Sudan University of Science and Technology
College of Graduate Studies
College of Languages

EFFECTIVENESS OF USING CORPUS LINGUISTICS ON VOCABULARY LEARNING AND RETENTION IN EFL CLASSROOM

فاعلية استخدام لغويات المتون في تعلم واستبقاء المفردات في فصول تعلم اللغة الإنجليزية لغةً أجنبيةً

(Thesis Submitted to the College of Graduate Studies for the Degree of PhD in Applied Linguistics)

Submitted By
Ahmed Fathy Faragallah Youssef

Supervised By
Dr. Mahmoud Ali Ahmed Omar

2020, October

Sudan University of Science & Technology

College of Graduate Studies

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

كلية الدراسات العليا

Ref. SUST/ CGS/AH

Approval Page

(To be completed after the college council approval)

Name of Candidate: Ahmed | Fathy | Fangello | Yousef

Thesis title: The Effectiveness of Using Corpus Linguistics on Vocabulary Learning and Retention in EFL Classrooms

Degree Examined for: PhD

Approved by:

1. External Examiner

Name: Dr. Kenya Ahmed Mohamed Nasr

Signature: ..... Date: 11/12/2020

2. Internal Examiner

Name: Nada Abd Ahmed M. Saad

Signature: ..... Date: 11/12/2020

3. Supervisor

Name: Prof. Mahmoud Ali Ahmed

Signature: ..... Date: 11/12/2020

cgs @ sustech edu. البريد الإلكتروني 83 769363 فاكس/ ص.ب 407

Sudan University of Science and Technology College

of Graduate Studies

DECLARATION

I, the under-signed, declare that I'm the sole author of the Ph.D. thesis entitled **"THE EFFECTIVENESS OF USING CORPUS LINGUISTICS ON VOCABULARY LEARNING AND RETENTION IN EFL CLASSROOM",** which is an original intellectual work. Willingly, I assign the copyright of this work to the College of Graduate Studies (CGS), Sudan University of Science and Technology (SUST). Accordingly, SUST has all the rights to publish this work for scientific purposes.

Candidate's name: Ahmed Fathy Farag-Allah Youssef
Candidate's signature:
Date: 10/2020

DEDICATION

This work is dedicated to my loving wife, Amani Abdel-hafiz, whose help and encouragement and resilience gave me the power and the will to resume this research successfully.

ACKNOWLEDGEMENTS

I give all praise and thanks to Allah who helped me through every stage of my thesis. I am so grateful to Dr. Mahmoud Ali, my Advisor, for his talent, benevolent, and great efforts during all stages of my research. I deeply appreciate his constructive comments, clear explanations, stimulating discussions, and patience. My deepest gratitude goes to Dr. Ahmed Al-Shobary, Assistant of Vice General Supervisor for Academic Affairs at Burayadah Colleges, for his great interest in my study and personal encouragement. Also, I really appreciate Dr. Waleed Radwan's support and help during the application stage of my research. Special thanks to the Committee members for their comments and questions during the discussion.

ABSTRACT

The rising number of available online corpora and the advanced concordancing software have been urging and boosting researches on potentials and uses of corpora in language analysis and teaching. In this study the researcher tried to fill in the gap apparent in the Saudi EFL context of poor or lack of corpora use. Therefore, the researcher has investigated the effect size of using corpus-based approach on vocabulary gains and retention. He has employed quasi – experimental study and randomly assigned two groups, the control (N= 29) and the experimental (N=25), to thoroughly investigate the study questions. To collect data, Vocabulary achievement pre-, post and delayed posttests have been employed with participants of both groups, the control which has been taught using conventional approach and the experimental which has been taught using corpus-based approach. In addition, the researcher has used a questionnaire to investigate learners' attitudes towards the use of corpus-based approach and another questionnaire to determine Saudi EFL teachers' awareness of corpus linguistics. At the end of the treatment test scores for both groups have been statistically analyzed, using descriptive statistics analysis and t-tests analysis. The results of posttests and delayed posttests showed a statistically significant differences at $P \leq 0.05$, in the mean scores of post and delayed post-tests in favor of the experimental group, which indicated the effectiveness of using corpus-based approach on vocabulary learning and retention. Findings also showed that the majority of learners (73%) had positive attitudes towards using corpus-based approach and they highly rated its benefits, while most of them (65%) illustrated facing some challenges in using corpora which could be met over the time and by much practice. Regarding teachers' awareness, the results revealed that the majority (63%) had low level of awareness of corpus linguistics and its applications in the classroom. Consequently, the researcher has recommended integrating corpus pre-service and in-service training for EFL teachers. He has also suggested enhancing friendly user interfaces for available corpora to be easily accessed and used by teachers and learners. In addition, combining conventional tools alongside with corpus-based method to meet some emerging challenges with corpora use. Lastly, the researcher concluded that supporting learners with sufficient training to possess technical skills of online corpora use and concordancing skills of linguistic analysis and interpreting is fundamental for successful integration of corpora in EFL classrooms.

مستخلص الدراسة ARABIC ABSTRACT

لقد أدى الانتشار الواسع للقواعد الالكترونية المفتوحة للمتون اللغوية إلى تحفيز وتشجيع الدراسات والأبحاث التي تتناول فوائد وإمكانات هذه القواعد المتوفرة الكترونياً في مجال التحليل اللغوي وتدريس اللغة.

في هذه الدراسة يحاول البحث سد هذه الفجوة الجلية في ميدان تعليم اللغة الإنجليزية في المملكة العربية السعودية المتمثلة في التطبيق الضعيف أو الاستبعاد التام لاستخدام طريقة التدريس المعتمدة على المتون اللغوية الالكترونية (المكنز النصي الالكتروني) في تدريس المفردات. ولذلك فقد قام الباحث بدراسة تأثير هذا الأسلوب التدريسي على اكتساب الطلاب للمفردات ومدى مستوى الاسترجاع للمفردات المتعلمة. وقد استخدم الباحث أسلوب البحث التجريبي والذي تضمن التعيين العشوائي للمشاركين على مجموعتين (تجريبية تضم 25 طالباً وأخرى ضابطة تضم 29 طالباً) وذلك من أجل الإجابة على أسئلة الدراسة، ولجمع البيانات فقد استخدم الباحث اختبار مفردات قبلي و بعدي و بعدي مؤجل مع جميع المشاركين من طلاب المجموعتين التجريبية والضابطة، كم استخدم أيضاً استبانات لاستطلاع آراء الطلاب حول فاعلية استخدام هذا الأسلوب التدريسي وتحديد مدى وعي وإدراك المعلمين بهذا الأسلوب التدريسي الحديث و مدى قناعاتهم بإمكانية تطبيقيه. وفي نهاية التجربة تم تحليل نتائج الاختبارات باستخدام اختبارات الفروق والاختلافات (t–tests) كما تم تحليل إجابات المشاركين على أسئلة الاستبانات المستخدمة باستخدام برنامج التحليل الاحصائي SPSS. وقد أظهرت نتائج تحليل نتائج الاختبارات أن هناك فروقات ذات دلالة إحصائية عن مستوى $P \leq 0.05$ في نتائج الاختبارات البعدية والبعدية المؤجلة بين طلاب المجموعتين لصالح المجموعة التجريبية (التي استخدم معها الأسلوب التدريسي المعتمد على النصوص)، وهذه النتيجة أكدت على فاعلية هذا الأسلوب التدريسي في اكتساب واستبقاء المفردات مقارنة بالأساليب والطرق التقليدية الأخرى. كما أظهرت نتائج تحليل استبانات الطلاب أن معظمهم (73%) لديهم انطباعات إيجابية وتقييم مرتفع لمزايا استخدام هذا الأسلوب في تعلم المفردات كما أنهم أيدوا وجود بعض الصعوبات أثناء الاستخدام ولكن الغالبية (65%) أكدوا على قدرتهم مقابلة هذه الصعاب مع مرور الوقت وكثرة الاستخدام. أما استبانات المعلمين فقد أظهرت أن الغالبية (63%) لديهم مستوى متدني من الوعي بهذا الأسلوب التدريسي ومزايا استخدامه في تعلم المفردات. وبناءً على ما سبق من نتائج الدراسة فقد صاغ الباحث بعض التوصيات والمقترحات والتي يمكن تلخيصها في النقاط الاتية: أولاً لابد من تدريب المعلمين قبل

وأثناء الخدمة على التطبيق العلمي لاستخدامات قواعد البيانات اللغوية الالكترونية والأسلوب التدريسي المعتمد على استخداماتها وأدواتها إلى جانب تطوير واجهة وأدوات هذه القواعد المتاحة أونلاين من أجل تسهيل استخدمها من قبل المعلمين والطلاب. كما يقترح الباحث الدمج بين الطرق والأدوات التقليدية وطريقة الاعتماد على المتون اللغوية في تدريس المفردات وذلك من أجل التغلب على بعض الصعوبات التي تقابل الطلاب في استخدام المتون بمفردها. وفي النهاية فقد توصل الباحث إلى أن دعم وتدريب الطلاب على المهارات التقنية والمهارات التحليلية والتربوية التي تمكنهم من الاستخدام الفعال لقواعد المتون الالكترونية يعتبر من أهم عوامل فاعلية تطبيق هذا الأسلوب المقترح المعتمد على المتون.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

NO.	ABBREVIATION	MEANING
1	ESL	English as a Second Language
2	EFL	English as a Foreign Language
3	ELT	English Language Teaching
4	CAT	Contrastive Analysis and Translation Approach
5	CLT	Communicative Language Approach
6	DDL	Data Driven Learning Approach
7	AWL	Academic Wordlist
8	SLA	Second Language Acquisition
9	L1	First Language
10	L2	Second Language
11	KWIC	Key Word In Context (A type of Corpus Query Tools)
12	EAP	English for Academic Purposes
13	COCA	Corpus of Contemporary American English
14	MICASE	Michigan Corpus of Academic Spoken English
15	BNC	British National Corpus
16	ACAD	Academic
17	FIC	Fiction
18	CALL	Computer Assisted Language Learning.
19	IELTS	International English Language Testing System
20	AFL	Academic Formulas List
21	CPT	Cambridge General Placement Test
22	SPSS	Statistical Program for Social Sciences Package (V.25).
23	ENG.	English
24	ANOVA	Analysis of Variances
25	BPC	Buraydah Private Colleges in Qassim Region, Saudi Arabia
26	EDMODO	An application and online platform for sharing, communication, collaboration, and coaching.
27	ESP	English for Specific Purposes.

LIST OF PUBLICATIONS

NO.	TITLE OF PUBLICATION	PUBLISHING JOURNAL	Volume and Date
1	The Effectiveness of Using Corpus-based Approach in Vocabulary Learning Gains and Retention Rates.	Journal of English Language Teaching and Applied Linguistics, ISSN (2707-756X) E-ISSN (2707-756X)	Vol. 2 (3) Aug., 2020
2	Learners' Attitudes Towards Using Corpus-based Approach in Vocabulary Classroom.	Journal of the College of Arts, BSU, ISSN (2090-9012) E-ISSN (2682-2776)	Vol. 5 (60) Jul.-Sept. ,2021

CHAPTER I

INTRODUCTION

1.0 Background

It can be seen that there is a high degree of agreement that vocabulary learning is an integral part of mastering a second language. The best way to attain solid and successful vocabulary learning, however, is still controversial, partially because it depends on a wide range of elements (de Groot, 2006). Consequently, it is not surprising to find teachers and learners sometimes confused about the right way to achieve it, especially as common coursebooks and teaching materials have not offered straightforward sufficient guidelines and descriptions.

Undoubtedly, vocabulary is fundamental to language and vital to the average language learner (Zimmerman,1997); while teaching and learning are considered major concerns for both EFL teachers and learners, solid acquisition of second language vocabulary is of a particular necessity for English as a second or as a foreign language learner (EFL/ ESL) who used to acquire dead abstract lexicons throughout long years of formal study (Hunt & Beglar, 2005).

Zhan (2008) ascertained a profound notion in this regard through emphasizing the impact of adopting convenient teaching methods and materials on getting learners acquire much deeper impression and richer information about the target words and making it easier to be saved and retained in a long time period. In the same regard, Laufer (2001) claimed that most learners acquire new vocabulary effectively through exposure to language input especially reading input, rather than the deliberate sophisticated process of memorizing words. It has been believed that one of the essential and vital resources for providing this language input with distinct features of authenticity and abundance is corpora-based approach which considered a revolutionary teaching trend in EFL classrooms (Sinclair, 2004). Over the last decades, there has been an increasing interest among EFL professionals, scholars, teachers, and material designers in proposing models, teaching methods, coursebooks, or dictionaries that are essentially based on corpus linguistics. Concurrently,

corpus linguistics as a discipline has been branching into sub-disciplines and methods that mainly pertinent to language teaching, like corpus-based approach, Data Driven Learning, Corpus-based Dictionaries…. etc. (Barbieri & Eckhardt, 2007).

