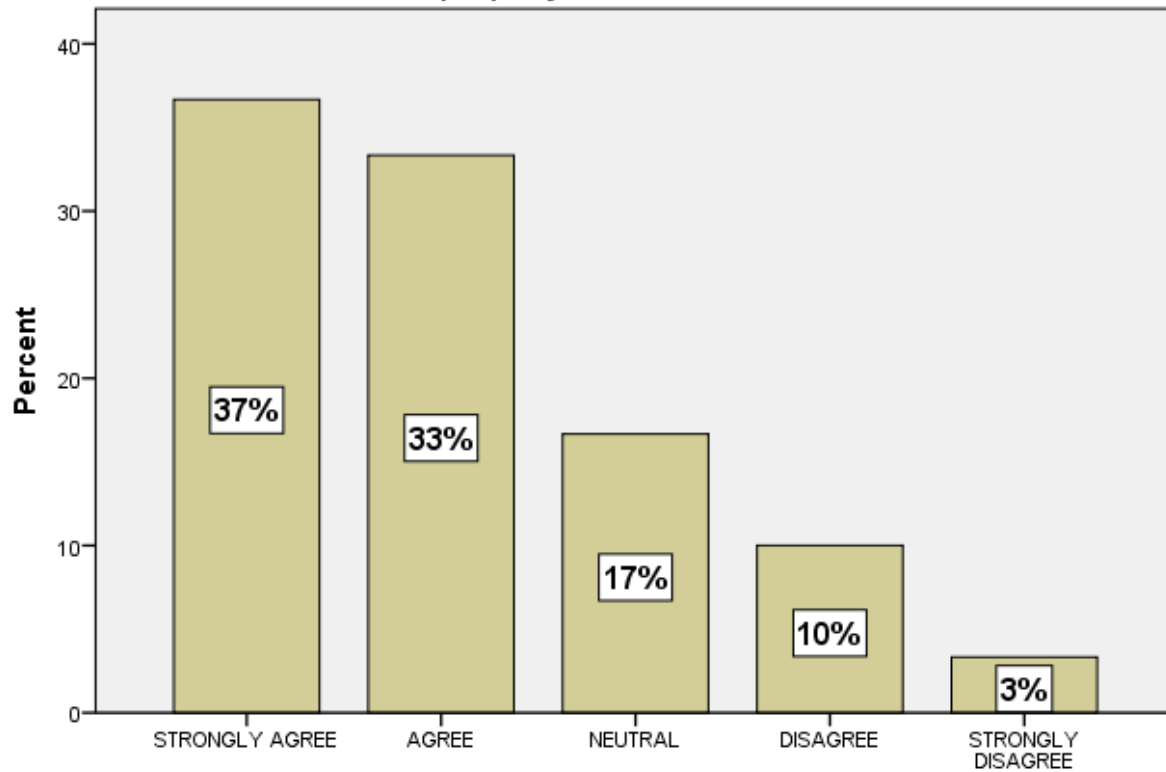
Regarding undergraduates' awareness of the significance of collocations, the vast majority of respondents (70%) agreed that some undergraduates are fully aware of the significance of collocations, so they properly learn them. On the other side, only 13% of the participants disagreed. It is worth mentioning that those who chose to stand in the middle are 17% of the respondents. This indicates that being aware of collocations and their significance is vital to achieving a proper learning process and hence achieving high level of language proficiency. The more aware of collocations the students are the better learning process and thus the better use of collocations.

Figure 4.11: Responses of participants on being aware of collocations and their proper use.

As demonstrated above, a relatively high score is recorded by *strongly agree* and *agree* (37% for each) while *disagree* and *strongly disagree* received quite low percentages (7% and 3% respectively). To be noted, 17% is given to *neutral*. This proves that the vast majority of participants agreed that some undergraduates are fully aware of the significance of collocations, so they properly use them. It can be deduced that being aware of collocations and their significance is crucial to achieving effective learning process as discussed in table 4.10. As a result, appropriate and native-like use of collocations could be achieved.

Figure 4.12: Responses of participants on whether undergraduates are exposed to collocations at schools.
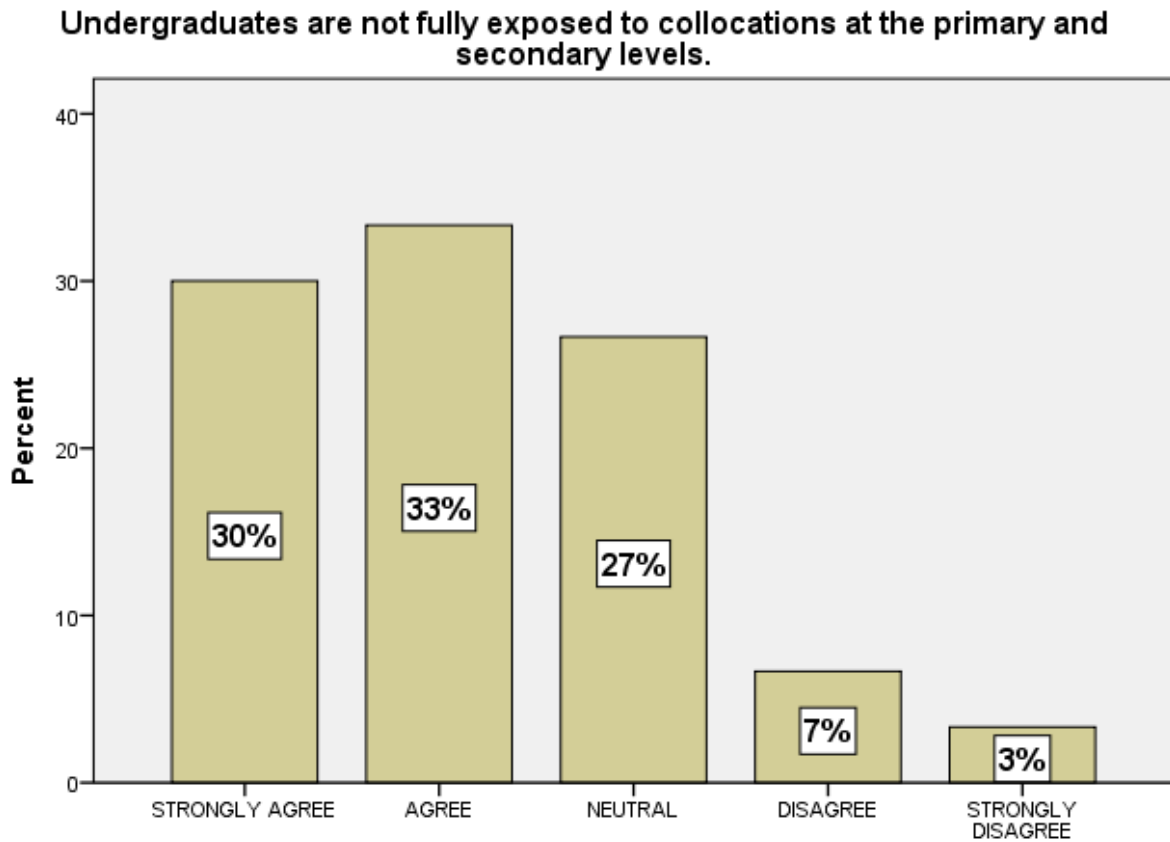
Figure 4.12 above projects the participants and their views on whether undergraduates had enough exposure to collocations at the primary and secondary schools. As the figure demonstrates, 63% of the respondents agreed that students have not had enough exposure to collocations whereas only 10% disagreed. It is worth notifying that the remaining 27% of the subjects were neutral. It could be claimed that being not fully exposed to collocations at schools is one of the major factors behind undergraduates' lack of collocational competence. This suggests that collocational knowledge is accumulative knowledge which should early be developed.

Figure 4.13: Respondents' opinions on collocation learning strategies used at universities and their role on developing students' collocational knowledge.

The learning strategies used for learning collocations at university are not fully effective for developing students' knowledge of collocations.
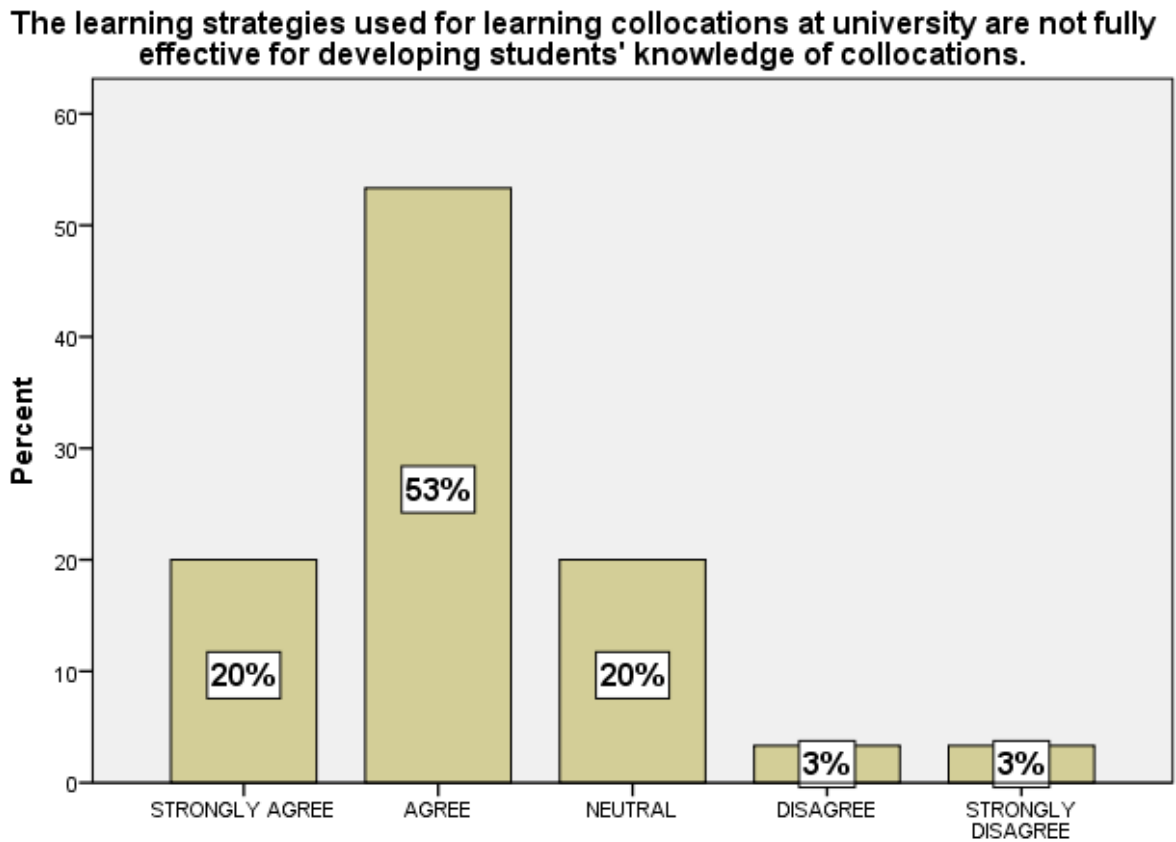
Figure 4.13 above shows respondents' views on collocation learning strategies employed at university. 73% of them responded with *strongly agree* and *agree*, 20% with *neutral* and only 6% for *disagree* and *strongly disagree*. The slightly high percentage given to strongly agree and agree substantiates that the strategies adopted for learning collocations at university are not actually effective in doing so. This shows the need for adopting effective and more beneficial strategies to language learning such as the use of corpora and corpus-based strategies.

Figure 4.14: Collocations and the university curriculum.
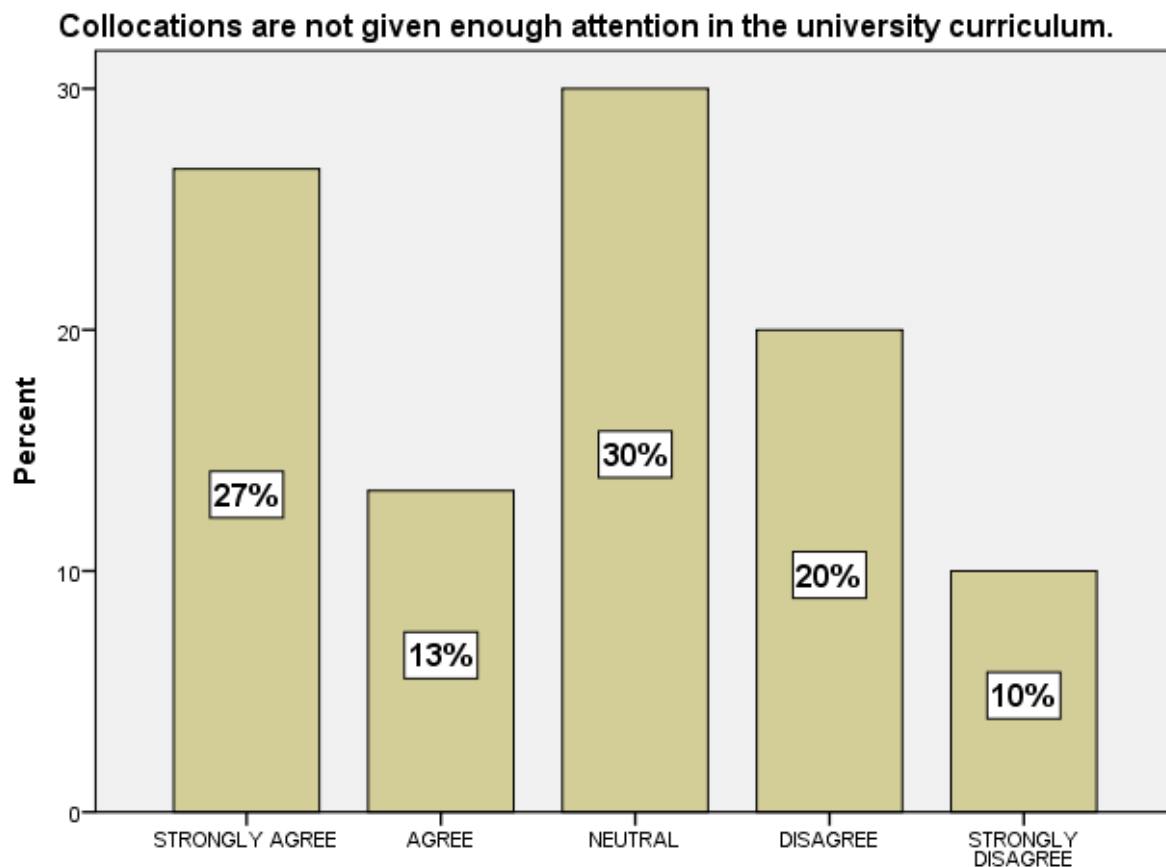
As clearly shown above, 40% of respondents agreed that collocations are not given enough attention in the university curriculum while 30% of them disagreed. The remaining 30% of the participants responded with neutral. As a matter of fact, collocations are given scant attention in the university curriculum and this little attention is given to collocation width rather than their depth, although they are vital to achieving the competence of any language. So it is suggested that corpus-based strategy is suitable for collocation learning as they enable learners get exposed to real language.