Corpora then can be utilized directly or indirectly in teaching English classes; it can be indirectly employed in designing teaching materials and forming a foundation for new methods in syllabus design (Römer, 2011). Furthermore, it can be directly integrated in teaching and learning by giving learners direct access to corpora inside or outside the classroom to do learning activities and tasks (Gavioli & Aston, 2001, Römer, 2011). In this approach, learners are believed to become independent researcher and autonomous language learners to explore language patterns and cooccurrences, rather than passive knowledge recipients (Conrad, 2005). Corpus linguistics is believed to provide authentic data which eventually help learners to explore a variety of linguistics and metalinguistics features of vocabulary and test their hypothesis about language. Regarding to vocabulary teaching and learning, Kennedy (1991) claimed that corpus linguistics has offered invaluable tools and innovative teaching methodologies to EFL teachers, professionals and materials designers

This study has been intended to investigate the effectiveness of employing this brand-new approach in vocabulary classrooms with university level students in Saudi Arabia, hoping to participate in innovating and enriching EFL teaching context in Saudi Arabia and coin an effective model for teaching and learning the central part of language mastering, vocabulary acquisition. The first spark that initiated the researcher's interest in this topic is attributed to his successful and rich experience with teaching legal translation to languages department four years back using parallel corpora to teach legal texts and terminology. The following paragraphs will be presenting all essential elements of the study starting with problem statement and ending with common terminology.

1.1 Problem Statement

Leaning vocabulary is considered an essential and inevitable part of learning a language that's why there's a consensus among researchers, ESL teachers, educators and language professionals on reinforcing vocabulary teaching approaches and tools in EFL or

ESL classrooms. Previewing the recent literature on vocabulary learning and retention, the researcher found out a great number of recommendations and tools that have been vindicated and verified to be utilized in language classrooms. Though how diverse and controversial they were, most of them agreed upon the significance and effectiveness of implicit contextualized vocabulary learning rather than the explicit decontextualized one. Corpus linguistics as a new and trendy discipline has been claimed to best achieve these notions of contextualization, implicitness, and authenticity through providing authentic data of abundant concordance lines with some searching tools. Subsequently, there has been an increasing interest in conducting studies on employing corpus linguistics in the field of language teaching which contributed to proposing corpus-based methods, tools and material design recommendations.

The first incident that initiated the researcher's interest in the study dated back to 2016 when he was teaching legal translation course to students of Languages Department and he was first using parallel corpora as a tool for teaching legal texts and terminology. It was a fruitful and rich experience to see how effective it was to use authentic abundant data that not only provides various linguistic features but also provides some meta-linguistic features for the target vocabulary. Then the researcher started to preview some prominent studies on corpora utilization in teaching vocabulary and surprisingly found no studies (at that time 2016) conducted on the Saudi context except very few ones about checking wordlists of texts and making specialized corpus for ESP course.

On the other hand, the researcher realized that most teachers seem not to know about or aware of corpora as a teaching tool, while most of them may be unknowingly using corpus-based coursebooks or resources, such as: *Collins COBUILD English Dictionary, Oxford Advanced Learner's Dictionary, Longman Dictionary of Contemporary English and Cambridge International Dictionary of English, Interchange Series, and Reverso Context Website.* This finding indicated that, on the one hand, most teachers do use corpus-based dictionaries but that, on the other hand, they are not aware of the corpus-linguistic background of these products. Therefore, the researcher encountered the second gap in the Saudi EFL context.

To fill in these gaps, the researcher has designed this study to thoroughly investigate the effectiveness of this approach, corpus-based approach, on Saudi learners' vocabulary learning and retention, hoping that this study could contribute to raising awareness of Saudi EFL teachers and professional of exploiting this new approach in their classrooms, coursebooks and teaching materials. Moreover, the action classroom research employed in this study could exemplify the procedures of the application process and giving solid samples of corpus-designed activities and tasks. Then, this investigation may bridge the gap in the Saudi context through an action classroom research and contribute to the corpus literature.

1.2 Objectives of the Study

The main aim of this study is to identify the effectiveness of using corpus-based approach in vocabulary learning and retention in Saudi EFL tertiary classrooms and to measure learners' and teachers' perceptions and views on corpora as a trendy learning tool in their vocabulary classes. Bearing this broad aim in mind, the following objectives have been formulated:

1- To investigate the effectiveness of using corpus-based approach in improving vocabulary learning in Saudi EFL classrooms.

2- To investigate the effectiveness of using corpus-based approach in improving vocabulary retention in Saudi EFL classrooms.

3- To determine the effect size of learners' English proficiency level in increasing the effectiveness of using corpus-based approach in vocabulary learning in Saudi EFL classrooms.

4- To investigate the attitudes of the experimental group's students towards using corpus-based approach in their vocabulary learning classes.

5- To determine Saudi EFL university teachers' awareness of using corpus-based approach in EFL classrooms.

1.3 Questions of the Study

The following research questions have been formulated:

1. What is the effect of using corpus-based approach on improving vocabulary learning in Saudi EFL classroom?
2. How far is corpus-based approach effective in improving vocabulary retention in Saudi EFL classroom?
3. What is the effect size of learner's proficiency level (as a moderator variable) in the effectiveness of using corpus-based approach in vocabulary learning?
4. What are the attitudes of the experimental group's students towards using corpus-based approach in vocabulary learning classes?
5. To what extent are EFL teachers in Saudi Arabia aware of using corpus-based approach in EFL classrooms?

1.4 Hypotheses of the Study

The researcher has made the following hypotheses:

1- There are statistically significant differences, at the level $P \leq 0.05$, between the mean scores of post-tests for control group' learners that received conventional vocabulary instruction and those of the experimental group that received corpus-based instruction in favor of the experimental group.

2- There are statistically significant differences, at the level $P \leq 0.05$, between the mean scores of delayed post-tests for control group's learners that received conventional vocabulary instruction and those of the experimental group that received corpus-based instruction in favor of the experimental group.

3- There are NO statistically significant differences, at the level $P \leq 0.05$, between the mean scores of post-tests for learners of different proficiency levels in the experimental group that received corpus-based learning treatment.

4- The attitudes of the experimental group's learners towards using corpus-based approach in vocabulary classrooms are relatively positive.

5- University EFL teachers in the Saudi context have relatively low level of awareness of using corpus-based approach in EFL classroom.

1.5 Significance of the Study

Few studies have been conducted in the Arab world context to investigate the impact of using corpus on vocabulary in EFL classroom. This study may contribute to raise awareness of both ELT professionals and learners about the feasibility and efficacy of corpora integration in EFL classrooms. The study has also suggested practical classroom activities and hands-on for corpus-based approach in vocabulary teaching that may benefit interested teachers. The results of the random sampling of teachers may act as a reminder and stimulus for teacher training programs administrations to cope with the trendy research literature on the effectiveness of corpora and integrate corpora training in their programs. Furthermore, the results of this study may contribute somehow in identifying the effectiveness of one of the most recent trends in language teaching and learning, corpus-based approach which may provide teachers, learners, ELT professionals, and curricula designers with authentic, practical and real-life materials.

To the researcher's knowledge this study is considered one of the very few studies to be conducted on corpora integration in Saudi EFL classrooms and in the Arab world context too. Therefore, having identifying the impact of corpus-based approach by an experimental and empirical comprehensive study is substantial at this stage. Finally, it represents a contribution to the second language acquisition literature and teacher training and preparation programs.

1.6 Limitations of the Study

The study is limited to identifying effectiveness of using corpora in vocabulary learning in tertiary EFL classroom in the Saudi context. It's also restricted to the foundation level at a private college in Saudi Arabia, Buraydah Private Colleges in Qassim Region. It has been conducted on two groups of (54) male students and (42) EFL male and female teachers at different universities and colleges (Qassim University, Buraydah Private Colleges, AL Rajhi University, and Taif University) in Saudi Arabia. The motivation in choosing this group stems from its distinctive features. It has a collection of teachers who

have a diversity of cultural, academic and educational backgrounds. Consequently, the results of this study could be more reliable and meaningful.

1.7 Definitions of Terms

This part will be handling the key terms used in this study as follows:

1.7.1 Academic vocabulary

This refers to vocabulary frequently used in the academic written contexts like scientific abstracts, text books, articles, manuals, handouts and formal reports. One of the most famous lists for academic words is AWL, a list of 570-word families that are commonly found in academic texts (Coxhead, 1998). It was selected by examining massive corpora (or collections) of written academic texts and articles and selecting the words that repeatedly occurred.

1.7.2 Lexical or Vocabulary Knowledge

This term is repeatedly mentioned in this study. It basically refers to both parts of vocabulary knowledge, receptive or definitional part which mainly refers to the conceptual part of vocabulary whereas productive or transfer part refers to the employment of vocabulary items in novel contexts and uses.

1.7.3 Definitional Knowledge of Vocabulary

This part of vocabulary knowledge is regarded as the minimum requirement of vocabulary gaining as it mainly focuses on knowing one or more meaning for each word, it can be described as the breadth or the quantity of vocabulary knowledge. The learners' definitional or receptive knowledge in this study has been measured by students' ability to match each word with its meaning or translate individual words or matching synonyms or antonyms. The achievement vocabulary tests used in the study have been equally divided into two parts, definitional and transferrable with 25 points for each.

1.7.4 Transferable Knowledge of Vocabulary

This essential part of vocabulary knowledge refers to the depth or the quality of vocabulary knowledge rather than the breadth or the quantity of vocabulary. It includes the skill of transferring the learnt words into new contexts and interpreting the meaning of words while reading authentic contexts. The researcher has weighed both parts of vocabulary knowledge through focusing and testing both of them in a balanced way. Transferrable or productive part has been tested in the Cloze part in vocabulary tests.

1.7.5 Vocabulary Retention Rate

This term refers to retaining vocabulary knowledge in a long-term period, while in this study it refers to retention rate of vocabulary knowledge in a time period of three weeks after the posttest. The researcher has designed a delayed posttest to measure this rate which can be measured by comparing learners' mean scores of immediate posttest and delayed posttest.

1.7.6 Learners' attitudes

In this study, it refers to experimental group's views on the application of corpus-based activities and tasks in their vocabulary classroom. These views have been collected through a questionnaire of 16 items of 5-point Likert scale.

1.7.7 Corpus-based Approach

This method mainly included activities and tasks that mainly based on corpora direct or indirect use whether through online use of corpora or employing corpus-based printouts. In this study, the researcher has adopted a gradual approach for corpora integration in the classroom starting with corpus-based printouts, then corpus-based activities and hands-on mainly led and executed by the teacher and students had a little to do, lastly activities and tasks that required direct access of groups and individuals to corpora.

1.7.8 Conventional Teaching Approach

This method refers mainly to the standard method of teaching vocabulary which mainly depends on the assigned coursebook and printouts and in this study the assigned coursebook is *Interaction Reading 1 Coursebook by Kirn, E and Hartmann, P. (2007).* Through this method many linguistic aspects of vocabulary can be taught through ready-made activities and tasks in the coursebook such as reading texts, exercises and tasks.

CHAPTER II

LITERATURE REVIEW

PART I: THEORETICAL FRAMEWORK

2.0 Introduction

The literature review consists of two main parts, theoretical framework and previous studies. The first part, theoretical framework, will be handling theories and arguments pertinent to corpus-based approach, teaching vocabulary methods, and Saudi EFL context, while the second part will be an elaborated discussion on and analysis for previous studies on Corpus linguistics integration in EFL or ESL classrooms at international, national and local contexts. Thus, theoretical framework will be handling various items to ensure thorough covering for all components of research topic. Specifically, it will preview discussions and analysis of the context of English language teaching in Saudi Arabia, followed by a brief history of corpus linguistics and Data Driven Learning approach and their relationship with SLA theories, especially CLT approach of teaching English. The last item of this part will be devoted for illustrating corpus linguistic use in vocabulary teaching and overall summary for the whole part.

2.1 The Situation of English Language in Saudi Arabia

As a result of globalization, English has acquired a prestigious status as a "lingua franca" or a widely international used language for communication, diplomacy, trade and education. As a natural result, there has been a severe need for learning English worldwide during the last decades. Saudi Arabia, as other nations, uses English in international communication which in turn enhances EFL context and promotes teaching English industry, which became a mandatory school subject for the 4^{th} grade of primary stage and on.

In the same vein, the discovery of oil in Saudi Arabia in the twentieth century was considered to be one of the major factors that forced spreading English language teaching accordingly. Oil experts has been brought from English-speaking countries to participate in

extracting oil and excavation operations, which eventually prioritize English language teaching as part and parcel of oil industry whether production processes or exportation. Recent linguistics justified this phenomenon under the recent coined term "Petro-Linguistics" to refer to the strong impact of oil industry in some oil producing countries on English language teaching (Mahboob & Elyas, 2014).

Mahboob and Elyas (2014) mentioned another major reason for the spread and increased interest of English teaching in Saudi Arabia; it was the development of military relationships between the United States and Saudi Arabia Therefore, teaching English has been promoted at the official level through curricula development, recruiting native speakers in some projects, prioritizing learning English at an early age, setting language minimum requirements for many university programs and public recruitment. All the above-mentioned reasons and others have desperately impacted EFL context in Saudi Arabia.

2.1.1The History of English Teaching and Methodologies in Saudi Arabia

In 1958, English language was introduced as a subject to be taught in the kingdom's public schools (Mahboob & Elyas, 2014). Form that date and on, English teaching has been receiving increased interest form the Saudi Ministry of Education to meet the urgent needs of economy, diplomacy, and military issues. This interest could be noticed in extending the studying years at schools, setting language proficiency requirements for some university majors, developing course books and curricula, extending scholarships grants to study English abroad, encouraging international schools, and setting language requirements for prestigious governmental and public jobs. All these taken steps have participated in promoting EFL context in Saudi Arabia and raised the overall language proficiency among Saudi citizens.

On the other hand, in the teaching or learning cycle one essential element received no or very little interest; it is teaching methodology which by so far affect the level of learners' proficiency. Mahboob and Elayes (2014) in their study stated that the most prevailing teaching method in Saudi EFL context is the tradition method, grammar-translation, which requires grammatical patterns and vocabulary items memorization. This

idea could interpret the considerable low learning outcomes of Saudi learners with such generous official support and availability of materials, equipped classrooms, and prestigious coursebooks. To recap, the cornerstone in an effective English language teaching context is a well-trained teacher that has the ability to employ modern effective teaching methodologies. The following paragraph will handle the challenges in the Saudi EFL context that could interpret low learning outcomes.

2.1.2 Challenges for Effective English Teaching in the Saudi Context

In his recent study Al-Nasser (2015) stated that the average Saudi learner has been taught English for 9 years at school, though, he in oftentimes cannot utter a complete meaningful sentence. This extreme and strict finding by Al-Nasser reflects how weak the Saudi EFL context is and How far it is from the normal expected outcomes. This finding draws the intention back to the main reasons for such weakness in proficiency levels, while there are high levels of equipment, courses books selection and language policies for university admission and for public recruitment. In his study Al-Nasser (2015) tried to probe this issue through investigating the potential challenges and difficulties in the Saudi EFL context. He, then, summarized the challenges five points: 1. The interference of the first language, Arabic, which led to defects in learning English. As Arabic and English have so many differences in their phonological system, morphology, spelling, and grammar, EFL learners get confused in acquiring certain linguistic aspects of word patterns or grammatical patterns, 2. As a religious and conservative community, Saudis have a strong feeling of identity which in some cases prevents from learning a language which could be regarded as a tool of weakening the identity and distorting culture, 3. Saudi English teachers are not well-prepared pre-service nor well trained in-service which led to poor language presentation and the use of outdated teaching methods, 4. Outdated curricula which included poor study plans, non-measurable and non-clear learning outcomes, poor assessment methods, and poor teaching methods.

In the same vein, Fareh (2010) illustrated other challenges in the Saudi EFL context for example: 1. Practical Teaching practices and activities are not directed to developing learner's language skills, 2. Activities and tasks assigned inside or outside the classroom

are not student-centered, they are teacher-centered, and 3. Lack or even absence of employing modern teaching material inside the classroom.

2.2 Vocabulary Knowledge

Knowing a word is not as simple as acquiring its meaning and spelling, but there are many aspects or elements of knowledge that have to be considered in learning or teaching vocabulary. From another viewpoint, vocabulary knowledge can be broadly divided into two integrated parts, the first is definitional or receptive knowledge which refers to knowing the meaning, pronunciation and spelling of a word, while the second is the transferrable or productive knowledge which refers to using the learnt words in novel contexts, written or spoken. It is assumed that the definitional knowledge of vocabulary is the first to be learnt then comes the much deeper knowledge, the productive. In the following paragraphs, the researcher will handle both types in details.

2.2.1 Definitional knowledge and transferable knowledge

Vocabulary acquisition is concerned with the depth and breadth of knowledge. According to Qian (1999, p.282), *"breadth of vocabulary knowledge"* is defined as the number of words which a learner has at least some minimum knowledge of meaning. On the other hand, "*depth of vocabulary knowledge"* is defined as a learner's level of knowledge of various aspects of a given word. This means that definitional knowledge of a word is the shallow level of understanding and it occurs at early stages of learning the word. Nagy (1997) believed that *"Definition-based learning typically involves memorizing (or attempting to memorize) brief definitions representing only a single meaning of the word to be learned, and hence lead to only a shallow level of word knowledge"* (P.73). However, memorizing word definitions cannot help students to understand and comprehend the word in different contexts. It is believed that the ability to transfer lexical knowledge to new contexts demands deep lexical knowledge, and transferable knowledge is possibly enhanced by setting optimum conditions for vocabulary learning.

According to Cobb (1997), learning words in different contexts can increase quality of word knowledge since students are able to transfer their new lexical knowledge to

reading comprehension. Similarly, Nagy (1997) suggests that if students are exposed to different contexts, they will gain and comprehend new vocab. He insists that most lexical knowledge is attributed to encounters with the words in various contexts since no single encounter with a word can lead to any great depth of word knowledge.

2.2.2 Incremental Nature of Vocabulary Acquisition and Retention

Vocabulary learning is considered to be a gradual process of building up pieces of constituent knowledge over a time period and through different encounters of words. This notion is called incremental nature of vocabulary learning that cannot be fully met with conventional teaching tools or methods. Each word is believed to be multi-dimensional and multi-faceted that needs to be studied from different perspectives and be encountered in a variety of contexts. Nation (2001) pointed out that there is so much to know about each word that one meeting with; it is not enough to know the definitional or simple aspects of a vocabulary items, but other deep components have to be learnt as well. He elaborated that this deeper knowledge of a word facilitates its use and availability on a long-term memory. Likewise, Schmitt (2000) noted that complete mastery of a vocabulary item requires a various linguistic aspects of vocabulary knowledge which cannot be learnt simultaneously, rather they can be learnt through many encounters of a word in different contexts. Accordingly, the more a learner meet a word, the better he learns it.

In other words, illustrated the incremental nature of vocabulary knowledge as a three-dimensional knowledge. The first dimension, learners can have knowledge ranging from zero to partial to precise. The second, depth of knowledge requires mastery of a number of lexical aspects. The last dimension is receptive and productive mastery. It can be concluded that a word or a vocabulary item can be learnt receptively before productively.

Based upon the above-mentioned discussion, learners need more opportunities to encounter target words in a variety of contexts and activities to fully understand them and develop productive knowledge of learnt words. In the same vein, Sokmen (19974) noted that re-encountering the word in different contexts has a significant impact on the retention rate and long-term memory. It can be concluded that the more learners encounter a word in various contexts, the better they learn and retain it.

To recap, the depth of vocabulary or lexical knowledge can be enhanced by words reencountering in a variety of contexts so that the acquired knowledge can be easily accessed, retrieved and used in novel contexts. Thus, worlds reencountering in different contexts deepen the attained knowledge and facilitate later retrieval processes.

2.3 Vocabulary Instruction in EFL Context

Since the field of English as a second/foreign language (ESL/EFL) was developed as a discipline in the 1950s, there have been various teaching approaches and methods, with different treatment for vocabulary from each approach. The perspective and the significance of vocabulary teaching in EFL context has varied and changed according to the principle and priorities of each approach. At early stages, vocabulary learning has been regarded as a supplementary element that could be learnt alongside other language components; i.e. grammar, functions, structures, and communication strategies. Therefore, in those traditional approaches, vocabulary has been seen as a subservient component to language analysis and language use (Zimmerman, 1997). For example, In the traditional approach, grammar-translation, phonology and syntax were prioritized, while in notional and communicative approaches communication and language use skills were predominant. In all these approaches, vocabulary has been regarded as a supplementary component that has been introduced while teaching grammar or functions or communicative texts.

From 1980s and on, an increased attention has been given to vocabulary learning as an essential part of language mastery. Schmitt (2000) and O'Dell (1997) noted that this increasing interest in vocabulary learning resulted from the effect of rapid technological advancements, especially the advent of computerized analyzing tools. As a result of these technological advancement, lexicographers and linguists were able to conduct extensive and objective studies and analyses based on large language corpora with far more details than ever. Knowledge from corpora enhanced knowledge of vocabulary 's value. Findings from the studies of real language usage reflected the perceived need for more detailed representation of the language. Some studies have given rise to considerable interest in the sense of large chunks of language, variously known as *lexical items*, *lexical phrases* and

prefabricated units. Lexical items are also claimed to be fundamental to language usage and should be fundamental to language instruction (Zimmerman, 1997).

These changes in language description and analysis have led to relevant changes in English language teaching discipline, especially in syllabuses design and teaching practices. O'Dell (1997) illustrated that four main new editions of ESL dictionaries were published in 1995 with substantial perspectives based mainly on lexical insights obtained from available huge corpora. Furthermore, it has been claimed that two basic tools for Teaching English as a second or foreign language have been developed with large corpora, the first was the compilation of accurate word frequency lists that allow more confidence in word selection and grading, the second was the obtained concordance outputs that manifest the actual uses of a word in different contexts and show collocations.

In summary, over the past two decades there has been an increasing belief in vocabulary significance in EFL and ESL contexts. With the reorientation in language description based on corpus-based studies, the perception of language nature and vocabulary roles has changed remarkably. Schmitt (2000) stated, *"Insights from corpus research have revolutionized the way we view language, particularly words and their relationship" (P.68).* Recently, vocabulary learning has gained a prominent status in ELT, and is no longer subservient to other components of language learning.

2.3.1 Approaches to Vocabulary Instruction

In modern lexical research, the emphasis in vocabulary instruction in EFL or ESL context were mainly on two extreme vocabulary teaching approaches. Coady's (1997) stated that one extreme approach is "*incidental learning*" which requires exposure to language use in different contexts, while the other extreme one is "*explicit learning*" in which a deliberate words teaching is provided. The "*strategy learning*" came in the middle of these two contrastive approaches of vocabulary instruction, which combines deliberate words teaching with exposure to language uses contexts. For incidental learning, there is not any deliberate words teaching rather students learn words naturally during their exposure to real language use. In contrast, explicit instruction or learning requires formal

deliberate instruction of words by using various techniques. A detailed description for each approach will be provided in the following sub-sections.

2.3.2 Incidental Learning

"*Incidental learning*" or what is called "*contextual learning*" is an approach of teaching vocabulary which imitates a natural process of L1 acquisition by exposing learners to a variety of contexts when their attention is not on the language itself, but rather on the use of language. Studying words in different contexts is believed to enhance comprehension of words deep meanings. Schmitt (2000) stated, *"Incidental learning can occur when one is using language for communicative purposes and so gives a double benefit for time expanded" (P. 120).* In Coady's (1997) study, the contextual acquisition research ascertained that most vocabulary knowledge comes from meaningful language encounters. Learning is more successful if the language is meaningful, rich in content, authentic, and the most important, comprehensible. In order to strengthen incidental learning, learners have to read a large number of texts in order to acquire new words.

On the other hand, the major drawback of this type of vocabulary learning, incidental learning, is claimed to be its imitation of natural acquisition processes which happen slowly and gradually. Thus, successful vocabulary learning takes a long time to take place, while learners normally have limited time for their study. For example, in learning language for academic purposes students cannot learn the necessary skills fast and efficiently enough if they try to adopt this time- approach. Moreover, learning from contexts will be successful only if the occurrences of target words are incidentally frequent enough and this does not happen especially with relatively infrequent words.

2.3.3 Strategy Learning

"*Strategy learning*" is related to the communicative approaches of the 1970s and 1980s. According to these approaches, context is a major source of vocabulary learning. This approach is a compensation for the limitation of incidental learning as it focuses on how well students can deal with contexts on their own. It emphasizes teaching specific

strategies to enable students to learn from contexts effectively and it also pay a great attention to inference skills which are the primary strategies to deal with new or unknown words. According to (Sokmen 1997), Learners are taught the strategies of "*inferring from contexts*" by recognizing clues in contexts, using monolingual dictionaries, and avoiding bilingual equivalences. Acquiring Vocabulary happens mainly through guessing words in contexts.

This approach is appealing to many scholars and it is greatly used in several EFL textbooks. However, Coady (1997) stated that some scholars claimed some disadvantages in using this approach such as the slow process of inferring a meaning from context which can also cause frustration to some students.

2.3.4 Explicit Learning

In this approach of explicit learning, students are deliberately taught the target vocabulary through using one or more of lexical teaching methods. It is direct attempt to teach the assigned lexical items using a variety of techniques or methods. This approach contrasts the implicit learning approach which requires unintentional vocabulary learning with direct lexical instruction. Various studies have investigated this approach and concluded that explicit learning has to be adopted alongside with other approaches to highlight the target words and deliberately learn some aspects of word knowledge (Sokmen, 1997). Moreover, as a result of the limited time of classes, learning has to be accelerated at some stages and learners need formal instruction to prepare themselves to cope with the demand of real language use.