Figure 4.15: Comparing corpora to other strategies to language learning.

Learning collocations through corpora is much better than via the other learning strategies.
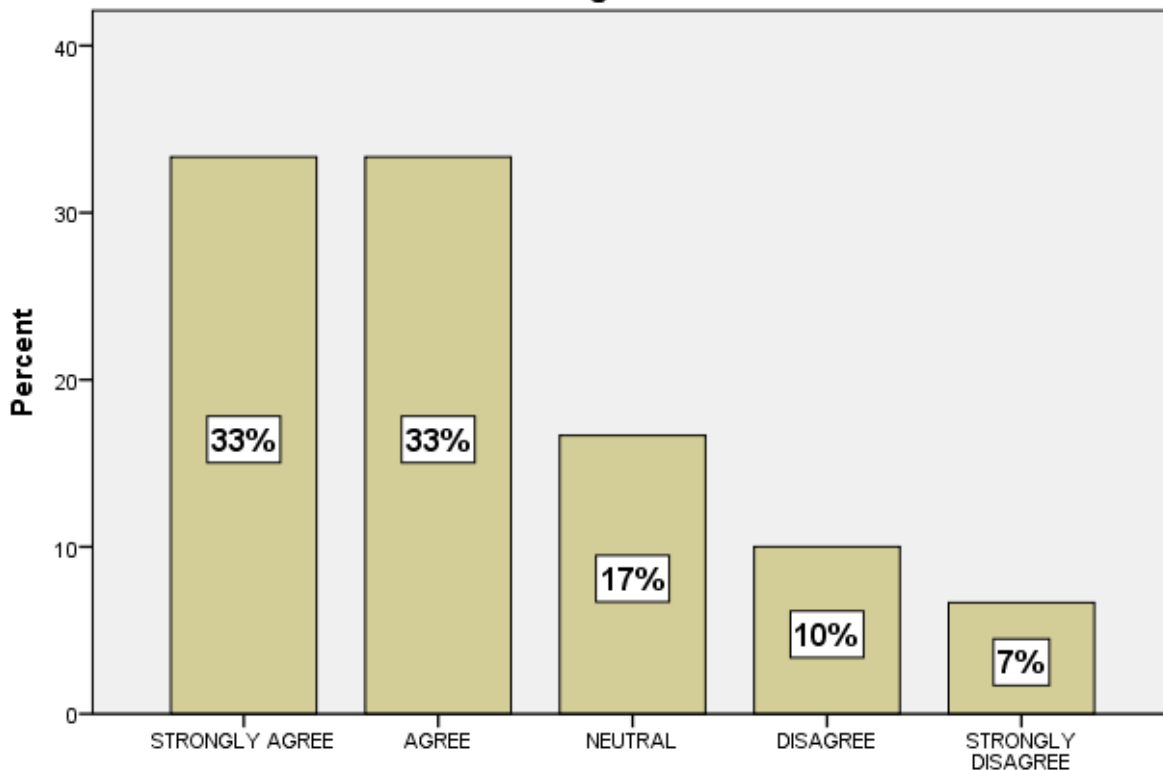
Undoubtedly, learning collocations via corpora is much better than through the other learning strategies. This claim is substantiated by the fact that 66% of the respondents reacted with *strongly agree* and *agree* as shown in figure 4.15 above. On the other hand, *disagree* and *strongly disagree* recorded 17% while *neutral* received 17%. As proved above, corpora are more effective than the other collocation learning strategies for a number of reasons; prime among them is the authentic language data provided by corpora, their inclusion of a wide range of registers, their availability and ease of accessibility and so on.

Figure 4.16: Comparing corpora to dictionaries in vocabulary learning.

The corpus is more helpful than the dictionary in vocabulary learning.
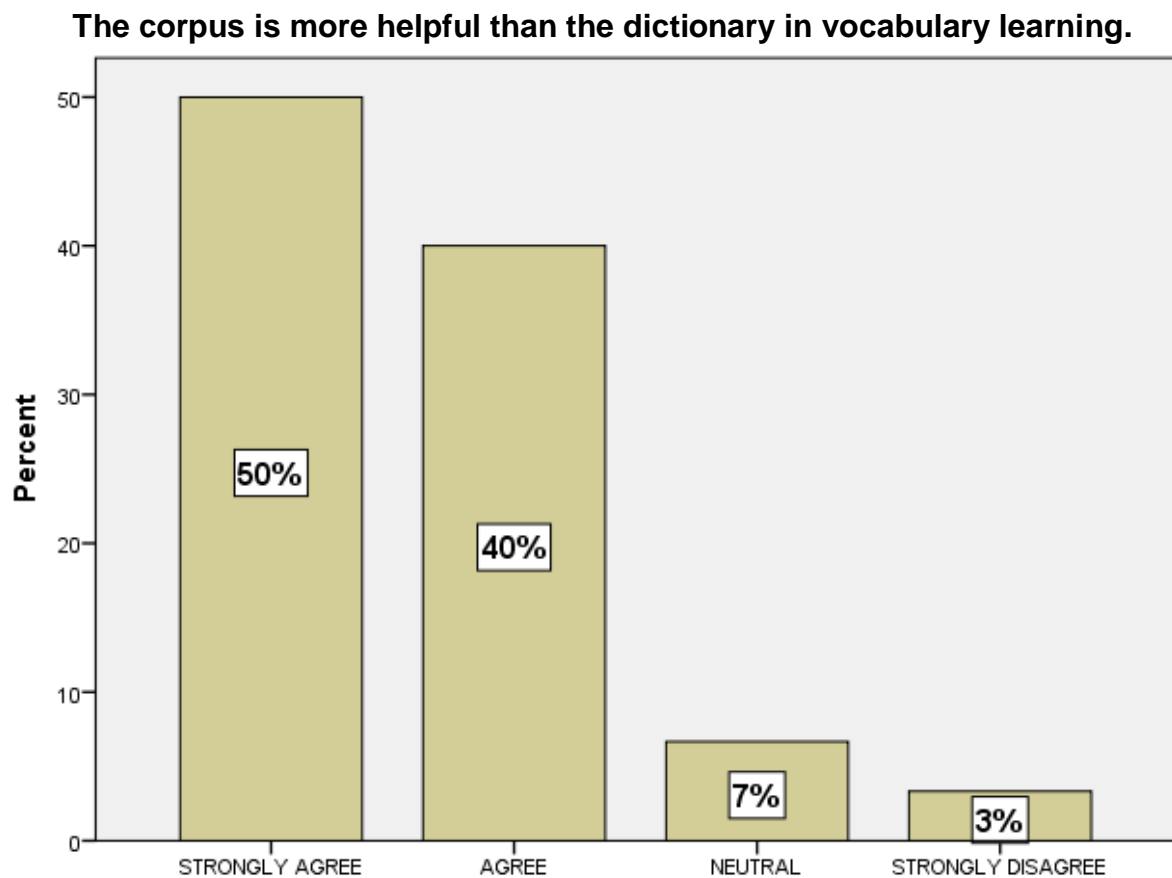
Figure 4.16 compares the use of corpus to dictionary in vocabulary learning. It is proved that corpora are better than dictionaries in vocabulary learning. This could be due to the fact that unlike dictionaries, corpora consist of large amount of language data collected from wide range of language sources. This data is used in language learning, teaching and investigation. On the other side, dictionaries are not as wide as corpora. Besides, the data in dictionaries are based on the intuitions of their writers which are not as effective as corpus data in language learning. This claim is supported by the figure above in which 90% of the participants agreed while only 3% disagreed. It is worth mentioning that *neutral* received only 7%.

Figure 4.17: Corpora in collocations learning and the authentic context they provide.
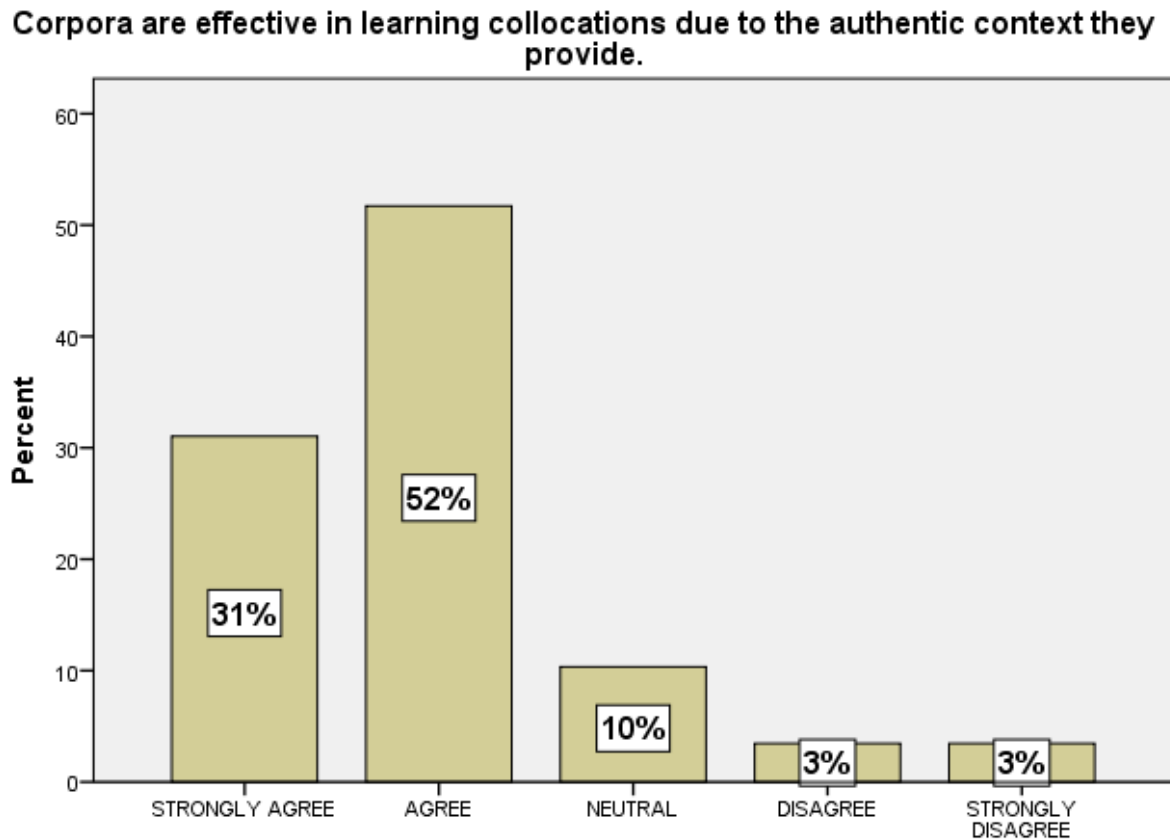
Figure 4.17 above displays collocation learning and corpora as authentic context. It is clear that 83% of the participants agreed, while only 6% of them disagreed. The other 10% of the respondents remained neutral. This is clear indication that corpora provide authentic context for learning collocations because of the fact that their data is naturally collected from different contexts of language use. These collections of language data show learners how the language is actually used by native speakers. A fact that many other strategies failed to prove.

Figure 4.18: Corpora and developing students' autonomous learning.

Corpora are helpful in collocation learning because they develop students' autonomous learning.
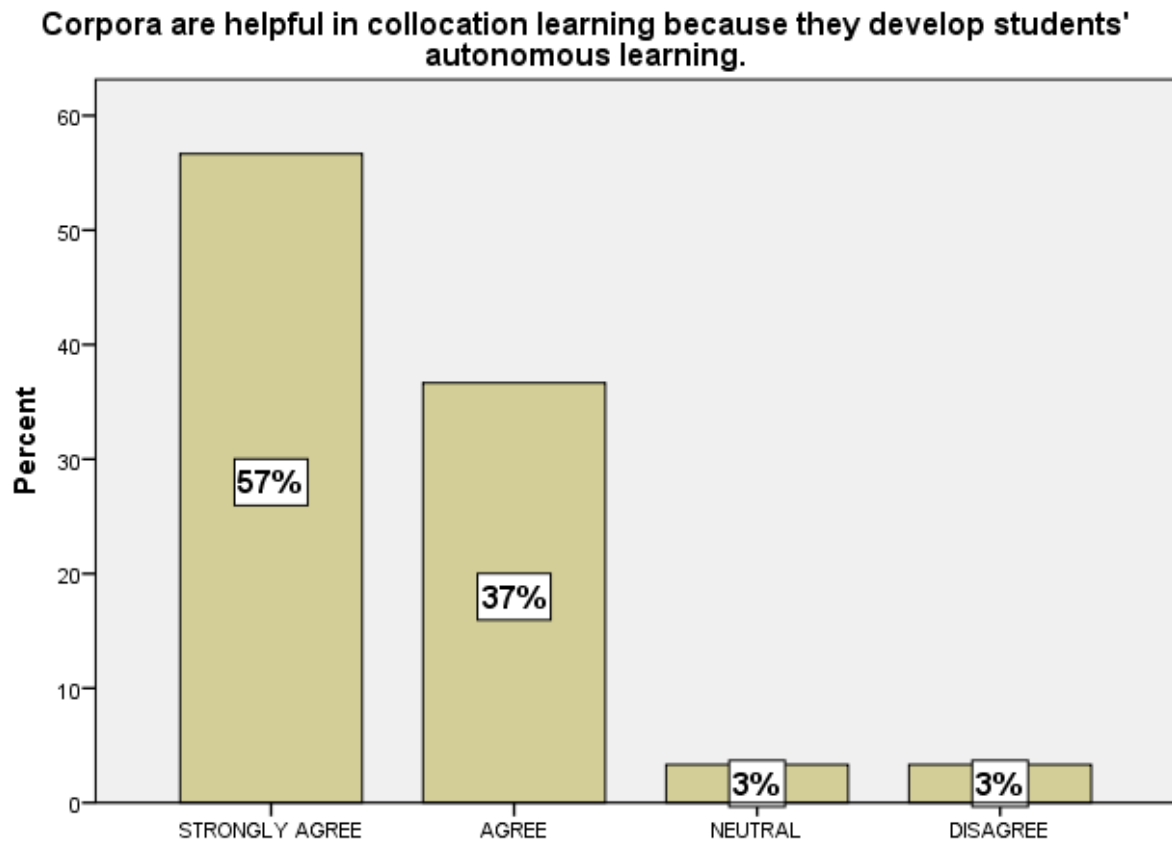
Figure 4.18 above presents corpora in collocation learning and developing students' autonomous learning. As shown, the vast majority of the participants (90%) agreed to the fact that corpora develop students' autonomous learning, while only 3% of them disagreed. To be noted, only 3% of the subjects stood in the middle. This clearly proves learners agreement upon the fact that corpora can develop their autonomous learning. This kind of learning can be developed through corpora because learners can individually access them anywhere anytime and use them in language learning and investigation.

Figure 4.19: Learning the language through corpora and the study of lexical and grammatical properties of words.

Learning collocations through corpora is beneficial as corpora enable the study of the lexical and grammatical properties of words.
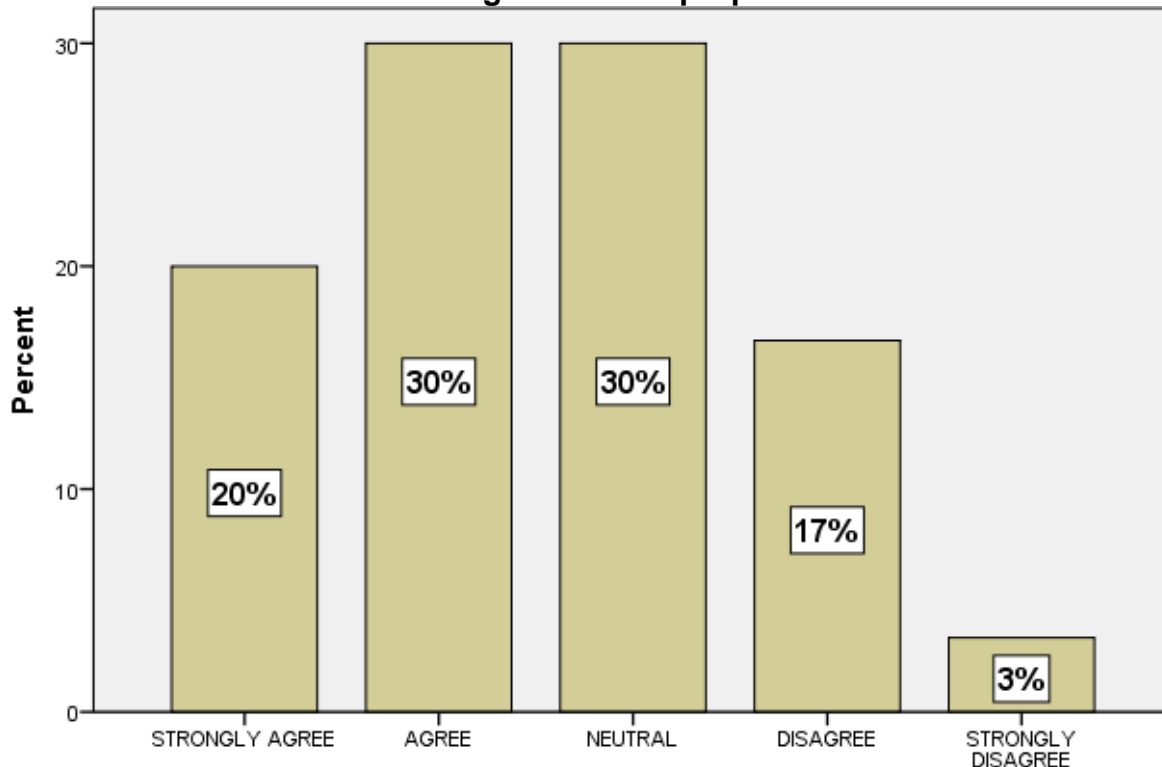
In fact, corpora have the potential of investigating the lexical and grammatical properties of words as they show how words are used within context. This is proved in figure 4.19 above by the fact that 50% of the subjects agreed, while 20% disagreed. It is worth mentioning that 30% of the participants reacted with *neutral*. With the advent of computers, large amount of language data can be stored for language analysis and investigation. Using this data, many linguistic features can be investigated such as the lexico-grammatical features of words. As grammar and semantics are complementary, each one can be used to investigate the other. For instance, two nearly synonymous words can be investigated by considering their associations with different grammatical structures, while two nearly synonymous

grammatical structures can be investigated by considering their associations with different collocations and classes of words.

Figure 4.20: Corpora and investigating senses and uses of words in different registers.

Learning collocations via corpora enables investigating the uses and senses of words in different registers.
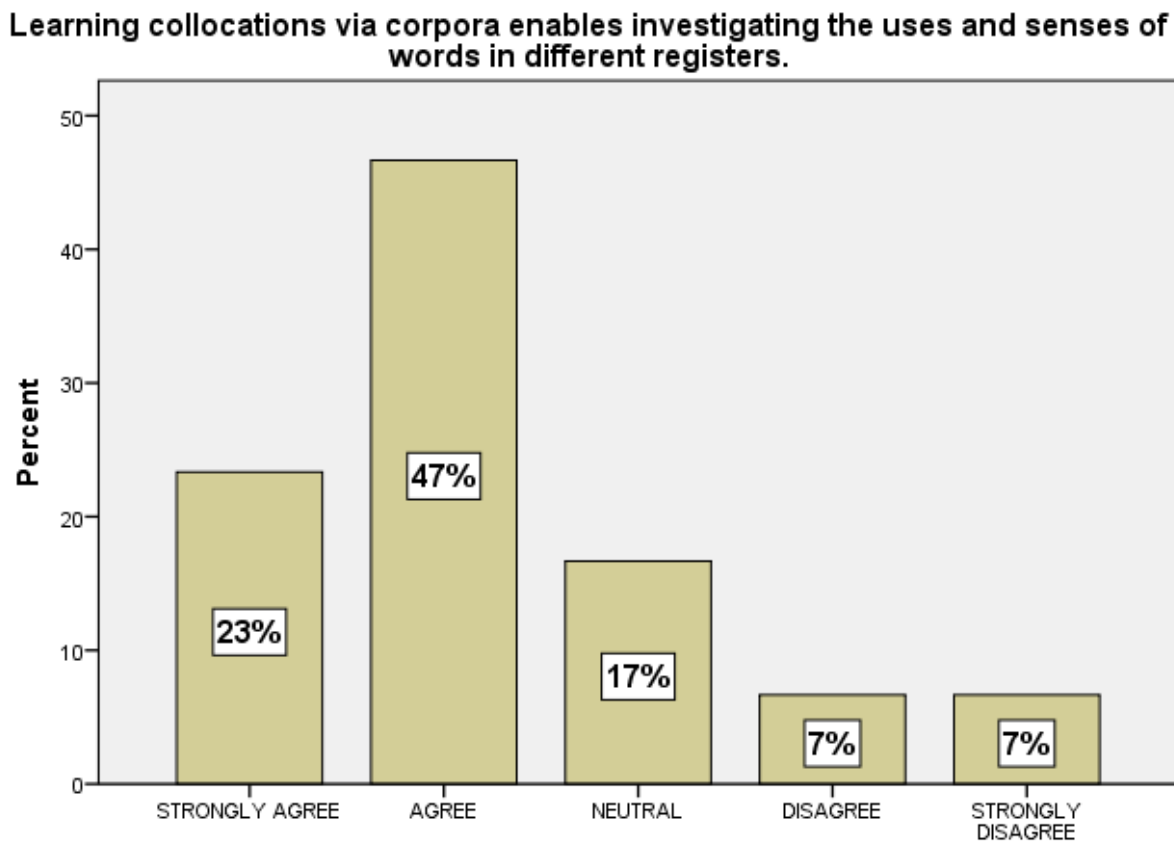
As shown in the figure above, 70% of the respondents agreed to the fact that corpora have the potentiality of investigating uses and senses of words in different registers, while 14% responded with *disagree*. The remaining 17% is dedicated to *neutral*. It is true that through corpora, word uses and senses can be investigated because corpus data are collected from different language registers meant to show how language is actually used by native speakers in different contexts of use. This fact turns out to be an area of strength of corpora because they provide language users with vital knowledge of how words are differently used in different contexts

and consequently develop their knowledge of collocations. This can be done via concordance lines through which left and right collocates of words can be investigated. It is true that concordance lines provide huge amount of results for the word under investigation which can be confusing to language learners specially those of low levels but this can be solved by the intervention of the teacher by limiting the search purposes and results.

Figure 4.21: Corpora and making the learning process fun and interesting.

Using corpora in collocation learning makes the learning process fun and interesting.
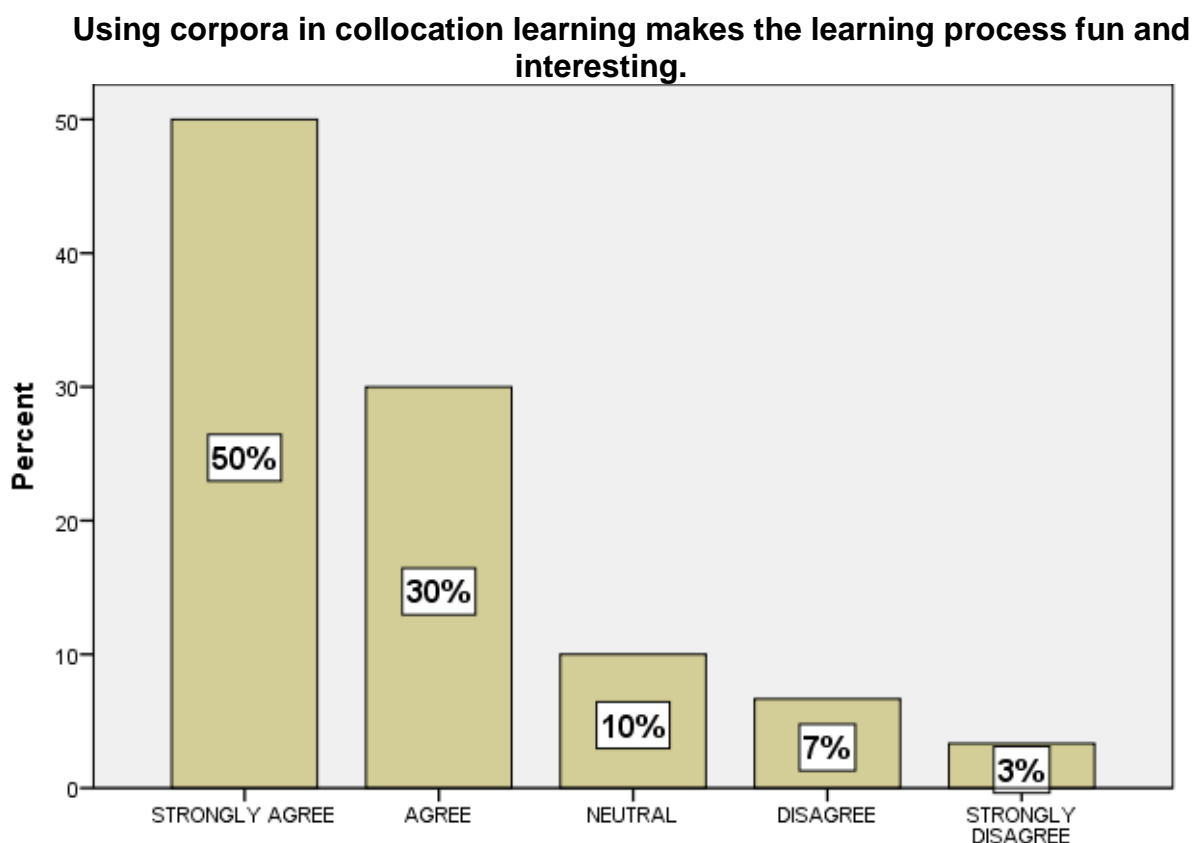
Figure 4.21 above shows the views of participants on corpora and making the learning process fun and interesting. As shown in the figure, *strongly agree* and *agree* scored a relatively high percentage (80%) while *disagree* and *strongly disagree* recorded a very low one (10%). The rest of respondents (10%) remained *neutral*. The vast majority of the subjects agreed that corpora make learning

collocations fun and interesting. This could be attributed to the fact that using corpora makes the learning process learner-centered rather than teacher-centered so that learning barriers could be minimized as peer work can be developed and hence learners relish learning and investigating the language.