There are many teaching techniques that can be used in explicit instruction method, even deliberate memorization of target lexical items, provided that they serve particular learning objectives. For example, incidental learning can be connected to this approach, but in a guided simple way such as glossing texts, using simplified readers etc. In addition, inferring strategies can also be explicitly taught alongside any other method. Recently, there are various techniques that enable students to learn vocabulary explicitly. Various tasks, activities and guidelines have been suggested in many studies related to vocabulary and lexicon. The most important is Nation's (2001) who recommended three essential steps

that could lead to vocabulary acquisition. These steps were '*noticing*', '*retrieval*' and '*creative (generative) use*'. The first step of '*noticing*' is to encourage learners to notice the word as a useful language item. The second step of '*retrieval*' is to put some activities for the new learned words so that they can be retrieved easily and at the same time the memory of those words will be strengthened. The last stage of '*creative or generative use*' is to provide activities and opportunities for learners to use new learned words. In addition, Coady (1997) illustrated three main principles underlying effective teaching and they are providing both definitional and contextual information, allowing students to process information, and facilitating multiple exposures of each word.

2.3.5 Learning Vocabulary Through Reading

Vocabulary learning approaches have gone back and forth in the fields of Applied Linguistics and Language Acquisition. As previously mentioned, the big change was the shift from the direct instruction of lexical items in the grammar-translation method to the incidental learning in the communicative approach, and now a reconciliation between implicit and explicit teaching. Actually, all these three current approaches should be studied as complementary to one another. It is possible to integrate these approaches by '*learning vocabulary through reading*' and this seems to be the best practice at present. Schmitt (2000) noted that learners could learn vocabulary from reading. Moreover, Coady (1997) noticed that *systematic vocabulary instruction strategy* together with *learning vocabulary through reading* is a more successful integrated approach than simply learning through contexts alone.

With the method of learning vocabulary through reading, the integration between different approaches of explicit learning, strategy learning and implicit learning is possible. Although implicit teaching is the most preferred approach at present, it cannot be applied to all teaching elements of lexical knowledge due to time limitation. Therefore, it is impractical for implicit approach to contextualize all assigned lexical items or utilize all the creative uses of a word for students to master them all.

Many techniques have been suggested in order to teach vocabulary through reading. Nation (2001) and Schmitt (2000), for example, suggest that certain words in mother texts

for reading may be made prominent, such as by extracting them clearly at the books' margins. Moreover, *intensive reading* of short texts is needed to facilitate text comprehension and to direct attention to other elements of text such as: grammar, vocabulary and discourse. By employing intensive reading strategy, a number of vocabulary and reading exercises could be added to each reading passage. Moreover, extensive reading is also possible by using *graded reading* with beginning students, *narrow reading* with intermediate students, and a wide variety of authentic texts with advanced students. *Graded reading* is some authentic books which are graded according to levels of readability whereas *narrow reading* means reading numerous authentic texts on the same topic, so vocabulary will be repeated throughout the course of reading. According to Schmitt (2000), one of the benefits of narrow reading is that it can facilitate access to authentic materials.

2.3.6 A Concordance-based Approach

A *concordance-based or corpus-based* is an approach that was initially employed in the fields of Lexicography and Corpus-Linguistics and then became a trendy proposed approach in teaching vocabulary and grammatical patterns in EFL/ESL context. This method involved compiling a corpus from authentic texts, then using a concordancing program for navigating this compiled corpus, and finally obtaining concordance outputs for instructional reasons. It is worth mentioning here that both terms *"corpus-based"* and *"concordance-based"* are used interchangeably as they both refer to the recent approach of employing online corpora of authentic compiled texts in teaching or learning English lexical items, grammatical items, or even translation skills.

This approach is claimed to be promising and revolutionizing, especially in vocabulary learning contexts. Firstly, it can provide a unique resource of language, so it can serve students' needs. Secondly, through this approach frequency wordlist can be quickly created to be used in selecting target words. Thirdly, thousands of words can be contextualized, obtained and searched through a concordancer. Through using this approach, the materials can be prepared much more conveniently and quickly in order to present target words in various authentic contexts with ample encounters of authentic

language samples. This enhances students' abilities to learn large number of target words in a short time. Therefore, vocabulary learning in different contexts can be facilitated and a sufficient number of word encounters can be achieved. Moreover, the corpus-based approach is assumed to be used alongside other methods of vocabulary teaching, i.e. implicit learning, explicit instruction, strategy learning, and learning through reading.

Corpus-based approach is compatible to almost all other methods of vocabulary instruction. For example, in explicit instruction, assigned lexical items can be highlighted through the display of *KWIC (Key Word In Context)* concordances where the key words are showed in the middle of concordance lines. Consequently, learners' attention is directed to the target words being learnt and at the same time they can easily observe lexical patterns in various contexts.

Similarly, it can be used with strategy instruction as it is noted to be helpful in making the clue words highlighted. For example, students may inductively discourse markers items by looking up words such as "Although" and "famous for", and then infer how these words give clues to the meaning of unknown words. This is to strengthen students' inference skills in guessing unknown words from context clues. In the same vein, corpus-based approach can be used with incidental learning through providing authentic texts and lines to be used as contexts for unintentional learning of key vocabulary.

In the current study, the researcher adopted much more flexible approach than the strict method of applying DDL approach which requires complete learners' autonomy and complete free selection of words being learnt. Therefore, he used a gradual integration of corpora in vocabulary instruction to ensure learners' familiarity to technical skills and concordancing skills required for using corpora. The main aim from this gradual introducing of corpus-based approach was ensuring learners' get familiarized with technical and inferring skills to be able to deal with obtained concordance lines.

2.4 An Overview on Corpus Linguistics

Corpus Linguistics is a modern discipline in the field of linguistics that has undergone rapid advancement and popularity since 1964 when the first online corpus, The

Brown Corpus appeared. Corpus linguists are mainly interested in descriptive or functional interpretations of language (Meyer, 2004), and study linguistic phenomena through the empirical analysis of large computerized databases of language called corpora (corpus, sing.). A corpus is *"a large and principled collection of natural texts"* (Biber, Conrad, & Reppen, 1998, p. 4), which is compiled so that it is representative of the language in general, a dialect, or any specific genres of a language. Corpora may contain linguistic data relevant to written texts, transcribed speech, academic or general or a combination of two or more register or genres. These texts are stored electronically, and then can be analyzed using computer software programs called *concordance generators*, *concordances*, or, generically, *concordancing software* (Conrad, 2005; Tribble & Jones, 1990).

The idea of collecting massive numbers of texts of different genres and registers for long time periods for the purpose of linguistic analysis wasn't realized as totally new when corpus linguistics arrived as a methodology. As Meyer (2009) pointed out in one of his recent papers that early dictionaries were based on a large body of published works and millions of citations slips of naturally occurring language. Furthermore, concordance lines *(i.e., a word displayed within a surrounding context)* as a format for displaying every instance of a word in a text or collection of texts has been around for centuries, as Tribble and Jones (1990) explain:

> *"In its original sense a concordance is a reference book containing all the words used in a particular text or in the works of a particular author (except, usually, the very common grammatical words such as articles and prepositions), together with a list of contexts in which each word occurs. […] Books like this have been in use since the Middle Ages, especially in Biblical Scholarship. […] The earliest known complete concordance of the Latin Bible was compiled by the Benedictine Hugo de San Charo in the thirteenth century. Hugo, it is said, was assisted by no fewer than 500 monks." (p. 7)*

Recently, in modern linguistics, the work of collecting and analyzing large databases of language, though still a time-consuming and tedious task, has been greatly simplified and largely automated by powerful computers and concordancing software programs, of which Laurence Anthony's AntConc (2013) is just one example. These powerful software programs have the capacity to swiftly and accurately identify each word in a text together with its surrounding context, including even those very common words

like articles and prepositions that were neglected, and often not included, in concordances that were collected manually, such as those by the monks of Hugo de San Charo. In addition to their capacity to search for keywords and their surrounding contexts, these programs also have the ability of calculating frequency information about words, which is often presented in the form of hierarchical lists (usually with the most frequently occurring word appearing at the top of the list). Furthermore, computerized corpus searches are not limited to word-level searches, users are also able to navigate for two or more words *(i.e., collocations),* phrases, clauses, or, if the corpus is tagged, for grammatical categories *(e.g., prepositions and articles).* This short preview of history of corpora illustrates a crucial point that corpus linguistics is not a coincident to the advent of computers or the integration of technology in the field of linguistics, instead it is considered an old linguistic research substantially supported and advanced by modern technology.

2.5 Corpora in Linguistic Research and Language Teaching

The beginning of computational corpora was essentially for linguistics analysis purposes "for finding out about language and texts" (Leech, 1997, p. 2). Today, nearly every subdiscipline within linguistics uses corpora, to a greater or lesser extent, to inform their studies. Despite corpora have been used by linguists for research purposes for over forty years, researchers who are also language instructors have begun to have increasing interest and eagerness in utilizing corpora in ELT classrooms. According to Leech (1997), corpora can have a direct or indirect effect on the language classroom. On one hand, they are indirectly impacting language classrooms by using materials developers to create improved reference materials (e.g., dictionaries, grammars, and thesauri) and textbooks. Furthermore, they are being employed by language instructors to inform syllabus and course design (Flowerdew, 1996), and to create tests (Coniam, 1994; Shillaw, 1994). Moreover, corpora have been used to create both general academic (Coxhead, 2000, 2002), and discipline specific (Wang & Guang-chun, 2008) wordlists. Wordlists like Coxhead's (2000) Academic World List (AWL) contain the most frequently occurring headwords of a discourse; in the case of the AWL, the words are those which occur most frequently in general academic discourse, regardless of discipline. Coxhead's list is based on three

principles: teaching the most relevant, useful, and frequent lexical items to students first. The list has substantially contributed to the prioritization of vocabulary selection for the EAP curriculum. However, while useful for prioritizing vocabulary instruction, wordlists need to be taught using a principled approach to teaching vocabulary accompanied by appropriate classroom techniques in order to assure that students acquire and are able to appropriately and smoothly use these words in their own speech and writing. Corpus-based methods and activities can help attaining this goal through authentic materials and useful software tools. This brings us to the other hand of this discussion of how corpora are having a direct effect on the language classroom. For illustration, there are two ways to directly engage second language learners in corpus work in the classroom: 1) they can be given direct access to a corpus and concordancing program on a computer; or 2) they can be given prepared print-outs containing the raw data, or concordance output, from a corpus. This study has handled both approaches of corpora integration in EFL classrooms putting into consideration the gradual application. Primarily, The researcher will mainly focus on teaching academic vocabulary to learners who had never engaged in corpus-based work prior to the study, and therefore needed to be exposed to this "first stage" of corpus consultation in order to become familiar with how to use and analyze concordance results (Leech, 1997; Chambers, 2007); then, in the second stage students will ultimately be given access and taught how to use online corpora to encourage and support autonomous language learning beyond the classroom. In these paragraphs the researcher is handling the term Data Driven Learning and its history in ELT research. The term *"Data-driven learning" (DDL)* has been coined to refer to activities in which language learners are given print-outs to figure out the targeted language patterns in their EL classrooms.

It was first coined by Tim Johns (1991) to describe a method he used, and was largely responsible for developing and popularizing (Johns, 2000a, 2000b). Johns' DDL is essentially based on the methods used by the linguists who were involved in the COBUILD project at Birmingham University. This project, directed by John Sinclair, extensively utilized key-word-in-context *(KWIC)* concordance data in order to provide a range of reference and teaching materials for English language learners, most notably dictionaries. At that time, this project was pioneering and inspiring as it was the first attempt by

lexicographers to create a comprehensive profile for each word entry in a dictionary based on empirical evidence of authentic native speaker use of the language. John Sinclair was encouraged by the benefits of this type of technology integration in language analysis to adopt this approach in language teaching classrooms. Therefore, he developed DDL in order to *"cut out the middleman as far as possible and to give the learner direct access to the [language] data"* (p. 30). Johns (2002) accurately and intensively differentiated between the goals of his early COBUILD project and his newly invented approach DDL in the following excerpt:

> *"From the start, it was clear that there would be a small but significant difference between the approach taken by our colleagues in COBUILD and what I wanted to do in the English for International Students Unit. In the COBUILD materials the data was to some extent 'hidden' from the learner by the team of researcher [sic] and lexicographers. My approach was rather to confront the learner as directly as possible with the data, and to make the learner a linguistic researcher. The metaphor I use with my students is that of the detective. (pp. 107-108)*

Corpus based approach and SLA theories. Over the years, a significant body of research has been developed which boosts employing corpus-based methods and activities, such as DDL, in ELT classrooms. Various solid linguists' arguments have been made which conceived corpus work aligned with the principles of second-language acquisition *(SLA)* and positioned it within the communicative language teaching *(CLT)* approach. Researchers and linguists who believe in the positive facilitative aspects of corpus-based methods in ELT argued that the same procedures followed by linguists in doing their language descriptive studies can be easily taught to and utilized by language learners to be given a direct access to corpora in their classroom for more effective SLA processes

This can lead to reassigning conventional classroom roles, whereby students become linguistic "researchers", and teachers become monitors or coordinators of research (Gavioli, 1997, 2001; Johns, 2000a, 2000b). However, Bernardini (2002) states that *"descriptive insights and research methodologies have not simply been borrowed from the descriptive paradigm, but have been adapted, reformulated, and often extended in various ways to fit pedagogic concerns and priorities"* (P. 29). In this newly structured classroom, students are encouraged to engage in linguistic research which involves raising questions about the target language and engaging in a process of hypothesis formation and testing of

particular rules of the language, a process by which interlanguage development is thought to progress (Conrad, 2005). Furthermore, Aston (1995, 1997) argued extensively that *"corpora can play a useful role in the acquisition and restructuring of schematic knowledge"* (P. 263). His argument is that concordance lines expose learners to authentic contextual repetition and variation of linguistic structures, promoting learners' competence and capabilities of synthesis and analysis of information, which, in turn is substantial for acquisition. Many researchers have also concluded that engaging students in corpus-based activities enhances noticing or consciousness-raising (Kettemann, 1995; Johns 1991a, 1991b; Conrad, 2005).