Figure 4.22: Using corpora in collocation learning and their ease of accessibility.
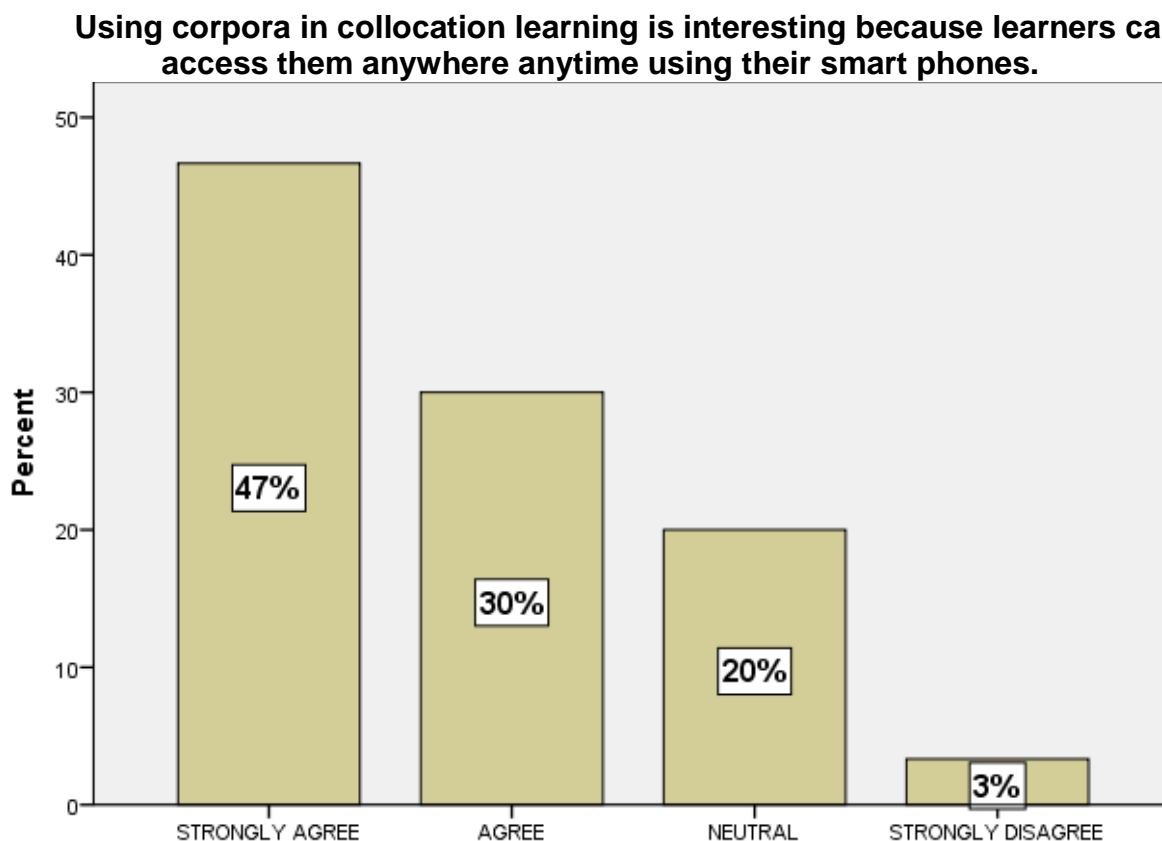
Using corpora in collocation learning is interesting because learners can access them anywhere anytime using their smart phones.

With regard to availability and ease of accessibility, corpora are available for use anywhere anytime and there is a wide range of online corpora that learners can freely access and use. These properties make corpora effective strategy to language learning, specifically collocation learning for learners can exploit them even outside the classroom and see how the language actually works. This claim is proved by figure 4.22 above by the fact that 77% of the participants agreed that using corpora is interesting due to the reality that corpora can be accessed by

125

learners via their smart phones anywhere anytime. It could be inferred that this free of accessibility makes learners free of classroom learning barriers so that they could enjoy learning the language and thus develop their competence of that language.

Figure 4.23: Corpora and developing peer work**.**

Learning the language through corpora develops peer work so that learners can negotiate meanings and uses of words.
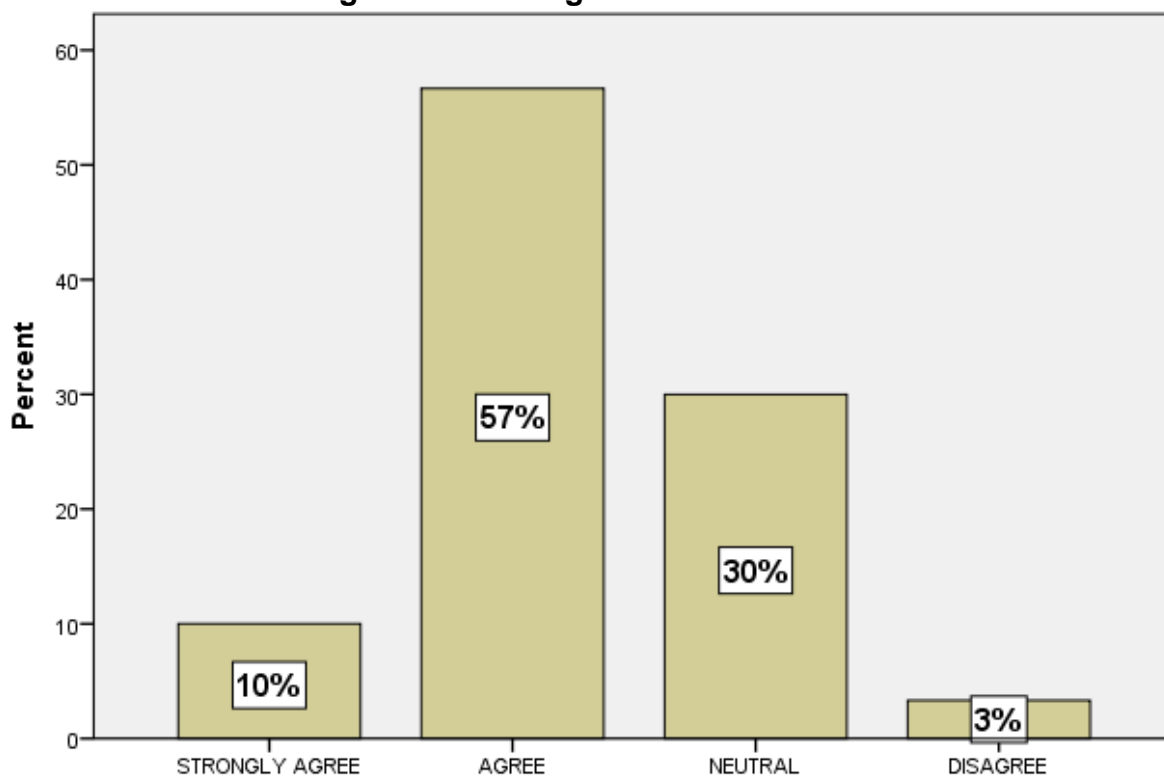
More than any other strategy to language learning, corpora have the ability of developing peer work so that learners can negotiate meanings and uses of words. This claim is substantiated by the fact that as shown in figure 4.23, 67% of the participants agreed while only 3% disagreed. It is worth notifying that the remaining 30% remained neutral. From these percentages, it could be claimed that corpus use makes the learning process learner-centered so that learners can develop their peer work skills and negotiate meanings and uses of words and thus develop

their knowledge of the language. In addition, learners can better learn from each other due to the lack of formality and that there are no barriers among them.

Figure 4.24: Using corpora in learning collocations and developing learners' grammatical knowledge.

Using corpora in learning collocations implicitly improves learners' grammatical knowledge.
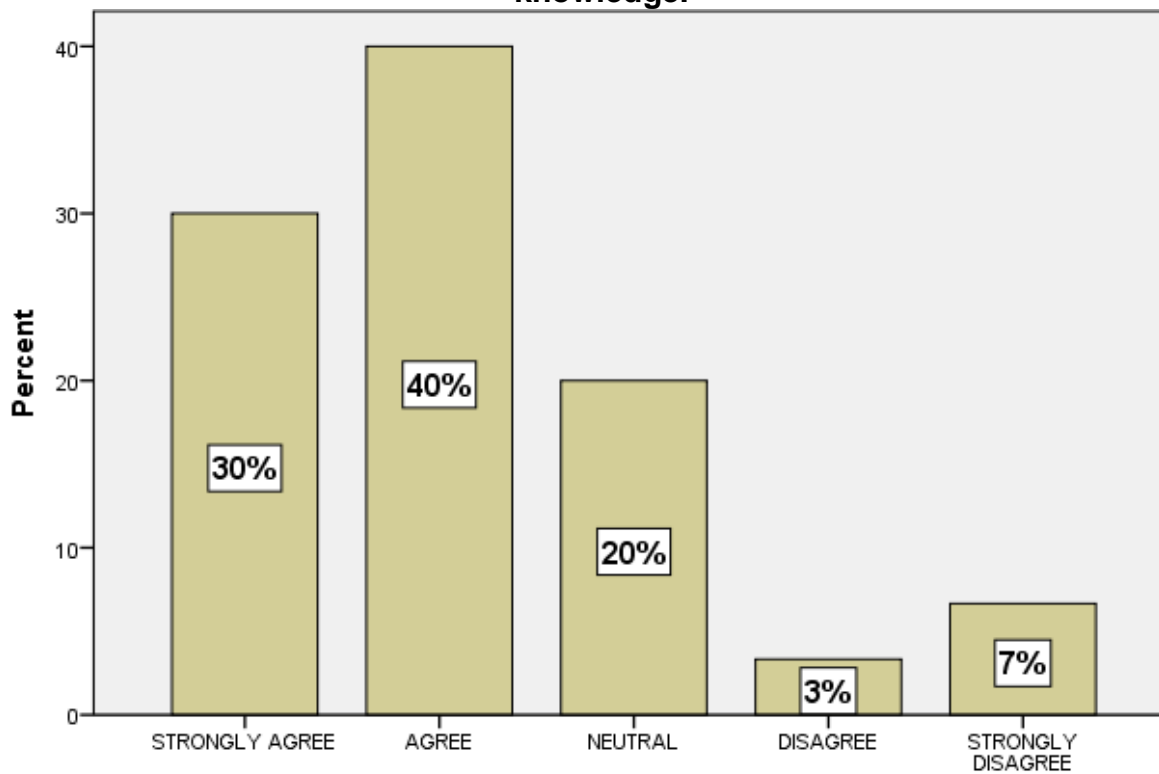
It is of no doubt that learning the language through collocations implicitly improves learners' lexical knowledge as well as their grammatical knowledge of words. This is due to the fact that learning the language in chunks always incorporates lexical and grammatical features of words. Through this strategy, learners' grammatical knowledge can be raised without directly recognizing that. In addition, learners cannot even easily distinguish what is a lexical investigation and what is a grammatical one. One leads to the other. This claim is substantiated by the above figure which shows that the vast majority of the subjects agreed that

using corpora in learning collocations implicitly improves learners' grammatical knowledge while only 10% disagreed. It is to be noted that 20% of the participants responded with *neutral*.

Figure 4.25: Understanding the natural texts in corpora.
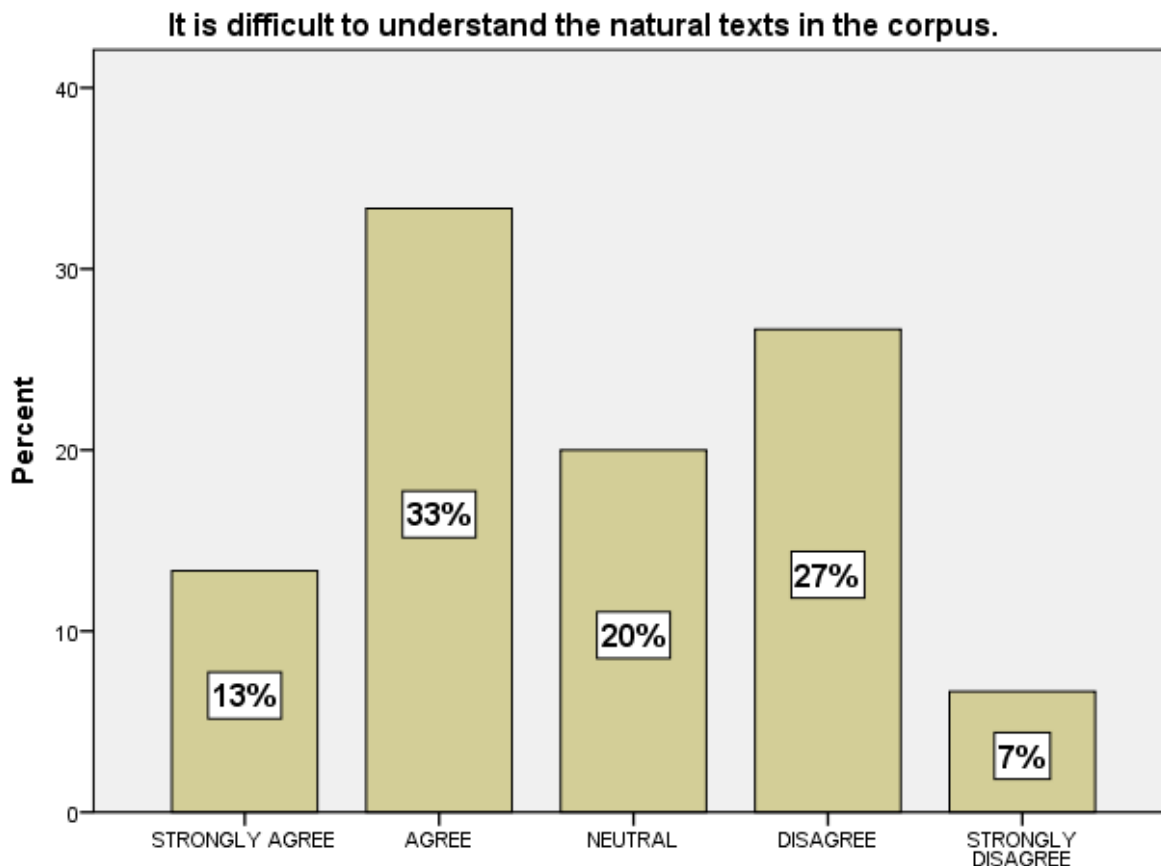
Figure 4.25 above presents the views of participants on the difficulty of understanding the natural texts in the corpus. As shown in the figure, *strongly agree* and *agree* recorded 46%. On the other side, *disagree* and *strongly disagree* received 34%. The rest of respondents reacted with *neutral*. As clearly shown, the participants who agreed to the fact that natural texts in corpora are difficult to understand are more than those who disagreed. It can be deduced that this difficulty resides in the fact that corpus texts are naturally collected and they are

real language used by native speakers. These texts reveal the complexity of English language. Unlike the other artificial language found in textbooks that learners have been accustomed to.

Figure 4.26: Corpora and guessing meanings of unfamiliar words from context.

Using corpora helps in guessing the meanings of unfamiliar words from context.
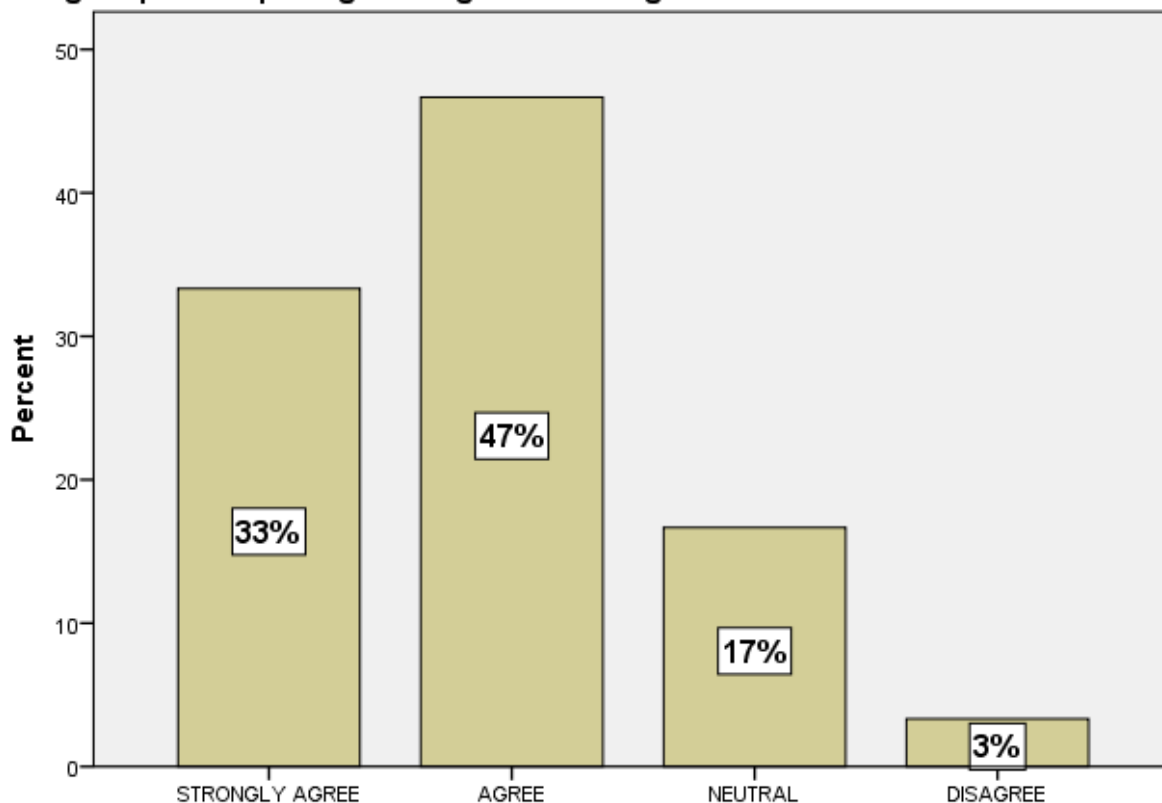
The above figure shows the subjects' views on the use of corpora in guessing meanings of unfamiliar words from context. As demonstrated above, a relatively high percentage (80%) is given to *strongly agree* and *agree* while *disagree* and *strongly disagree* recorded only 3%. To be notified, the other 17% of the subjects responded with *neutral*. As corpora are texts (not lists of individual words), they provide the linguistic contexts of words so that they assist in guessing meanings of unfamiliar words. Unlike dictionaries and thesauruses where words are listed with

their definitions and some intuition-based examples for each word, corpora provide more realistic examples of how native speakers naturally use their language.

Figure 4.27: Corpus-based activities as a new strategy to language learning and teaching.

Corpus-based activities should be adopted as a new strategy to language learning and teaching.
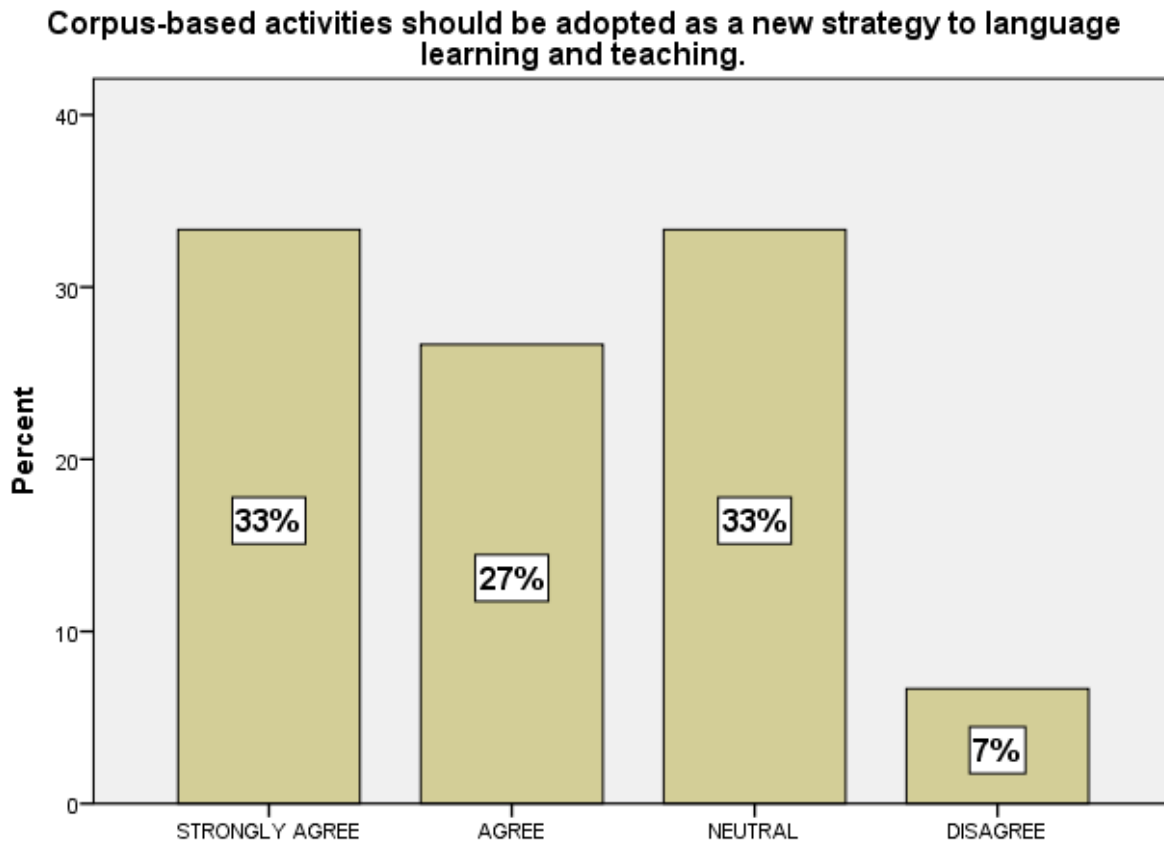
Figure 4.27 above displays responses of respondents on using corpus-based activities as a new strategy to language learning and teaching. As shown in the figure, 60% of the participants agreed while only 7% disagreed. It is worth notifying that 33% of the participants responded with *neutral*. As the percentages proved, the majority of subjects are in favor of employing corpus-based activities in language learning and teaching. Indeed, one of the important roles of corpora is providing more realistic examples of language use that reflects the complexities of natural language. Hence, a successful learning process could be achieved.

130

Figure 4.28: Corpora and the skills of investigating the language.
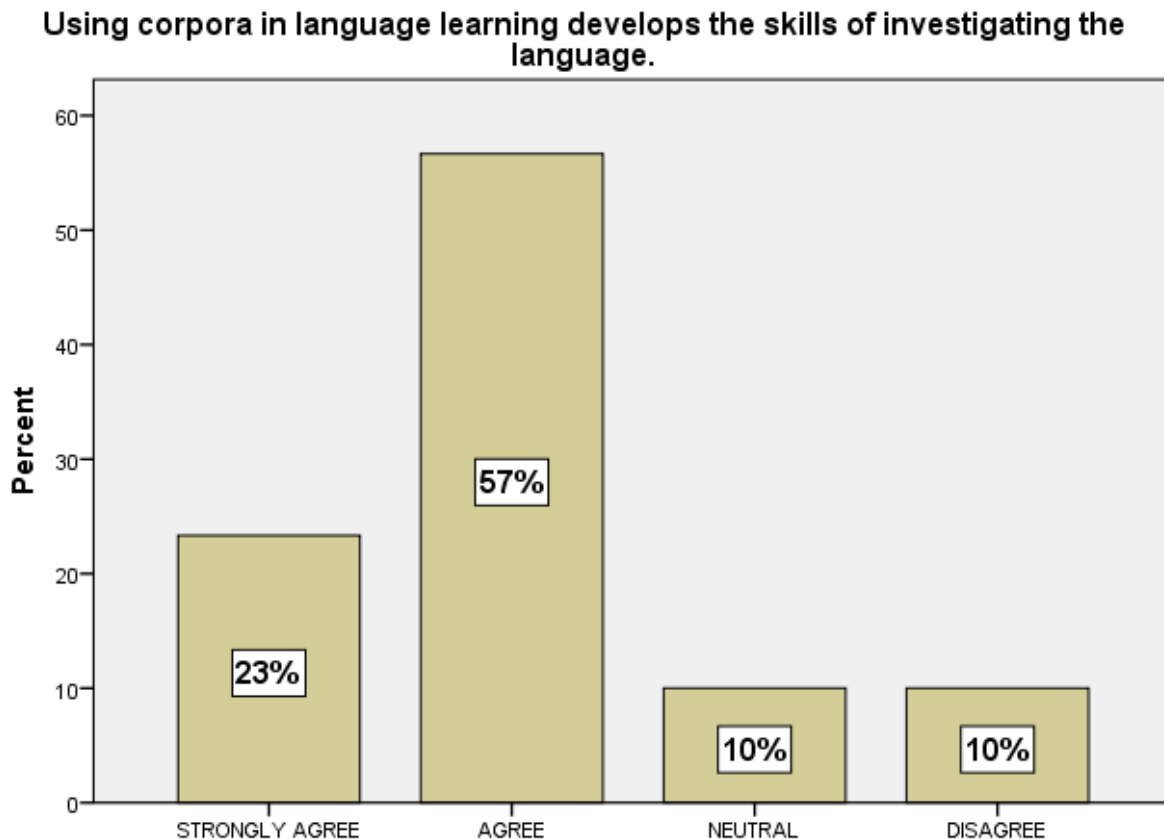
The figure above demonstrates the views of participants on using corpora in language learning and developing their skills of investigating the language. As the figure shows, those who agree scored a slightly high percentage (80%) while those who disagree recorded a very low one (10%). Those who remain neutral scored only 10%. From the percentages provided above, it could be inferred that learners can autonomously investigate the language through corpora and hence become autonomous learners as they can individually access and use them.

Figure 4.29: Corpora and real language use, frequency and collocations.
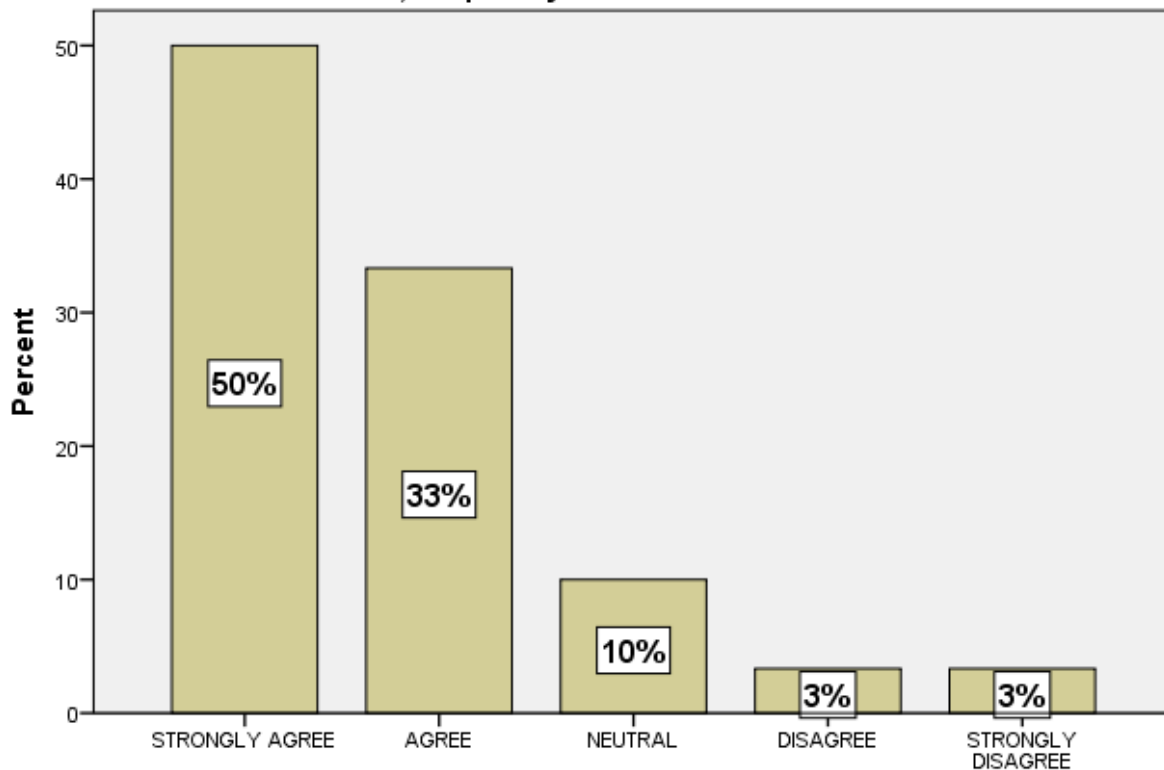
As figure 4.29 above displays, 83% of the participants agreed that corpora have the potentiality of providing learners with easily accessible information about real language use, frequency and collocations. On the other hand, only 6% of them disagreed. It is worth mentioning that 10% of the participants responded with neutral. In fact, there is a wide range of corpora available for use and some of them can freely be accessed. These corpora provide language learners with how language is actually used, frequency of words (what common or uncommon words in a language), and collocations.