2.5.1 Corpus-based Activities and CLT Approach.

In addition to facilitating second language acquisition, corpus-based activities are deemed as being aligned with various principles and essentials of Communicative Language Teaching approach *(CLT),* which currently dominates English language teaching (ELT) discipline. First, concordance data exposes learners to linguistic features in authentic contexts, which enable them to induct linguistics and grammatical patterns.

Furthermore, the new pattern of learner-teacher roles enhances student-centered classroom and shifts learning control from teachers to learners. Hence, corpus-based activities are thought to increase learner autonomy *"as students are taught how to observe language and make generalizations rather than depending on a teacher"* (Conrad, 2005, p. 402). The next paragraphs will preview the changing roles of teachers and learners in corpus-based approach.

2.5.2 Teachers' Roles and Learners' Roles in Corpus-Based Approach

Undoubtedly, there are dramatic changes in the assigned roles for both teachers and learners in corpus-based classrooms compared to their conventional roles in traditional EFL classrooms. Corpus-based approach encourages and emphasizes the active role of learners, moving the focus from teacher-centered mode to learner-centered. Teachers' roles are limited to monitoring, guiding, encouraging, organizing and facilitating the process of learner's knowledge constructing and discovery. On the other hand, the learners' role

broadens to be language researcher, patterns and meanings discoverer, problem solver and task achiever. This dramatic change in roles ensures infotainment, authenticity of learning and positivity in learners' side since they move from passive recipients into active constructors and discoverers of knowledge. The learning process makes them get the satisfaction of success, enhance their self-confidence, and further stimulates their interest in learning. For instance, they are supposed to sum up the concordance lines to come up to language patterns under the guidance of their instructors. It assures the student-centered exploratory learning approach, which essentially changes the conventional top-down teaching mode to bottom – top.

2.6 Theories and Approaches Pertinent to Corpus-Based Approach

Many theories and approaches of Second Language Acquisition and ELT have been used to theorize for and lay out the basics of corpus-based approach. Theories of Constructivism, Lexical Grammar, and DDL and Problem-solving approaches have been used as a foundation and justification for the viability of Corpus-Based Approach. It is quite obvious in Zhen (2005) description of Corpus-based approach as an *"autonomic learning", "authentic language input", "self-discovery" and "bottom -up inductive learning"*. The next paragraphs handle all these theories and approaches to clarify the shared basics and foundations with corpus-based approach starting up with theories followed by approaches.

2.6.1 Constructivism and Lexical-Grammar Theories and Their Relationship with Corpus-Based Approach

According to (Feng-ran, 2009) *Constructivism* as a worldview and widely- known theory assume that people build up their learning through gradual active autonomous process in which each learner actively constructs or creates his own representations through adding new information to prior knowledge, thus mental representations are subjective. According to the viewpoint of constructivists, the process of learning is a personalized constructive process in which each learner builds up his own internal perception of knowledge. This personal perception is constantly amenable, its structure and linkages

constitute the base to which new knowledge structures are linked. This viewpoint of knowledge posits that all we know of the world are human perceptions of our experience of the world. Conceptual growth comes from the sharing of various perspectives and the simultaneous changing of our internal representations in response to those perspectives as well as through cumulative experience.

For all the aforementioned details Zhang (2010) assumes that The Theory of Constructivism is the theoretical framework of "the English teaching model based on computer and classroom" (P.59). it has presumably believed that Data-driven learning incarnates the basics of Theory of Constructivism through enabling language learners to be language researchers and construct their own knowledge.

Lexical Grammar Theory is an essential contribution of the linguist Sinclair for the literature of language teaching and learning. Sinclair (2009) assumes that practically language system cannot be divided into its constituents, lexis and grammar, as grammatical features are determined and defined by lexis and lexis are normally used in grammatical patterns. He notes that the meaning of a word can be identified by combining both lexical and grammatical components as both of them substantially contribute to the intended meaning of a word or a phrase. This Theory assures that teaching a language should begin from lexis and their basic meanings and the most typical collocations, especially for highly frequent ones. In this theory, the meaning of word depends on their surrounding lexical items, the context of the words, which can be formed by word's collocation, colligation, semantic preference and semantic prosody. All the aforementioned elements can be well achieved through immersing learners in authentic texts of corpora.

2.7 Considerations for Corpora Integration in A Classroom.

Though there are several benefits for corpora employing in ELT, there are some caveats that have to be considered in an early stage before actually engaging them in this type of work; the main point is training learners to identify, analyze and categorize linguistic data. Gavioli (2001) explains that training learners to work with corpus data is difficult because *"Unlike dictionaries, grammars and textbooks [concordance data] does not offer explanations; it merely provides raw data which it is up to the user to explain"* (p.

110). It's presumably hard to train nonnative learners as they are less intuitive for making generalizations or recognizing linguistics and grammatical patterns than native learners doing. In addition to making the process harder, a lack of intuition about the target language often leads nonnative speakers to make overgeneralizations about the language. However, nonnative-speaking students were also found to adopt alternative strategies to compensate low intuitions about target language, such as forming and testing hypotheses, since they can't only rely on their intuitions. It has been concluded that with additional guidance from the language instructor, nonnative speakers can substantially benefit from analyzing concordance data of corpora.

2.8 Corpus Based Approach and Vocabulary Instruction

As corpus-based activities, such as DDL, have solid theoretical interpretations supporting their integration in EFL or ESL classrooms and because these activities are deemed as being consistent with modern language teaching approaches, there has been a growing number of studies investigating a wide range of its uses in ELT classrooms. Vocabulary instruction through corpus informed activities have received a great bulk of these studies. Many researchers have investigated the uses and benefits of using DDL or corpus activities for vocabulary instruction in ELT classroom (Cobb, 1997, 1999, 2007; Horst, Cobb, & Nicolae, 2005; Pickard, 1994; Stevens, 1991a, 1991b; Thurstun & Candlin, 1998).

These studies showed how to create corpus informed vocabulary printouts (e.g., fill-in-the-blank or matching exercises) that have the added advantage of being based on authentic texts and also exposing learners to multiple, novel contexts at one time (Stevens, 1991; Thurstun & Candlin, 1998). They investigated the efficacy of complex online self-access vocabulary packages for extensive vocabulary study using concordance data alongside more traditional reference materials (Pickard, 1994; Horst, Cobb, & Nicolae, 2005). In addition to supplying teachers with ideas for creating corpus-based vocabulary activities, researchers have outlined the facilitative effects of using corpus-based materials. For example, Stevens (1991) found that concordance-based vocabulary exercises can be more easily solved by learners than traditional gap-filler exercises, suggesting they should

be used *"if the purpose of the exercise is to reinforce the vocabulary, as opposed to testing, and if the teacher's tendency is to engender a sense of confidence and well-being in the students"* (P. 55). However, Stevens could not make any claims about the efficacy of the activity because he did not empirically test learning outcomes from engaging in DDL. On the other hand, Tom Cobb's (1997, 1999) study has empirically tested the effectiveness of corpus-based techniques to teach vocabulary. Also, Cobb (1997) compared the vocabulary learning outcomes of his students when new words were learned by viewing multiple concordance lines vs. a single sentence accompanied by a short definition of the word, and found that viewing concordance lines lead to small but consistent gains in his students' vocabulary knowledge. Furthermore, in a follow up study, Cobb (1999) found that viewing concordance lines also facilitates the acquisition of transferable word knowledge, supported by the fact that these students were able to apply their knowledge of the word in novel activities and contexts. This empirical evidence suggests that corpus-based activities has an important and meaningful place in the vocabulary teaching curriculum.

However, many studies have compared corpus informed approach to other traditional vocabulary teaching activities and materials in an effort to see which leads to greater learning outcomes and gains. Though doing this proves that corpus-based approach is an effective and beneficial learning tool, it's presumably regarded to some extent as a substitute for all other conventional methods and activities for teaching vocabulary which have been used for decades in ELT classrooms. Consequently, the researcher thinks that instead of promoting DDL or corpus-based activities over other traditional vocabulary teaching methods and activities, it should be considered as a supplementary tool in a varied and comprehensive plan to teach vocabulary and used in aligned with other traditional methods and activities. Therefore, the main goal of the current study was to determine how corpus-based approach can be used to enhance other traditional vocabulary learning methods (i.e., with dictionaries and by guessing the meaning of a word from its context).

In other words, what can DDL teach students about a word that traditional methods either cannot or fail to? A second goal of the study was to determine how best to design offline corpus-based activities in order to support students' analyses and to teach collocations. Since much of the research has either focused on teaching vocabulary with

online concordancing (Cobb, 1997, 1999) or by using concordance lines to change patterns of traditional vocabulary exercises (e.g. gap-fillers) (Stevens, 1991), there is a need for research which seeks to discover how to best design and exploit corpus-based activities in their purest, offline form *(i.e. as raw language data extracted from an appropriate corpus and subsequently given to language learners for analysis).* Furthermore, this study contributes to the growing body of literature which is seeking to discover how the theoretical benefits of corpora are being realized through the collection of students' own accounts of their experiences working with concordance data (Boulton, 2008; Chambers, 2007; Götz & Mukherjee, 2006).

2.8.1 Corpus-informed Teaching Materials

As previously mentioned, that many studies have proved the effectiveness of using corpus-based approach in language teaching. In the same vein, another major benefit of corpora is designing teaching materials and glossaries which considered of great significance for effective language learning as well. This kind of corpus-based materials and books presumably provide authenticity and reality straight into the language classroom. However, it can be noted that some coursebooks claim falsely to incorporate authenticity as Wong (2002) stated:

> *...textbook conversations fail to match findings from empirical studies. The mismatch between textbooks and naturally occurring language has implications for teachers and the writers of teaching material, especially since dialogues of the sort analyzed frequently appear in textbooks marketed as offering authentic natural language, or language which is true to life. (P. 37)*

Cullen & Kuo (2007), in their study, investigated twenty-four textbooks to identify their coverage of spoken grammar; the results were disappointing regarding their representativeness of real used grammar. However, all the analyzed coursebooks were reputable and famous in the British ELT context. In their recent publications Bennett (2010) and Reppen (2010), provided full explanations for uses of corpus data and gave teachers valuable and practical ides and activities for integrating corpora into their classrooms. They both provided full descriptions of using MICASE corpus and COCA corpus, besides other corpora. Another example is Reppen (2010) who proposed *Your Turn*

Activities in his textbook which give students opportunities to communicate with the presented material. Given that many teachers could be anxious about or unsure of the right way to use corpora as a tool in his classroom, these publications are considered of great value and benefit.

2.9 Types of Corpora

There many categorizations of corpora *(general vs. specialized; static vs. monitor; native vs. learner; native vs. translated; monolingual vs. multilingual; parallel)* all of which play an important role in different types of research and they can give priceless data and findings. The most significant type of corpora is the general monitor as it is the most interesting type to a very wide range of linguists and researchers. Recently, most of corpora that can be used by learners and teachers in language learning are available online and easily accessed. One of the prominent examples is the *British National Corpus or BNC* (https://www.english-corpora.org/bnc/), and the *Corpus of Contemporary American English or COCA*, with more than one billion words (https://www.english-corpora.org/coca/). The Corpus of Contemporary American English (or COCA) includes over 1000 million words of text and is evenly divided among all the genres and registers, i.e. spoken, fiction, T.V shows, movies, newspapers, TV shows, and academic texts. It includes about 20 million words annually from 1990 to 2019 and it is being updated at a regular base. You can set comparisons among genres and registers or periods of time.

Regarding BNC, it has about 100 million words which is considered ten times smaller than COCA. It is not as balanced in its constituents as COCA is, as the great bulk of its data taken from written resources, about 90% of the texts, as opposed to only 10% of spoken text. The written part of the BNC is taken from national and regional newspapers, specialist periodicals and journals for all interests and ages, popular fiction and academic books, published and unpublished letters and memoranda, university and school essays, besides other kinds of text. The spoken part consists of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to

radio shows and phone-ins. The *British National Corpus* is not as up-to-date as COCA. Building this corpus took a long time of hardworking of huge teams as the work began in 1991 and completed in 1994. Given all the above-mentioned differences, the researcher has assigned COCA corpus as the tool of corpus-based approach in his study. The following paragraphs will provide much information about reasons for choosing COCA and detailed description of using COCA.