Figure 4.30: Corpora, looking at language from a critical angle and increasing learners' knowledge of the language.

Using corpora in language learning helps learners look at language from a critical angle and thus increases their knowledge
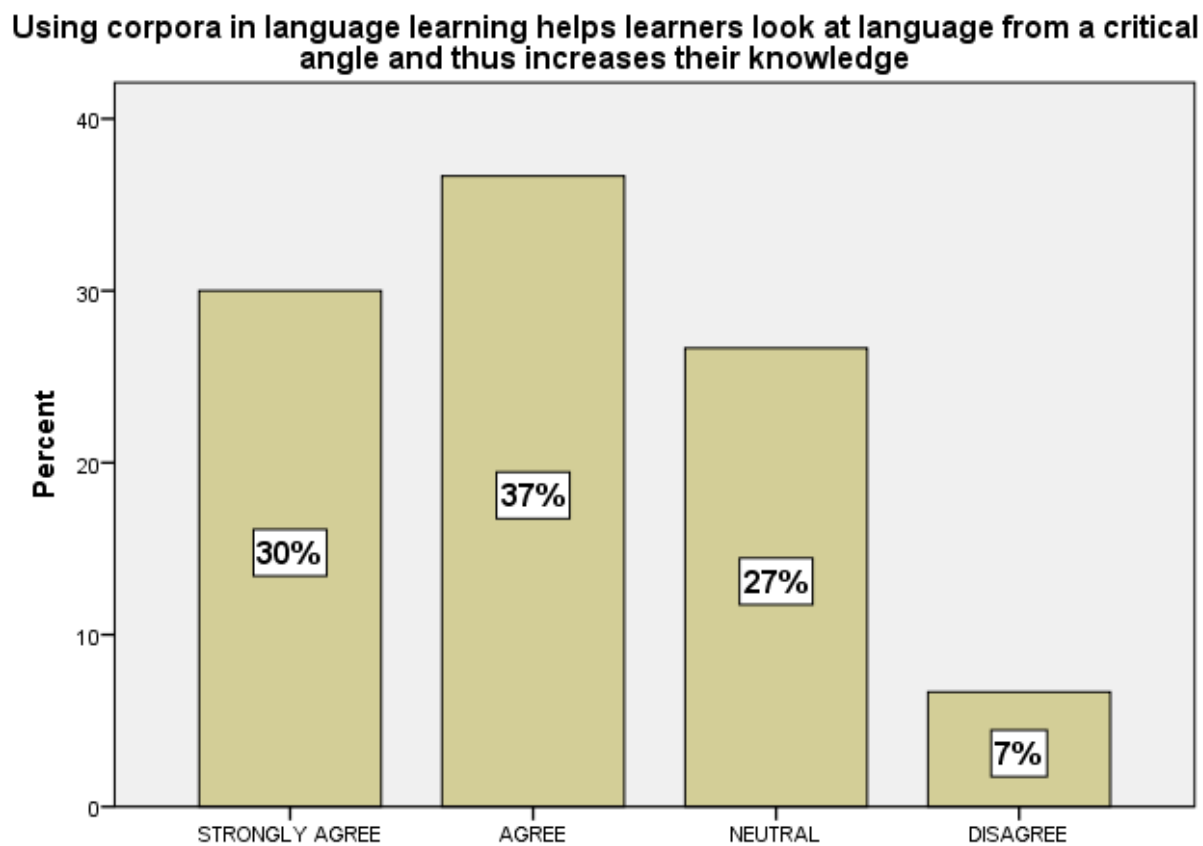
As figure 4.30 displays, 67% of the participants agreed that using corpora in language learning helps learners look at language from a critical angle and thus increases their knowledge of the language while only 7% of them disagreed. The remaining 27% of participants responded with *neutral*. Unlike the other strategies to language learning and teaching, through corpus-based strategy, learners have the chance of looking at how language works and how native speakers exploit the resources of their language to communicate.

4.3 Teachers' Interview Analysis

This part of the chapter explores the views of five EFL university teachers from different higher education institutions on corpus usage in collocation learning by Sudanese undergraduates. These teachers were selected to take part in the study because of their knowledge and experience of corpus linguistics and their availability and wellingness to participate.

The interview aims to explore the following:

- The definition of corpus.

- Corpora and their helpfulness in teaching English language, especially collocations.

- How often corpora are used by the interviewees in English language teaching, collocations in particular.

- Teaching collocations through corpora in comparison to other vocabulary teaching strategies.

- Using corpora for other purposes besides teaching collocations.

Concerning the definition of corpus, nearly all the interviewees defined corpus as a collection of written or spoken texts that meant to represent a language or variety of a language. One of them claimed that corpus is purposefully collected to provide an archive of authentic linguistic data, while another one added that corpus helps in learning authentic language.

Regarding corpora and their usefulness in teaching English collocations, it is believed by most of the interviewees that corpora are of a great help to English language teaching, especially collocations in many ways. To begin with, corpora are considered to be an authentic source of language data for language teachers. In addition, they provide real texts in which words and their collocates are actually used by native speakers as well as they show the different uses and senses of words

in different contexts or situations of language use. Finally, they influence our way of thinking about language and what we use in language teaching.

With regard to how often corpora are used by the interviewees in teaching English collocations, four of the interviewees stated that they do not use corpora in collocation teaching but one confirmed that he uses them to do so from time to time.

Concerning teaching collocations through corpora in comparison to other vocabulary teaching strategies, all the interviewees claimed that corpora are better than the other vocabulary teaching strategies for many reasons. In the beginning, corpora are better in terms of contextualization, because unlike the other strategies, corpora have the potential of providing vocabulary items within context. Secondly, they have authentic data and they are easy to use and modern. Thirdly, for teachers, corpora are considered to be the most important source of information for they help them consult, categorize and then know how to use certain words in different situations or contexts. Fourthly, through the corpus-based approach, learners could take part in the learning process by investigating collocates of words and looking at how words are differently used in different contexts. In conclusion, through using corpora, learners have the chance of becoming autonomous learners, so that they can better learn the language.

Regarding using corpora for other purposes besides teaching collocations, two of the interviewees confirmed that they use corpora to look at how language is actually used by native speakers. On the other hand, the third interviewee stated that he uses corpora in translation and the fourth interviewee uses them in conducting his on-going Ph.D. project due to the facts that they have authentic data and they are easy to use and modern. It is worth notifying that the fifth interviewee confirmed that he does not use corpora in teaching collocations but as far as he knows they are used to expand students' knowledge of the language.

4.4 Verification of the Study Hypotheses

This research aims at investigating the impact of using corpora in solving collocation learning difficulties encountered by Sudanese undergraduates. More specifically, the study tries to validate the following research hypotheses:

- Corpora to a far extent help in solving collocation learning difficulties encountered by university students.

- It is supposed that English language teachers have positive attitudes towards using corpora in solving collocation learning difficulties encountered by university students.

- It is expected that there is a significant difference between those who study through corpora and those who do not.

4.4.1 The First Hypothesis

In order to validate the first research hypothesis (corpora to a far extent help in solving collocation learning difficulties encountered by university students), a student attitude questionnaire was used in the data collection process. It was administered to participants of the experimental group (second year students of Biomedical Engineering at Bayan College of Science and Technology) because they were taught collocations through corpora so that they can judge whether corpora are effective in learning collocations.

Results of the questionnaire proved that corpora help in solving collocation learning difficulties encountered by university students. In almost all of the questionnaire statements devoted to the role of corpora in language learning, strongly agree and agree overweight neutral, disagree, and strongly disagree. For instance, statement 17 which states: 'Corpora are helpful in collocation learning because they develop students' autonomous learning'. In this statement, 83% of the participants agreed while only 6% disagreed and the remaining 10% were neutral.

This clearly indicates that corpora provide authentic contexts for learning collocations because of the fact that their data are naturally collected from different contexts of language use. These collections of language data show learners how the language is actually used by native speakers. A fact that many other strategies failed to prove. Another example is statement 21 which says: 'Using corpora in collocation learning makes the learning process fun and interesting'. Reacting to this statement, strongly agree and agree recorded relatively high scores (50% and 30% respectively) while disagree and strongly disagree scored very low percentages (7% and 3% respectively). These scores could be attributed to the fact that using corpora makes the learning process learner-centered rather than teacher-centered so that learning barriers could be minimized as peer work can be developed and hence learners enjoy learning and investigating the language. In this concern, Choorit et al. (2014) proved that learners can explore the language patterns and search for naturally-occurring samples through the use of corpora. In a similar context, Agnes (2012) on her corpus-based study, stated that the questionnaire results confirmed that the instruction was effective in increasing students' knowledge of the three causatives. This demonstrates that corpora have been proved to be an effective strategy to English language learning and teaching because of their inclusion of large and authentic language data. These data are believed to be effective in investigating words and their grammatical environments which provide a baseline for understanding collocates of words. It is believed by this study that understanding the use of individual words (such as periphrastic causatives, delexical verbs, etc.) cannot be achieved without their context of use which can best be provided by corpora/corpus-based activities. Consistantly, using interview to discover the attitudes of students towards using corpora, Girgin (2011) stated that the analysis of students' interviews showed more positive attitudes towards using corpus-based activities in the learning of five target structures in

English grammar. Additionally, according to the present study, the results of the post-test showed that there is a statistical difference between the control group and the experimental one, which means that the experimental group performed better than the control group. From these results, it could now be inferred that corpora help in contributing a solution to collocation learning difficulties encountered by Sudanese unviersity students.

In the researcher's own view, corpora have the potential of solving collocation learning difficulties for a number of reasons. In the beginning, corpora improve learners' lexical as well as their grammatical knowledge of words. This is due to the fact that learning the language in chunks always incorporates the lexical and grammatical features of words. Through this strategy, learners' lexical and grammatical knowledge can be raised without directly recognizing that. Consequently, learners cannot even easily distinguish what is a lexical investigation and what is a grammatical one because one leads to the other. This claim is substantiated by figure 4.25. In addition, corpora have the potential of guessing the meanings of unfamiliar words from context. They provide the linguistic contexts of words so that they assist in guessing meanings of unfamiliar words. This is confirmed by figure 4.26. Furthermore, the use of corpus-based activities in language learning and teaching provides more realistic examples of language use that reflects the complexities of natural language. Thus, a successful learning process could be achieved. In addition to that, corpora provide learners with easily accessible information about real language, frequency of words and collocations. In fact, there is a wide range of freely available corpora that learners can access and see how the language behaves. This claim is confirmed by figure 4.29. In conclusion, corpora suit the needs of learners (specially learners of English

as a foreign and second language) for mastering their collocational competence and hence achieving high level of proficiency in the language.

From the discussion above, it could be said that the first research hypotheseis that claims 'corpora to a far extent help in solving collocation learning difficulties encountered by university students', is proved.

4.4.2 The Second Hypothesis

For the purpose of verifying the second hypothesis of the study (it is supposed that English language teachers have positive attitudes towards using corpora in solving collocation learning difficulties), a teacher interview was conducted on five experienced EFL teachers from different Sudanese universities.

Results of the interview showed that most of the interviewees agreed that corpora to a far extent help in English language teaching, particularly collocations in many ways. For instance, they are considered to be an authentic source of data for language teachers. The results also revealed that corpora provide real texts in which words and their collocates are actually used by native speakers and they show the different uses and senses of words in different contexts or situations of language use. It is also proved that corpora influence our way of thinking about language and what we use in language teaching.

The interview results also showed that through the corpus-based approach, learners could take part in the learning process by investigating collocates of words and looking at how words are differently used in different contexts. The interviewees added that, via using corpora, learners have the chance of becoming autonomous learners, so that they can better learn the language. In this concern, Biber et al. (1998) stated that a corpus allows us to investigate the meanings of words through the use of concordance listings, which can provide an exhaustive list of all

occurrences of a word in context. In a related context, Girgin (2011) showed that the analysis of the students' interviews revealed more positive attitudes towards using corpus sources in English grammar learning. This proves that not only teachers have positive attitudes towards using corpora in language teaching and learning, but students as well.

As an EFL teacher, the researcher views that the contribution of corpora to language learning and teaching is undeniable as they directly and indirectly assist in EFL/ESL contexts. They can directly be accessed and used by language learners and teachers in the learning and teaching processes, specifically collocations. Such as the use of concordance lines (which display the different uses and senses of words in different contexts of language use) and many other corpus-based techniques/activities. With regard to the indirect use of corpora in language learning and teaching, they assist in decision making concerning what to teach and how to teach it.

According to the above discussion, it could now be stated that the second study hypothesis which claims 'English language teachers have positive attitudes towards using corpora in solving collocation learning difficulties' is verified.

4.4.3 The Third Hypothesis

In order to verify the third study hypothesis (it is expected that there is a significant difference between those who study through corpora and those who do not), two tests (a pre-test and post-test) were employed to collect the data. Participants of the tests were second year students of Biomedical Engineering from Bayan College of Science and Technology. Results of the post-test proved that there is a significant statistical difference between the two groups which means that the experimental group performed significantly better than the control group. This indicates that corpora are effective strategy in learning collocations in English language. Similarly, Ashouri et al. (2014) proved that the learners who were in the

experimental group, got aware of the existence of collocations, used them and learnt them properly and they also started to find the collocation of every other word which they learnt during the term by themselves because the treatment appealed to them. In a similar study conducted by Agnes (2012) on the effectiveness of a corpus-based instruction in deepening EFL learners' knowledge of periphrastic causatives, the results of the study indicated that the experimental group improved and outperformed the control group significantly in the post-test. Li (2017) also confirmed that the results revealed that while both groups (control and experimental) showed improvements in their academic writing, the students in the experimental group showed a significant improvement in the use of collocations, including a higher rate of accuracy, or naturalness, and an increased use of academic collocations and fixed phraseological items.

In the same line, as shown in the previous studies section, Alharbi (2017) on her research 'acquisition of lexical collocations: a corpus-assisted contrastive analysis and translation approach', found that the corpus-assisted CAT (contrastive analysis and translation) group significantly outperformed the other two groups on all the tests. Also in consistence with the present study, Barabadi et al. (2016), investigating the effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary, proved that learners in the experimental group outperformed those of the control group. They added that, the better performance of the experimental group can be attributed to the fact that learners could take a more active role in the process of learning in which self-discovery, inductive and bottom-up processes were emphasized.

The results displayed above are in line with the current study in terms of the claim that 'corpus-based strategy to language learning and teaching is better than the other non-corpus-based strategies. As the researcher sees it, this is true for many reasons. To begin with, comparing corpora to dictionaries, corpora consist of

large language data collected from a wide range of language sources used in language learning, teaching and investigation. On the other side, dictionaries are not as wide as corpora. Besides, the data in dictionaries are based on the intuitions of their writers which are not as effective as corpus data in language learning. In addition, corpora provide authentic context for learning the language, especially collocations because of the fact that their data are naturally collected from different contexts of language use. These collections of language data show learners how the language is actually used by native speakers, a fact that many other strategies failed to prove. Furthermore, corpora can develop learners' autonomous learning. This kind of learning can be developed through corpora because learners can individually access them anywhere anytime and use them in language learning and investigation. Another important factor about corpora is that, through them word senses and uses can be investigated because their data are collected from different language registers meant to show how language is actually used by native speakers in different contexts of use. This fact turns to be an area of strength of corpora because they provide language users with a vital knowledge of how words are differently used in different contexts and consequently develop their knowledge of the language, particularly collocations.

With regard to availability and ease of accessibility, corpora are available for use anywhere anytime and there is a wide range of online corpora that learners can freely access and use them. These properties make corpora an effective strategy to language learning, specifically collocation learning because learners can exploit them even outside the classroom using their smart phones and see how the language actually works. This free of accessibility makes learners free of classroom learning barriers so that they could enjoy learning the language and thus develop their proficiency of that language. Moreover, corpora have the ability of developing peer work so that learners can negotiate meanings and uses of words

and hence, makes the learning process learner-centered so that learners can better learn from each other due to the lack of formality and that there is no barriers among them. Last but not least, corpus-based strategy is better than the other strategies due to the above mentioned factors and many more.

From the results reached by the studies projected above and the mentioned facts about corpora and how they help in language learning and teaching, we can confess that the hypothesis ' it is expected that there is a significant difference between those who study through corpora and those who do not' is verified.

4.5 Summary of the Chapter

This chapter has presented results of the study, discussion, and findings. Then, in the light of the research findings, the study hypotheses have been verified. The following chapter will be dedicated to summary of the study, conclusions, recommendations, and suggestions for further studies.

Chapter Five

Summary, Conclusions, Recommendations and Suggestions for Further Studies

Chapter Five

Summary, Conclusions, Recommendations and Suggestions for Further Studies

5.0 Overview

This chapter presents summary of the study, conclusions, recommendations, and suggestions for further studies.

5.1 Summary of the Study

This study mainly aims to investigate the impact of employing corpora to solve collocation learning difficulties encountered by university students. More specifically, the research attempts to answer the following questions:

- To what extent do corpora help in solving collocation learning difficulties encountered by university students?
- What are the attitudes of English language teachers towards the utilization of corpora in solving collocation learning difficulties encountered by university students?
- To what extent is there a significant difference between those who study through corpora and those who do not?

The study adopts the mixed method approach and four tools of data collection were employed; a pre-test, post-test, student attitude questionnaire, and teacher interview. The population for the study were Sudanese EFL learners and teachers. To begin, participants of the pre-test and post-test were (60 and 30 respectively) students from second year, Biomedical Engineering at Bayan College of Science and Technology. Those students study English as a university requirement.

Regarding respondents of the student attitude questionnaire, they were 30 students who had an experiment using corpora in collocation learning, so that they could judge whether the corpus is effective in raising students' awareness of collocations. With regard to participants of the teacher interview, five Sudanese EFL teachers from different Sudanese universities and university colleges were interviewed. The study findings proved that using corpora in language learning is effective. The findings also showed that English language teachers have positive attitudes towards utilizing corpora in English language learning and teaching. Based on these findings, the study recommended that corpus-based strategies should be adopted as corpora represent real language that is used by native speakers so that students collocation use sounds more natural and native-like. It is also recommended that EFL teachers should be trained on how to exploit corpora and use them in language teaching and learning. In the light of the research findings, it is suggested that more research needs to be undertaken on the difficulties encountered by Sudanese undergraduates in using English collocations, especially those which do not have equivalents in Arabic. Furthermore, as it is hard to find EFL teacher using corpora in the English language classrooms in Sudan, future research should therefore concentrate on investigating the underlying factors.

5.2 Conclusions

Many findings have been reached by this study. They are as follows:

- Using corpora in language learning, especially collocations is quite effective. This is attributed to the authentic data that corpora provide and many other factors.

- Using corpora is the best strategy that suits the needs of students for mastering collocations (particularly delexical verbs) for they enable the investigation of word use within context.

- Using corpora in language learning and teaching is much better than using the other non-corpus-based strategies. This is due to a number of factors such as the fact that through using corpora learners could take a more active role in the process of learning.

- Corpora develop learners' autonomous learning. This kind of learning can best be developed via using corpora because learners can individually access them anywhere anytime and use them in language learning and investigation.

- English language teachers have positive attitudes towards utilizing corpora in language learning, collocations in particular, but nearly all of those teachers do not practically use corpora in their classrooms.

5.3 Recommendations

Based on the study findings, the following recommendations are provided by the researcher:

- EFL teachers should be trained on how to practically use corpora in their classrooms.

- EFL teachers should adopt corpus-based strategies/activities that represent the real language used by native speakers, so that students' collocation use sounds more natural and native-like.

- EFL learners should properly be trained on how to exploit corpora and use them in language learning and investigation.

- Curriculum designers should rely on corpora rather than their own intuitions about the language for corpora provide them with real and authentic contexts of language use. This could assist in fostering students' collocational awareness and thus improves their language proficiency.

- Using corpora in language learning and teaching is recommended but to make best use of them, the availability of language laboratories and the technological equipment is a must.

5.4 Suggestions for Further Studies

Taking into account the study instruments of data collection and the findings obtained, the following suggestions for further studies are provided by the researcher:

- As this study investigated corpus usage in collocation learning, specifically, verb-noun collocations, it is suggested that further research investigating the other types of collocations be undertaken.

- More studies need to be carried out on the significance of collocations to achieving high level of proficiency in the language.

- It is suggested that further experimental investigations on exploiting corpora in language learning, teaching and investigation are needed as corpora have a lot to offer.

- Further research needs to be undertaken on Sudanese undergraduates' difficulties in using English collocations, especially those which do not have equivalents in students' native language (Arabic).

- As it is hard to find EFL teacher using corpora in the English language classrooms in Sudan, future research should therefore concentrate on investigating the underlying factors.

5.5 Summary of the Chapter

This is the concluding chapter of the thesis. It has presented summary of the whole dissertation and conclusions, it has come up with recommendations based on the study findings, and finally it has presented suggestions for future studies.

References

Agnes, L. (2012) *The effectiveness of a corpus-based instruction in deepening EFL learners' knowledge of periphrastic causatives*. Retrieved from https://tesol-international-journal.com/wp-content/uploads/2013/11/A6_V6.pdf [Accessed on 02/04/2018].

Akbar, A., Hashemian, M., & Alipour, S. (2013). *A* corpus-based approach toward teaching collocation of synonyms. *Theory and Practice in Language Studies*. 1(1799-2591), 51- 60 http://dx.doi.org/ 10.4304/tpls.3.1.51-60 Retrieved from

http://www.academypublication.com/issues/past/tpls/vol03/01/08.pdf

Alharbi, R. M. (2017). *Acquisition of Lexical Collocations: A corpus-assisted contrastive analysis and translation approach*. (Doctoral dissertation).

Newcastle University. Retrieved from https://theses.ncl.ac.uk/dspace/bitstream/10443/3660/1/Alharbi%2C%20R.M.S.%202017.pdf [Accessed on 07/09/2017].

Ali, A. (2017). *Investigating the Sudanese EFL undergraduates knowledge and use of collocation* (Doctoral dissertation).

Retrieved from http://repository.sustech.edu/bitstream/handle/123456789/18590/Investigating%20the%20Sudanese%20EFL....pdf?sequence=1 [Accessed on 11/07/2017].

Alzi'abi, S. (2017). *Guessing verb–adverb collocations: Arab EFL learners' use of electronic dictionaries*. http://lexikos.journals.ac.za Retrieved from file:///C:/Users/wn7/AppData/Local/Temp/1394-2237-1-PB.pdf [Accessed on 02/04/2018].

Ashouri, S., Arjmandi, M., & Rahimi, R. (2014). The impact of corpus-based collocation instruction on Iranian EFL learners' collocation learning. *Universal Journal of Educational Research* 2(6), 470-479 http://dx.doi.org/10.13189 Retrieved from https://files.eric.ed.gov/fulltext/EJ1053887.pdf

Bahns, J. (1993). *Lexical collocations: A Contrastive View*. ELT Journal, 47(1), 56 – 63.

Barabadi, B., & Khajavi, Y. (2016). The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. *Cogent Education*. 4 (1283876). http://dx.doi.org/10.1080/2331186X.2017.1283876. Retrieved from https://www.cogentoa.com/article/10.1080/2331186X.2017.1283876.pdf

Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics*, 1(1), 1-37. Available at https://benjamins.com/catalog/ijcl.1.1.03bar

Benson, M., Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of*

English: A guide to word combinations. Amsterdam: John Benjamins.

Bernardini, S. (2002), *Exploring New Directions for Discovery Learning*.
Retrieved   from
http://booksandjournals.brillonline.com/content/books/b9789004334236s015
[Accessed on 10/10/2017]

 Bibber, D., Conard, S., & Reppen, R. (1998). *Corpus linguistics: Investigating*

language structure and use. Cambridge: Cambridge University Press.

Breyer, Y. (2011). *Corpora in language teaching and learning: Potential,*
evaluation, challenges. Retrieved from
https://www.researchgate.net/publication/264499058 [Accessed on
27/04/2018].

Bussmann, S. (2006). *Routledge dictionary of language and linguistics*. New York:
Routledge.

Carmen, M., Belles-Fortuno, B. & Gea-Valor, M. (2010). *Corpus-based*

approaches to English language teaching. London: Continuum

International Publishing Group.

Choorit, K. & Supakorn, P. (2014). A corpus-based study of phrasal verbs: carry

out, find out, and point out. *International Journal of Research Studies in*

Language Learning. 3 (7), 73 – 88 Retrieved from

http://consortiacademia.org/wp-
content/uploads/IJRSLL/IJRSLL_v3i7/820-3121-1-PB.pdf

Cohen, L., Manion, L., & Morrison, K. (2005). *Research methods in education*.

London & New York: ROUTLRDGE FALMER & Taylor & Francis

Group.

Collocation (2014). *Oxford concise dictionary of linguistics* (Third edition).

Oxford: Oxford University Press.

Collocation (2007). Oxford Advanced Learner's Dictionary (Seventh edition).

Oxford: Oxford University Press.

Collocation. In Oxford Online English Dictionary. Retrieved from

https://www.thefreedictionary.com/Oxford+English+Dictionary+Online

[Accessed on 11/02/2018].

Conard, S. (2010). What can a corpus tell us about grammar?. In A. O'Keeffe & M.

McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 227-

240). New York, NY: Taylor & Francis.

Corpus (2014). *Oxford concise dictionary of linguistics* (Third edition). Oxford,

Oxford University Press.

Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed

methods approach*. Singapore. SAGE Publication Asian-Pacific Pte.

Ltd.

Creswell, J. W. (2012). *Educational research: Planning, conducting and

evaluating quantitative and qualitative research*. Boston: Pearson

Education, Inc.

El-Dakhs, S. (2015). Collocational competence in English language teaching. *Arab

world English journal, 6 (1), 68-82. Available at

https://www.researchgate.net/publication/282364488

Evison, J. (2010). What are the basics of analysing a corpus?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 122-135). New York, NY: Taylor & Francis.

Firth, J. (1957). *Papers in linguistics*: 1934-1951. London: Oxford University Press.

Girgin, U. (2011). *Corpus-based activities at lower levels of EFL proficiency: The effectiveness of using concordance lines on grammar learning*. (Master thesis). Bilkent University. Available at

http://www.thesis.bilkent.edu.tr/0005040.pdf [Accessed on 12/06/2017].

Greaves, C. & Warren, M. (2010). What can a corpus tell us about multi-word units?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 212-226). New York, NY: Taylor & Francis.

Hong, Y. (2014). *A study of collocation errors among Chinese learners of English (with reference to Chinese college students of Tongji University in China).* (Doctoral dissertation). University of Malaya.

Available at http://studentsrepo.um.edu.my/4698/1/YE_HONG.pdf

[Accessed on 12/06/2017].

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. New York: Addison Wesley Longman Limited.

Kashiha, H. & Heng, C. S. (2014). Discourse functions of formulaic sequences in

        academic speech across two disciplines. *GEMA Online Journal of*

        Language Studies. 14 (2), 15-27 Retrieved

        from http://ejournal.ukm.my/gema/article/view/4555

Kothari, C. R. (2004). *Research methodology: Methods and techniques*. India:

        New Age International Publishers.

Lee, D. Y. (2010). What corpora are available?. In A. O'Keeffe & M.

Lewis, M. (1993). *The lexical approach*. Hove: LTB Publications.

Lewis, M. (1997). *Implementing the lexical approach: Putting theory into practice*.

        Hove: Language Teaching Publications.

Li, S. (2017). Using corpora to develop learners' collocational competence.