2.9.1 Reasons for Selecting COCA Corpus in this Study

The researcher has decided to employ the Corpus of Contemporary American English (COCA) in this study for its unique attributes that could facilitate learners use and teacher's adaptation in EFL classrooms. It's believed that COCA amidst other available English Corpora is by far the most widely-used whether by researchers, teachers or even learners. Recently, it has undergone beneficial changes regarding its scope, size, friendly interface, additional functions and stats, easiness and speed of searches. It encompasses roughly one billion words for the time period (1990- 2019), from various of genres equally represented in the available data. The following table has been adapted from the website of COCA to illustrate the distribution of included genres which substantially indicates COCA as a large, recent and varied corpus. In march 2020 COCA has added three new genres to its massive data; they are T.V shows and movies subtitles and blogs and web texts which represent available addition to the informal available data. The following table illustrates these genres in details.

Table 2.1: Genres Distribution in COCA Corpus.

Genre	# texts	# words	Explanation
Spoken	44,803	127,396,932	Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Oprah)
Fiction	25,992	119,505,305	Short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books 1990-present, and fan fiction.
Magazines	86,292	127,352,030	Nearly 100 different magazines, with a good mix between specific domains like news, health, home and gardening, women, financial, religion, sports, etc.

Newspapers	90,243	122,958,016	Newspapers from across the US, including: USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle, etc. Good mix between different sections of the newspaper, such as local news, opinion, sports, financial, etc.
Academic	26,137	120,988,361	More than 200 different peer-reviewed journals. These cover the full range of academic disciplines, with a good balance among education, social sciences, history, humanities, law, medicine, philosophy/religion, science/technology, and business
Web (Genl)	88,989	129,899,427	Classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages (by Serge Sharoff). Taken from the US portion of the GloWbE corpus.
Web (Blog)	98,748	125,496,216	Texts that were classified by Google as being blogs. Further classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages. Taken from the US portion of the GloWbE corpus.
TV/Movies	23,975	129,293,467	Subtitles from OpenSubtitles.org, and later the TV and Movies corpora. Studies have shown that the language from these shows and movies is even more colloquial / core than the data in actual "spoken corpora".
	485,179	**1,002,889,754**	

Note: This table has been adapted from the official website of COCA Corpus in March,2020, at https://www.english-corpora.org/coca/

The balance between the involved eight genres make it quite easy to get tangible information about the frequency of the target words or phrases or grammatical strings across all genres, regardless of the register of occurrence. It can be very formal like academic articles and lectures or very informal like T.V. scripts or movies subtitles or somewhere between like newspapers and magazines. To illustrate, the following charts demonstrate the frequency of the word *plan* across all genres, as well as the verb" *donate*".

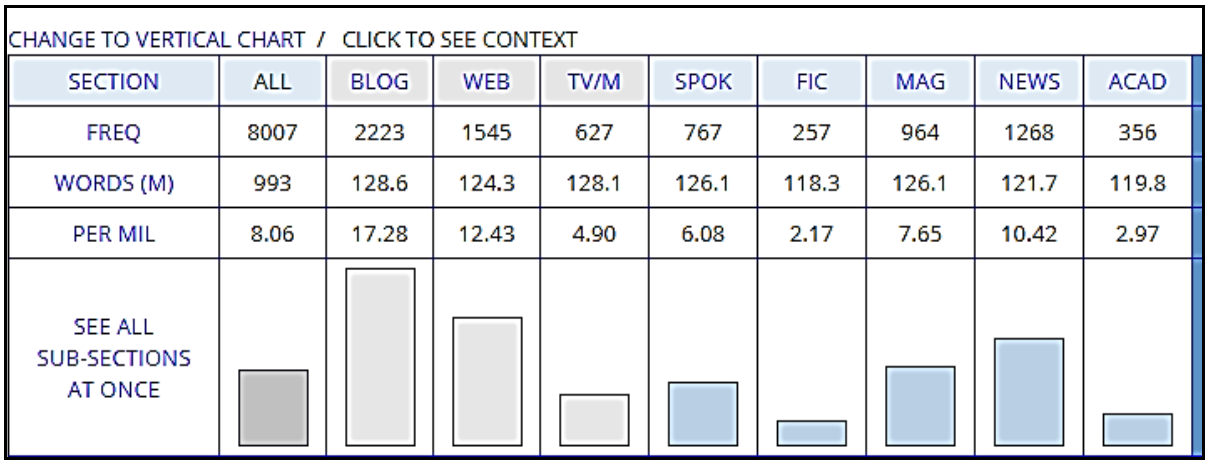


Figure 2.1: Frequency of the verb "*denote*"

CHANGE TO VERTICAL CHART / CLICK TO SEE CONTEXT

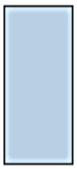
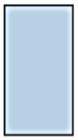
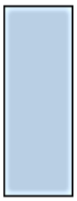
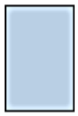
SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	159633	20701	22591	18113	24818	8116	20488	29011	15795
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	160.75	160.96	181.81	141.43	196.76	68.59	162.49	238.30	131.86
SEE ALL SUB-SECTIONS AT ONCE									

Figure 2.2: Frequency of the noun "*plan*".

This intensity and balance of data can be exploited in investigating also the frequency of related phrases or words, i.e. collocates or clusters. For example, the following figure illustrates clusters or collocates of "*nutritious+ NOUN"*.

CLUSTERS	NUTRITIOUS
228	nutritious food
126	nutritious foods
89	nutritious meals
65	nutritious diet
48	nutritious meal
37	nutritious breakfast
20	nutritious snacks
17	nutritious for
14	nutritious eating
14	nutritious lunch

Figure 2.3: Clusters for Adj "Nutritious + Noun"

We can also compare two genres (or sets of genres) to identify frequent collocates of the target word. For instance, there has been clear differences in frequency of the collocates (nearby words) of *chair* in across the two genres *academic* and *fiction*. Furthermore, it can be used to compare synonyms of a given word in different genres. In addition to comparing across genres, you can also compare words, to "tease out" differences between related words. For example, the words that occur with *deep* and

profound, showing that (for example) *deep breath* is common but *profound breath* never occurs, or that *profound effect* is common but *deep effec*t is not. You can also compare other related words, such as adjectives used with the words *men* and *women*, or verbs occurring near *Obama* or *Trump*. Just as we compared collocates (nearby words) in different genres above (for *chair* in ACAD and FIC), we can also compare collocates in different periods, to look at semantic change. For example, searches results show that collocates of *green* in the 1990s and the 2010s are quite different in their frequency and popularity.

Another significant attribute is that the attention it gives to the **top 60,000 words** in the corpus, and the wide range of information for each word, including frequency information, definitions, synonyms, WordNet entries, related topics, concordances (new display in COCA), clusters, websites that have the word as a "keyword", and KWIC/concordance lines. A unique feature of COCA, which makes it very useful for language learners and teachers, is the ability to ***browse*** through a list of the ***top 60,000 words*** (lemmas) in the corpus, and then to see an extremely wide range of information on each of these words (See Figure 2.4).

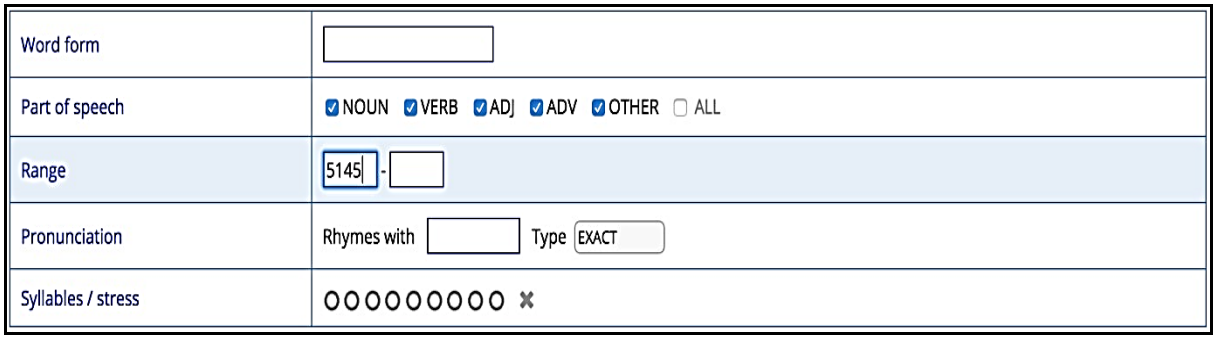


Figure 2.4: Search Box in the 60000 Top Words (The most Frequent words)

For example, the following are just a few examples of high frequency words (about word #5000 in the 60,000-word list), medium frequency (~25,000), and low frequency (~45,000) words. For each word in the list, users can hear the word pronounced, see videos with that word in the text, find related images from Google Images, an see a translation for their preferred target language.

	RANK	FREQ	Word	PoS	Audio	Video	Image	ZH-CN
1	5200	11377	blogger	NOUN				
2	5201	11374	utterly	ADV				
3	5202	11372	trouble	VERB				
4	5203	11368	texture	NOUN				
5	5204	11365	head	ADJ				

	RANK	FREQ	Word	PoS	Audio	Video	Image	ZH-CN
1	25201	577	unappealing	ADJ				
2	25202	577	tryst	NOUN				
3	25203	577	urologist	NOUN				
4	25204	577	scrabble	NOUN				
5	25205	577	demarcate	VERB				

	RANK	FREQ	Word	PoS	Audio	Video	Image	ZH-CN
1	45107	112	incubus	NOUN				
2	45108	112	unquantifiable	ADJ				
3	45110	112	husk	VERB				
4	45111	112	hunky-dory	ADJ				
5	45112	112	jacked	ADJ				

Given the proven results of both quantitative and qualitative studies on corpus-based vocabulary instruction in EFL or ESL classrooms, it is predominantly clear that employing authentic adequate context for learning target words dramatically increases retention rate of new learnt vocabulary as EFL learners study new English vocabulary autonomously, their attention maintains at an optimal rate. For Example, in his quantitative research Davies, R. (2014) concluded that essential benefits can be gained through employing computer programs in vocabulary learning classrooms equipped with meaningful, adequate, and appropriate data. In another study, Ma (2013) has reported that using a computer program, that provide learners with much practice for the target vocabulary in various contexts, has proved far high level of learners' retention rate. Another clear example is computer-mediated English dictionaries for text processing have proven to be helpful for improving vocabulary learners' receptive vocabulary knowledge (Li, 2010). For example, an individualized vocabulary building program called *The First 4000 words* (Seward, 2012) that provides read-along activities with voice-recognition feedback, underlined reading sentences, and repeatable listening to sentences was tested,

and significantly improved vocabulary gains were reported for the experimental groups (Fehr et al., 2012). Positive outcomes such as significant vocabulary gains, improved reading comprehension, and fast word recognition were confirmed through a direct CALL learning program for frequent vocabulary (Tozcu & Coady, 2004).

On the other hand, another significant factor has to considered in handling corpus-based approach in vocabulary learning and retention which is viewed as a substantial element of corpus integration, is target words rehearsal or occurrences in various authentic contexts. The learner would frequently encounter the same word. This has been proved by many researchers (Hulstijn, 2007; Nation, 2001; Stuart Webb, 2007), to positively influence learner's receptive and productive vocabulary use and retention. Positive effects of rehearsal chances on second language vocabulary learning have been investigated (Barcroft, 2007), noted that about 10% improvement in vocabulary learning for learners exposed to retrieval-oriented and control conditions. Rehearsal opportunity for vocabulary learning showed that L2 learners' improved retention (Slamecka & Graf, 1978). The positive impact on learning of encountering new words in three consecutive sentences rather than one single sentence has been found in an experiment with 32 Iranian students (Baleghizadeh & Shahry, 2011). However, just repeating a new word aloud several times was found to be insignificant in the acquisition of new words (Abbs, Gupta, & Khetarpal, 2008). The effect of three repeated exposures to new words, and different presentation factors in encountering new words were also investigated by Laufer and Rozovski-Roitblat (2011), and their findings reveal that effective presentation benefits learners in retaining new words better than just the frequency of meeting new words. The further effects of repeated word occurrence (i.e., 1, 3, 7, 10 occurrences) on vocabulary learning suggest that more learning gains can be expected as repetition numbers increase (Stuart Webb, 2007).

2.10.1 Corpora Integration in Vocabulary Classroom

The lexical competence of learners has been proved to be important to second language learners (Carter& McCarthy, 1988. P.97) and at the same time learners believe that it is a big a challenge for mastering a language (Coady, 1997; Cobb, 1999). Recently, the development of technology has motivated a new vision for language teaching through

integrating computers as effective instruments in teaching classrooms, thus facilitating the learning process.

Different lexical materials have been made available online through various online corpora equipped with some tools for categorization and navigation through corpus data. The utilization of these tools has proven to be helpful to ESL teachers, especially for teaching and testing vocabulary, rather than using dictionaries. The following paragraphs will show potentials of corpora integration in vocabulary instruction.