        Language learning and technology. 21 (3), 153-171. Retrieved from

        https://scholarspace.manoa.hawaii.edu/bitstream/10125/44625/1/21_03_li

        Pdf

Manafi, S. & Ghaffarof, S. (2013). The effect of collocational competence on

        translation accuracy of translation trainees. *Journal of Advances in English*

        Language Teaching. 1 (3), 76 – 84

        Retrieved   from   file:///C:/Users/wn7/AppData/Local/Temp/277-843-1-
PB.pdf

Marczyk, G. DeMatteo, D. & Festinger, D. (2005). *Essentials of research design*

and methodology. Canada: John Wiley & Sons, Inc. McCarthy (Eds.), *The*

Routledge handbook of corpus linguistics.

Mcmillan, J. H. (1996). *Educational research: Fundamentals for the consumer*.

The United States of America: HarperCollins CollegePublishers.

Menon, S. & Mukundan, J. (2012). Collocations of high frequency noun keywords

in prescribed science textbooks. *International Education Studies*. 5, (6)

http://dx.doi.org/10.5539/ies.v5n6p149 Retrieved from

https://files.eric.ed.gov/fulltext/EJ1067079.pdf (pp. 107-121). New York,

NY: Taylor & Francis.

Meyer, C, F. (2004). *English corpus linguistics: An introduction*. Cambridge:

Cambridge University Press.

Mohammad, S. & Akbarian, I. (2007). Language proficiency and collocational

competence. *The journal of Asian TEFL*, 4 (4), 35-58.   Available at

file:///C:/Users/wn7/AppData/Local/Temp/AsiaTEFL_V4_N4_Winter_2007

Nasr, O. (2011). *Assessing English collocational knowledge among Sudanese EFL*

university students. (PhD dissertation). Sudan University of Science and

Technology.

Nation, I. S. P. (2001) *Learning vocabulary in another language*. Cambridge:

Cambridge University press.

Neville, C. (2007). *Introduction to research and research methods*. Available at

https://www.unrwa.org/sites/default/files/introduction-to-research-and-research-methods.pdf. [Accessed on 27/05/2018].

Okeeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. New York: Cambridge University Press.

Oxford Concise Dictionary of Linguistics. (2004). Sixth ed. Oxford University Press.

Pandey, P. & Pandey, M. M. (2015). *Research methodology: Tools and techniques*. Romania: Bridge Center.

Rabie, E. (2017) *An investigation into the problems of understanding and using idiomatic expressions and collocations among university EFL students majoring in English language*.[Doctoral dissertation]. Sudan University of Science and Technology. Retrieved from http://repository.sustech.edu/handle/123456789/18656 Accessed on 05/05/2017].

Reppen, R. (2010). Building a corpus: what are the key considerations?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 31-37). New York, NY: Taylor & Francis.

Salih, A. F. (2016). *A corpus-based study of academic collocation use and patterns*

in postgraduate computer science students writing. (Doctoral

dissertation). University of Essex. Available at

http://repository.essex.ac.uk/16426/1/FINAL%20AFNAN%20PHD%20THESIS-%20submit.pdf [Accessed on 01/08/2017].

Sinclair, J. (2004).*How to use corpora in language teaching*. Amsterdam: John
    Benjamins.

Singh, Y. K. (2006). *Fundamental of research methodology and statistics*. India:

    New Age International Publishers.

Wu, S., Li, L., Witten, I. & Yu, A. (2016). Constructing a collocation learning
    system from the Wikipedia corpus. *International Journal of Computer-
    Assisted Language Learning and Teaching*. 6 (3). Retrieved from
https://pdfs.semanticscholar.org/ffd3/815a1b36fbc5a01749f58b2062947f3a1c09.pdf

Appendices

Appendix 1: Student's Pre-test

Part One: Participant's Information

Name (optional): ....

Gender:

   a) Male [   ]   **b)** Female [   ]

Part Two: The Test Questions

Please use the appropriate delexical verb (<u>make</u>, <u>do</u>, <u>have</u> or <u>take</u>) and change the form of the verb if necessary to fill in the following:

Section One: <u>Make</u> and <u>Do</u>

1 - Will you ……………. me a favor if you don't mind? I want you to carry these bags for me.

2 - It is your turn to .... the cooking.

3 - I can't stay with you till tomorrow because I have a lot of things to ....

4 - You should always check your work carefully in case you have .... some mistakes.

5 - What do you .... for a living?

6 - Please be responsible and ……………..………… something positive.

7 - The rain has .... a lot of damage to the city.

158

8 - Sara certainly .... an impression last night! All my friends are asking about her.

9 - I have to quit. I want to …………….. my own business.

10 - I am just passing by to …………………. sure that everybody is doing very well in the exam.

11 - They will be …………….……. priority decisions about key issues.

12 - Hamburger is my favorite. My wife ……………… it to me every day.

13 - They …………… me an interesting offer with a good salary in Saudi Arabia last year.

14 - She …………….. everything to win his heart but she couldn't.

15 - I have to ………………… a reservation early in the morning for my wife.

Section Two: <u>Have</u> and <u>Take</u>

1 - Why don't you .... a nice cool drink before you leave.

2 - The governor is expected to .... a decision concerning foreigners in the coming few days.

3 - Suzan .... an argument with her husband yesterday morning.

4 - I want to .... a picture with the tourists with my camera.

5 - We were .... a long conversation before we come to an agreement.

6 - I will .... a taxi so that I can get there fresh and cool for the interview.

7 - I have been studying English for ten years now. I feel like I have to .... a break.

8 - The weather is amazing. Would you like to .... a walk?

9 - You are old enough to .... responsibility of your actions.

10 - They .... a serious fight because of the new governmental system.

11 - I will be .... IELTS test in the coming few weeks.

12 - It is always useful to .... a nap especially when you are having a long working day.

13 - They are .... a wedding party of their colleague on the 17th of June in Bahri.

14 - The government has to .... immediate action concerning the robbery crime.

15 - We are .... a barbecue by the bend of the Blue Nile next week.

Thank you very much indeed for your cooperation

The researcher,

Ibrahim Alsafi Abduldafi (Ph.D. Candidate)

Sudan University of Science and Technology

Appendix 2: Student's Post-test

Part One: Participant's Information

Name (optional): ....

Gender:

   a) Male [   ]   **b)** Female [   ]

Part Two: The Test Questions

Please use the appropriate delexical verb (<u>make</u>, <u>do</u>, <u>have</u> or <u>take</u>) and change the form of the verb if necessary to fill in the following:

Section One: <u>Make</u> and <u>Do</u>

1 - Can you wait a minute? I have to ………….. three phone calls right now.

2 - This week, you ……………… the cooking and I will do the cleaning and the washing up.

3 - They are going to marry without ………………….. any arrangements.

4 - All I can do is to ……………….. things clearer and easier than before.

5 - Going to college has really .... a difference to how I think.

6 - I ………………... coffee every morning. I can't start my day without it.

7 - If you have a lot of things to do, you should …………………….. a list.

8 - Don't worry about getting everything correct. Just .... your best.

9 - Sudanese …………………….. delicious traditional food.

10 - In the last English exam, I …………… only two mistakes in the whole exam.

11 - Which job are you going to take? You need to .... a choice.

12 - He's fed up with .... everything himself. He wants to get married soon.

13 - I've been studying English for 4 years now, but I don't feel like I'm .... progress.

14 - I'm .... a research on the difficulties of learning English language.

15 - She is outgoing and loves .... a lot of friends.

Section Two: <u>Have</u> and <u>Take</u>

1 - I ………………...... advantage of the fact that my friend works in a clothes boutique.

2 - I'm thirsty. I'm going to .... a drink of water.

3 - I .... a good breakfast before I left home.

4 - We use to .... a good time every day by the shore of the river.

5 - Don't …………………… any notice of what he says. It's all nonsense.

6 - I .... a nice relaxing holiday in Mecca.

7 - I …………………… my doctor's advice and stopped drinking alcohol.

8 - It is always a good idea to .... your time when making an important decision.

9 - It is really helpful to ……….…………........... notes during classes.

10 - I want to .... a swim in the lake before it gets hotter.

11 - We .... a nice chat over coffee.

12 - They usually .... a bus to college when it is early.

13 - You need to …………………. care of yourself so that you won't get sick again.

14 - I'm planning to …………………….. an English course this summer.

15 - Mountaineers ……………………... many risks when they climb high mountains.

Thank you very much indeed for your cooperation

The researcher,

Ibrahim Alsafi Abduldafi (Ph.D. Candidate)

Sudan University of Science and Technology

Appendix 3: Student Attitude Questionnaire

Dear Student

The researcher wishes to seek your opinions towards the difficulties that face students in learning English collocations, students' motivation towards learning these word combinations, and the use of corpora in learning and teaching collocations in English language.

This questionnaire is part of a study entitled "*Impact of Employing Corpora in Solving Collocation Learning Difficulties Encountered by University Students*" to be submitted to Sudan University of Science and Technology for the fulfillment of Ph.D. degree in English language.

To assure you, the information in this questionnaire will confidentially be used for research purposes only. Your cooperation is highly appreciated.

Please carefully respond to the statements that follow.

Thanks for being cooperative.

The researcher,

Ibrahim Alsafi Abduldafi (Ph.D. Candidate)

Sudan University of Science and Technology

Part One

Personal Information of Participant:

Name: ....

Gender:

   a) Male [   ]   **b)** Female [   ]

Part Two

The Questionnaire Statements

Please put a tick [√] in the right box that represents your attitude towards each of the following statements: (The following statements investigate the learning of English collocations via corpora, the difficulties that face students in learning this word combination, and students' motivation toward learning these collocations).

Statement	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1 - Learning the language through collocations easily develop students' receptive knowledge.					
2 - Learning the					

language via collocations effectively develop students' productive knowledge.				
3 - The inappropriate use of collocations is attributed to the influence of EFL learners' native language.				
4 - Collocations are difficult for undergraduates to properly use because they are not properly taught by teachers at schools.				
5 - Collocations are difficult for undergraduates to properly use because they are not properly taught by teachers at				

university.					
6 - Collocations are confusing for students because they are not given enough attention in the classroom.					
7 - Understanding collocations properly requires solid cultural and linguistic background of learners' L1 (Arabic) and English language.					
8 - To properly use collocations, learners must have cultural and linguistic background of both their L1 (Arabic) and English language.					
9 - Learners' lack of					

practice is one of the factors affecting learners' collocational competence.				
10 - Some undergraduates are fully aware of the significance of collocations, so they properly learn them.				
11 - Some undergraduates are fully aware of the significance of collocations, so they properly use them.				
12 - Undergraduates are not fully exposed to collocations at the primary and secondary levels.				
13 - The learning strategies used for learning				

collocations at university are not fully effective for developing students' knowledge of collocations.				
14 - Collocations are not given enough attention in the university curriculum.				
15 - Learning collocations through corpora is much better than via the other learning strategies.				
16 - The corpus is more helpful than the dictionary for vocabulary learning.				
17 - Corpora are effective in learning collocations due to the authentic context they				

provide.					
18 - Corpora are helpful in collocation learning because they develop students' autonomous learning.					
19 - Learning collocations through corpora is beneficial as corpora enable the study of the semantic and syntactic properties of words.					
20 - Learning collocations via corpora enables investigating the uses and senses of words in different registers.					
21 - Using corpora in collocation learning makes the					

learning process fun and interesting.				
22 - Using corpora in collocation learning is interesting because learners can access corpora anywhere anytime using their smart phones.				
23 - Learning the language through corpora develops peer work so that learners can negotiate meanings and uses of words.				
24 - Using corpora in learning collocations implicitly improve learners' grammatical knowledge.				
25 - It is difficult to understand the				

natural texts in the corpus.					
26 - Using corpora helps in guessing the meanings of unfamiliar words from context.					
27 - Corpus-based activities should be adopted as a new strategy to language learning and teaching.					
28 - Using corpora in language learning develops the skills of investigating the language.					
29 - Corpora provide learners with easily accessible information about real language use, frequency and collocations.					

30 - Using corpora in language learning helps learners look at language from a critical angle and thus increases their knowledge of that language.					

Appendix 4: Teacher's Interview

Dear Participant

This interview is part of a study entitled "*Impact of Employing Corpora in Solving Collocation Learning Difficulties Encountered by University Students*" to be submitted to Sudan University of Science and Technology for the fulfillment of Ph.D. degree in English language.

To assure you, the information in this interview will confidentially be used for research purposes only. Your cooperation is highly appreciated.

Thank you very much indeed for your cooperation

The researcher,

Ibrahim Alsafi Abduldafi (Ph.D. Candidate)

Sudan University of Science and Technology

Part One

Participant's Personal Information:

1 - Name of Participant

....

2 - Name of University

....

Please Tick the Following:

3 - **Academic Qualifications in English Language:**

   a) B.A / B.Ed. **[   ]**   **b)** M.A. **[   ]**   **c)** Ph.D. **[   ]**

4 - **Experience in Teaching English:**

   a) 1-5 (years)   [   ]   **b)** 6-10 (years) [   ]

   c) 11-15 (years) [   ]   **d)** more than 15 (years) [   ]

5 - **Professional Qualifications:**

   a) ELT Diploma   **[   ]**

   b) Diploma in TEFL   **[   ]**

   c) CELTA (Certificate in English Language Teaching to Adults) **[   ]**

   d) DELTA (Diploma in English Language Teaching to Adults) **[   ]**

   e) Other   **[   ]**

   f) None   **[   ]**

Part Two

The Interview Questions

1 - Define what is corpus?

....

....

....

....

....

....

....

....

....

2 - Can corpora be helpful in teaching English language, especially collocations? If so, in what ways?

....

....

....

....

....

....

....

....

....

....

3 - How often do you use corpora in teaching English language, collocations in particular?

....

....

....

....

....

....

....

....

....

....

4 - What do you think of teaching collocations through corpora in comparison to other vocabulary learning strategies?

....

....

....

....

....

....

....

....

....

....

5 - Do you use corpora for any other purposes, besides teaching collocations? Why/why not?

....

....

....

....

....

....

....

....

....

....