2.10.2 Providing Speedy Searching Tool

Using concordance method works as an effective speedy searching tool. In traditional learning, dictionaries are used as the main resource to find the meanings of words; however, this task is often too tedious and time-consuming (Cobb, 2003). through employing corpora query types to search for word contexts, students are engaged in far more fast and rich language learning context.

2.10.3 Providing Authentic and Meaningful Language Samples

Cobb (2003) has suggested that language learning takes place when adequate examples are noticed and processed by learners; for example, when learning a word, learners need to find more examples or definitions or contexts in order to fully grasp the assigned lexical items. Conventional glossaries are not sufficient for fully understanding unfamiliar words. Using a concordance to search for word examples enables learners to find more examples including word clusters, collocations, contexts, and frequency, rather than simple and prefabricated examples. So, Learners can develop the breadth and depth of their lexical knowledge.

2.10.4 Encouraging Student-Center and Independent Learning

Conventional approaches for vocabulary learning are considered to be a *'passive way of learning';* regardless of the adopted approach in the classroom, explicit or implicit learning, the learning experience remains inactive and teacher-centered. As students receive lists of words or fabricated reading passages, then they are asked to look up word

definitions, practice and memorize these words, and likely they remember them or oftentimes they get depressed as they forget the words and go through the same cycle endlessly.

On the other hand, through designing and using activities that get learners engaged in noticing and exploring the target language, teachers can engage learners in a *'content decision making'* learning situation (Hadley, 2001). This technique is called *Data-Driven Language Learning.* It is suggested by Krishnamurthy (2004) that a word often has many meanings; however, the actual meaning should be determined by its surrounding contexts. By using the data-driven technique, learners can choose, explore and determine the language from various resources that the computer found. The classroom therefore becomes student-centered and learners have active control of their own learning (Nation, 2001). In the following paragraphs, the researcher will show some suggested activities and tasks for integrating corpora in a vocabulary learning classroom.

2.10.5 Creating Concordance Sheets for Young Learners or Beginners

The main problem is that young learners do not know more about computers. Accordingly, they find themselves frustrated because of the massive obtained data from query results (Hadley, 2001). However, teachers can help students to meet this challenge and overcome its distorting effects through creating a concordance sheet which have more relevant and suitable data for learners.

After distributing the concordance sheet, teachers ask their students to pick some examples and keep them in their language diaries. Through this activity, learners can be engaged in noticing and exploring the different language contexts; during the process of writing down selected examples, students are supposed to go through a cognitive process of digesting and grasping the language input. Hadley's study (2001) proves that students' writing skills and lexical knowledge have been proved by using this teaching skill.

2.10.6 Classroom Project-10 Words a Week

In this activity, Teachers could engage learners in a classroom project. They can pick some fun stories for students as homework; they could ask them to use highlighters to list unfamiliar lexical items on vocabulary cards during the reading process. Teachers have to prepare two boxes inside the classroom; one of them for unknown words and the other for learnt ones. Students can place their colorful word cards in an unknown word box at any time they come to the class, and teachers can choose at least ten words from an unknown word box as a weekly learning project in which students are asked to look for examples from corpora and share the queries obtained results with the whole learning.

2.10.7 Task-based Learning-Filling Gaps

One of the most common activities and pair work tasks is *Filling Gaps* exercises that could be designed by the teacher through adapting concordances lines for target words. When learners are engaged in such tasks or activities, teachers can monitor their learning progress and have control over the class. Teachers can use the authentic contexts searched from corpora to compose a gap-filling sheet for students to work on; for example, if the target word for the class is 'medicine', teachers may want to present several collocations such like use medicine, take medicine, and prescribed medicine. The gap-filling task may look like this:

> • If you wear contact lenses, ask your health care provider if you should wear them while you are _____ the eye medicine. Also ask if the medicine you have been _____ (prescribe) will stain your contact lenses.
>
> • If your doctor _____ a new antibiotic, take all of the medicine as prescribed.
>
> • Many antihistamines cause drowsiness, so you may want to _____ the medicine only at bedtime.

(Search results of 'medicine' from COCA online Corpus: Available at: https://www.english-corpora.org/coca/)

By encouraging learners to search for answers for the gaps through using online corpora which could form interesting meaningful learning experience (Bracewell & Laferriere, 1996). Furthermore, the task-based learning could also improve students trust in their abilities once they achieve the assigned tasks.

2.10.8 Developing Study Materials

Similarly, corpora could help in improving learners' analysis, noticing, inferring, and generalizing skills and enhance their linguistic skills including the knowledge of synonyms, polysemy, homonymy, antonyms, etc. As Barlow (1996) illustrated:

> *"Corpora can reveal not only the range of patterns of a language that the learner must assimilate, but also their frequency, which is an important factor in materials development and syllabus design. Moreover, the use of corpora has the potential to radically alter the field of second language learning, and perhaps Linguistics as well". (P. 2)*

As it has been illustrated in Barlow's (1996) quotation that ELT coursebooks OR textbooks which are mainly corpus-informed provide far more authentic and real-life examples and exercises than those conventional textbooks that offer prefabricated and made up examples. Likewise, Dash (2008) supported this viewpoint by ascertaining that *"corpus-based teaching materials are reliable and authentic because they give attention to the common choices of usage than those that are less common".* (P. 119)

2.11 Summary and Conclusion

A number of corpus linguists (Hunston, 2002; Meunier, 2002; Tribble, 2001) has stated that corpora have great effect in facilitating language learning and teaching and the use of corpora is also in consistent with the common beliefs in pedagogy that discovery learning and problem-solving would encourage students to be active autonomous learners. However, as previously mentioned, there are some potential challenges in using corpus-based approach that have to be taken into consideration; otherwise the potentials of corpora use could be less effective. Thus, it was proposed that corpora have to be pedagogically mediated by meeting some possible challenges in students' proficiency levels, low technological skills, students' anxiety of trying new things, and dealing with massive data of concordances.

I addition, it has to be taken into consideration that corpus-based activities should not wholly replace other conventional tools, rather it should be used alongside these conventional tools to get the maximum benefits (Meunier, 2002). Another significant point is that blueprints of using corpora must not be technology-driven, but pedagogically driven.

In other words, we should not use corpora just because it is a modern trendy tool available and easily accessed, but its use have to be *"vindicated to the extent that it agrees with what we know about language and language acquisition, and can be shown to be an effective learning tool"* (Johansson, 2009, p. 42). Consequently, there is a severe need for experimental classroom-based action studies conducted by professional and teachers who are aware of the potentials as well as the limitations of corpus-based approach (Seidlhofer, 2002). This type of action research could figure out more underlying issues on corpora use and could test its effectiveness in ELT context.

Some researchers and ELT professionals argue for potential challenges and difficulties for incorporating corpora in language teaching for a variety of reasons. Some believed that it is technically challenging and time-consuming while classroom time and even technological resources could be limited. Other argued that it is confusing as learners would be encountered with a massive amounts of concordance lines which could be frustrating and boring. Another raised issue is some types of corpora and corpus tools or software are not free and need paid subscription. However, researchers believed that all these raised issues could be properly handled through wise use of corpus-based approach which requires gradual application, well-trained teachers, learners training on technical and concordancing skills, teacher's constant feedback and support, using corpus-based printouts, and finally combine other required conventional tools alongside with corpora to fully achieve the assigned objectives.

In summary, the convenient and principled use of corpus-based approach should combine both technical skills and pedagogical skills, with more focus on the pedagogical skills. The pedagogical skills include all required skills for teachers and learners to deal with corpus obtained data, corpus-informed activities designing and doing tasks, and so on. Thus, it can be concluded that designing evidenced-based blueprints for applying corpus-based approaches with different ages or proficiency levels in vocabulary instruction classroom is significant and of great importance for getting the benefits of corpora.

PART TWO: PREVIOUS STUDIES

2.12 Introduction

In the first part of this chapter, the researcher has showed significant theories, approaches and discussions related to corpora use in language teaching. Generally speaking, there is an agreement among most of scholars and researcher on the effectiveness of corpus-based approach on vocabulary gains and retention rate for EFL or ESL learners. In addition, it is believed that this approach of corpus use is wholly consistent to modern trends in teaching and pedagogy which call for learners' autonomy, learner-centered classroom, discovery learning tasks, problem-solving activities, and authenticity and meaningfulness of input. In the second part of this chapter, the researcher will show previous studies at three levels, international context, Arabi context, and Saudi context. The end of this part, the researcher will show an overall view on results of the previous studies and the gap that he will investigate at the Saudi EFL context.

2.12.1 Studies Conducted on the International Context

The researchers Boontam and Phoocharoensil (2018) has investigated the effectiveness of using Data Driven Learning on learning three prepositions (during, between, and among). They also investigated learners' attitudes towards learning through DDL approach. The participants of the study were 30 Thai young learners at the 4th grade at a school in Nakhonsawan Province, Thailand. The study took 6 weeks, one class of 50-minutes per week. The researchers used pre and post-tests to measure the effectiveness of DDL approach on vocabulary gains. The pre and post-tests were identical with some changes in items order, and consisted of 15 gap-filling items and 5 sentence building items. To investigate learners' attitudes, the researchers used questionnaires and interviews. At the end of the experiment, the researchers statically analyzed students pre and posttests scores to measure vocabulary gains. They also statistically analyzed questionnaires responses and qualitatively analyzed their responses to the interview's questions. Results showed a noticeable improvement in learners' vocabulary learning after going through a six-week

treatment using DDL approach. In addition, learners showed positive attitudes towards the utilized activities and tasks of DDL approach and they described it as helpful, fun and challenging.

Another experimental study done by Çelik and Elkatmiş (2013) which aimed to compare the effectiveness of using corpus-based approach to the effectiveness of using lecture-based approach to learn Turkish punctuation marks. The sample of the study consisted of 171 university learners studying at Primary Education Department at University of Kirikkale, Turkey. The target knowledge that assigned for learning and testing was comma, full stop, colon, semi colon, triple exclamation mark, quotation mark, dot, hyphen, and apostrophe. To collect data, the researcher utilized achievement tests (pre and posttests) to measure their punctuation knowledge, a questionnaire and focal interviews to investigate learners' attitudes towards corpus-assisted approach. Posttests and pretests scores for the assigned four groups have been analyzed using descriptive analysis and using T-tests to determine the effect size of corpus-assisted approach compared to conventional lecture-based approach, while learners' responses to questionnaires were analyzed using descriptive statistics to investigate their attitudes towards corpus-assisted approach. Analysis of pre and post-tests scores showed that experimental groups overperformed their counterparts, the control ones in Turkish punctuation skills. In addition, the analysis of learners' responses to the questionnaire and the interview questions revealed that learners had positive attitudes towards corpus-assisted approach. They also showed their desire to use this approach in the future as teachers.

Cotos (2014) tried to enhance ESL graduate students' skills of linking adverbials in an advanced academic writing course. The researcher used experimental design in which he applied pre and post-tests to measure the effect size of using two types of DDL activities based on two types of corpora, native speaker corpus and learner corpus. Posttests and pretests scores for the assigned groups have been analyzed using descriptive analysis and using T-tests to determine the effect size of the two types of DDL activities. The results showed significant improvement in learners using of linking adverbials in their writing tasks. Their use of adverbials was noticed to be more accurate, varied and frequent. The researcher concluded that the more significant improvement was observed in the written

production of the students who worked on the corpus that their own writings were included in.

The prominent linguist Mukherjee (2004) conducted a study on teachers' perceptions on and awareness of corpus linguistics and its relevance to their teaching and pedagogy. The sample of the study was 248 qualified English teachers at secondary schools in North Rhine-Westphalia in Germany in the time period (2001-2002). He was holding workshops at secondary schools for English teachers aiming at three things: first, investigating teachers knowledge about corpus linguistics and its relevance to their teaching profession; second, intruding basic principles and terms of corpus linguistics; third, introducing corpus linguistics pedagogical implications and applications for classroom practices to get teachers familiarized with corpus-based approach and tools. participants teachers were asked to answer some questions before the commencing of the workshop to investigate their knowledge and perceptions on corpus linguistics. During the workshops, they were introduced to the concepts and principles of corpus linguistics and trained on using corpus tools, designing corpus-based activities and tasks' they also trained on the pedagogical implications of corpus-based approach such as the representativeness of generated data, the uses of word-lists and concordances, the applications of corpus data in classrooms. At the end of the workshop, they were asked to answer some questions about their perceptions on corpus linguistics. Mukherjee (2004) concluded that before the workshops teachers showed very low awareness of corpus linguistics concepts and its applications in their English language classroom, while after attending the workshops, about 95% of them were well aware of the potentials and uses of corpus linguistics that could be gained from the utilization of corpus-based activities and tasks.

In another study Levchenko (2017) also investigated learners' perceptions on corpus-based activities in learning English for Academic purposes. He conducted his study on 10 doctoral students (60 males and 4 females) studying at different departments at Samara University, Russia. The study aimed to improve their academic writing competencies through using corpus-based activities. The study took 15 weeks in which students attended two hours weekly for academic writing classes. The researcher utilized both corpus-based printouts and online corpora activities. To collect the data on their

perceptions on corpus-based approach, the researcher used three surveys at three different stages in the treatment; at the beginning, in the middle, and at the end of the experiment. To analyze the obtained data, both qualitative and quantitative methods were used. The results revealed that learners perceived corpus-based activities as helpful and effective tools for developing academic writing skills. In addition, they believed that corpus-based activities enhanced their confidence towards the tasks of academic writing. They also stated that they could realize the high-frequency academic words that they should incorporate in their future writings.

In the same vein, MacArthur and Littlemore (2008) investigated the effect of using corpus data on facilitating learners' working out the meanings of the peripheral senses of denominal verbs in English and Spanish. The researcher used experimental design in which he applied pre and post-tests to measure the effect size of using corpus data on learners' knowledge of the peripheral senses of denominal verbs in English and Spanish. Posttests and pretests scores for the assigned groups have been analyzed using descriptive analysis and using T-tests to determine the effect size of the use of corpus data on peripheral senses of denominal verbs. The results showed that they were able to work out not only the basic senses of denominal verbs but also the different senses of these verbs from the corpus obtained data.

Chang and Sun (2009) investigated the effect size of using web-concordancers and scaffolding on learners' proofreading skills. The researcher employed a quantitative method in his study. The sample of the study consisted of 26 second-year senior high school EFL students at a private school in central Taiwan. They were placed into two groups, experimental (using web-concordancers and scaffolding tools) and control (using conventional tools). the researcher used pre and post-tests; pre-test was used to prove the comparability of the assigned groups, and posttest to measure learning gains after going through two different teaching approaches. Tests scores were analyzed quantitatively using descriptive statistics analysis and T-tests. The results revealed that students had positive attitudes towards the effect of corpus-based and scaffolding activities. Likewise, learners proofreading skills with concordance support demonstrated significant improvement ($t=-3.65$, $p<0.01$. on the other hand, students expressed their desire to receive feedback and

support during corpus consultation stage, especially in early stages of corpora use. This result indicated that a sort of guidance is required during students' corpora use in activities and tasks, especially at early stages of corpora consultation.

Varley (2008) investigated the effectiveness of using direct corpus consultation alongside the conventional dictionaries or grammar text books in EL context. He conducted a qualitative design study to answer the study's questions. The sample of the study consisted of 19 advanced ESL students (IELTS 6 and above) who were studying B.A in a public institution in New Zealand. The main task that participants asked to do was to carry out an individual project in which they had to choose a genre of spoken or written English and use a Concordancer (Wordsmith Tool) to investigate two to four lexical or grammatical items. Data collection tools were students' reflective logs in which they were asked to write down their perceptions and notes while working on their projects. The researcher qualitatively analyzed students' reflective blogs to figure out their perceptions on corpus consultation method. The findings illustrated that students had positive attitudes towards corpus use, especially in the areas of learning vocabulary and syntactic pattern awareness. Some students stated, in their reflective logs, that they were able to access collocation and cluster information that conventional dictionaries could not offer.

Todd (2001) conducted a study on 25 Thai students at university level. This study aimed at examining students' induction ability as it arose from a concordancing activity. First, the teacher identified lexical items which had a strong potential to cause errors in writing. The students were asked to make small concordances of the lexical errors; they then applied patterns from the concordance in the larger process of completing the self-correction of their own errors. The results showed that learners acquire valid patterns from their self-created concordances, also they became aware of self -correction to their own errors. This study proved the interrelationship between induction and concordancing skills. The researcher concluded that concordancing and corpus technology could play a role in helping Thai students to learn English.

2.12.2 Studies Conducted on the Arab World Context

One of the most relevant studies that conducted in the Arab context, specifically in Egypt, was Elsherbini, S., and Ali, A. (2017). It was a quasi-experimental study investigated the impact of corpus-based activities on enhancing EFL students' grammar and vocabulary and explored their attitudes toward using corpus. The researcher assigned two classes of 104 freshmen taking business English course at Sadat Academy for Management Sciences; with 54 in the experimental group and 50 in the control group. The mixed-methods design was used to gather data from a grammar test, a vocabulary test, a questionnaire, and semi-structured interviews. Students in the experimental group were trained to use corpus and were taught through corpus-based activities for 11 weeks; whereas control group students were instructed using the course book only. The differences in students' performance in grammar and vocabulary were examined between their pre- and post- tests scores. The results of the t-test indicated that there was a significant difference between the mean scores of the pre- and post-tests of grammar and vocabulary of the experimental group in favor of the posttests. Also, experimental group's students outperformed their counterparts in the control group in the mean scores of the grammar and vocabulary posttests. After the treatment, the attitudes of experimental group students toward corpus were explored through administering a questionnaire and semi-structured interviews. Data analysis of the questionnaire revealed that learners' attitudes were generally positive towards using corpus-based activities in learning both grammar and vocabulary. Analysis of the interviewees' responses supported the findings of the questionnaire and offered some insights into using the corpus.

In collaboration with Noor and Amir, Al-Mahbashi (2015) conducted a study on 60 Yemeni female students from the second level in the English Language Department, Faculty of Education, Sana'a University in the time period (2014-2015). The study aimed to investigate the effect of the Data Driven Learning (corpus-based approach) over the conventional tools like dictionaries on improving receptive knowledge of lexical items and collocations. The participants were randomly assigned into two groups: a control group, called (Dictionary Group) and an experimental group, called (DDL Group). the researchers

assigned COCA corpus to be the main source for the designing activities, tasks and assignments for the experimental group, while the control group was using conventional dictionaries for learning lexical items alongside the textbook. Due to the lack of computers and internet connection at the department, the researcher employed corpus-based printouts, rather online corpora use. The researchers employed three achievement tests, pre-test, post-test, and delayed post-test, to measure the effect of using DDL activities compared to conventional dictionaries on learning receptive knowledge of words and collocations. The three achievement vocabulary tests were identical except for the order of test items was different in each test. After applying the three tests, the researchers analyzed the scores of pre, post, and delayed post-tests to measure the effect of using corpus-based approach on vocabulary receptive knowledge improvement and retention rate. The results revealed that the experimental group outperformed the control one in collocational and lexical receptive knowledge in both posttests and delayed posttests which indicated higher vocabulary gains and retention rates in favor of the corpus-based approach. The researchers concluded by stating that "*The long and short-term impact of DDL method could be attributed to learning the target words in rich contexts through numerous exposures to them that would subsequently lead to vocabulary acquisition through what is called discovery learning".*

2.12.3 Studies Conducted on the Saudi EFL Context

One of the most recent studies that investigated the use of corpora in Saudi EFL classroom for learning verb-noun collocations was conducted by Alruwaili (2020). The target items were selected through a multi-level approach which consisted of three levels: a corpus-based approach, a phraseological approach and a pedagogical approach. The main aim for this study was investigate the effect of training learners, on corpora activities to learn verb-noun collocations, on enhancing the application of corpus-based approach in EFL context. The sample of the study consisted of 51 intermediate-level students studying general English at Foundation Year, at a Saudi university. The study took 5 weeks, two classes weekly one hour for each. She had three sessions of training for the participants on using corpus resource (ANTCONC), and on concordancing skills and technical skills. She also employed a tracking software (The All-In-One Keylogger) on all students' PCs which

enable the he to track all the students' textual logs and screenshots. For data collection, she used tracking logs, activity sheets, reflective forms and interviews. Therefore, the study adopted a mixed approach design to collect data. All the obtained data from activity logs, activities sheets, reflection forms and interviews, were analyzed using quantitatively and qualitatively. The results revealed that learners were able to apply DDL approach the same way as they were trained on which indicated that successful training for learners is essential for successful corpus-based approach application in EFL classroom. The researcher also concluded that over time and with practice the students spent less time in activities. In addition, it showed that learners had positive attitudes towards using corpora in EFL classroom. In conclusion, this study stressed on the significance of learners' corpus training, which includes both technical and concordancing skills (noticing, determining, analyzing, interpreting, generalizing), for effective application of corpus-based or DDL approach in EFL context. It also provided an empirical application for tracking learners' concordance activities.

Another study by the same researcher Alruwaili (2018) investigated teachers' attitude towards corpus-based approach after attending a workshop on the use of corpora in EFL classroom. Sample of the study consisted of 56 in-service Saudi teachers. To collect the data, the researcher used two questionnaires (pre-course and post-course) and interviews. Obtained data from questionnaires responses and answers to the interviews were quantitatively and qualitatively analyzed. The results revealed that teachers had moderately positive attitudes towards the use of corpus-based approach in their classrooms. Furthermore, the findings illustrated that there were some factors that could influence teachers attitudes towards the use of corpora such as the training course, computer and technical skills, their perceptions of their roles and the learners' roles in communicative approach. The researcher concluded that corpora integrating in the Saudi EFL context is possible after meeting the potential challenges that may hinder or distort the application of corpus-based approach.

Another study by Alqarni, A. (2019) that aimed to investigate the common formulaic patterns in EFL textbooks specifically used to enhance academic English for many Saudi university students in their Foundation Years through utilizing corpus-based

approach. In his study, Alqarni intended to identify the most frequently occurring 3-, 4-, and 5-word formulas in the textbook whose pragmatic functions were specified on the basis of Simpson-Vlach and Ellis' taxonomy (2010). Furthermore, the corpus-based approach was used to compare the extent to which the Academic Formulas List (AFL) (Simpson-Vlach & Ellis, 2010) appeared in the textbook, and the number of times AFL formulas occurred. The main findings of the study showed that the textbook included a total of 342 formulas: 217 formulas; 101 formulas; and 24 formulas for the 3-, 4-, and 5-word formulas, respectively. In addition, the textbook incorporated only 143 formulas out of 400 AFL formulas. Finally, the majority of the identified formulas functioned as constructing stance expression (46%), followed by referential expressions (36%), and lastly as discourse organizing expressions (18%). these results could enhance understanding of the provided input in an ELT textbook, and to raise the awareness of the significance of formulas inclusion in ELT textbooks. The findings could therefore help L2 researchers, L2 instructors, and also ELT publishers to not only incorporate academic formulas into ELT textbooks but also to integrate formulas with high frequency for L2 learners. The researcher concluded that L2 formulaic patterns or collocations should not be randomly included, but in a way that is consistent with scholarly criteria.

In her study, Alharbi. R. (2017) examined the effect of three Focus-on-Forms instructional approaches on the passive and active acquisition of non-congruent collocations: 1) the non-corpus-assisted contrastive analysis and translation (CAT) approach, 2) the corpus-assisted (CAT) approach, and 3) the corpus-assisted non-contrastive analysis and translation approach. To thoroughly investigate the proposed combined condition (i.e. the corpus-assisted CAT) and its learning outcomes, a control group received no treatment to be a baseline for the experimental groups. The target lexical items for the study was 30 collocations non-congruent with the learners' first language (Arabic Language). The sample of the study was 129 undergraduate university learners in a Saudi university. The participants were randomly assigned to the experimental three groups and the control group. The three experimental groups received three different instructional approaches; the first experimental group (Corpus-assisted CAT) were performing tasks and activities using Arabic/English Corpus data, the second one (non-corpus-assisted CAT)

were performing tasks and activities based on texts using contrasted analysis and Arabic translation from the teacher for the target collocations, the third one (corpus-assisted non CAT) were performing gap-filling and multiple choices tasks using monolingual corpus data. To collect data, the researcher used passive and active recall tests immediately after the treatment and three weeks later to measure retention rates. Data obtained from tests was statistically analyzed by using t-tests and descriptive statistics analysis to measure the gains for each group. The results revealed that the first experimental group (corpus-assisted CAT) outperformed the other three groups in all tests (passive recall, active recall, immediate and delayed). These results were interpreted by the researcher in the light of the hypotheses of *'noticing'*, *'involvement load'* and *'pushed output'*.

A relatively pertinent Saudi study, Almutairi, N. (2016), has focused on the using specialized compiled corpora for teaching Personal Statement Letters writing as a type of academic writing for university students. The main aim of the study was to investigate the effectiveness of using a compiled specialized corpus on teaching personal statement writing for university students in Saudi universities. He chose to compile a specialized corpus from personal statement letters of law students, and employed a concordance called Sketch Engine to compile a corpus of 67 letters and a total word count of 50691 words. Then, he used the concordancer to analyze the compiled data and determine the common lexico-grammatical features in writing a personal statement letter. The obtained results from this analysis were used in designing exercises, activities, and tasks to be taught for university students in academic writing classes. The lexical findings, frequency and tag lists showed that personal nouns, adjectives, and nouns were frequently used and had significant uses. In addition, they revealed that using personal pronouns in this type of writing is dominantly preferable as it served self-promotion purposes. Grammatical findings, there were a dominance for the use of past tense in listing qualifications, and the use of passive voice in narrating things. Lastly, she concluded that these results could help teachers design various classroom activities to equip students with the appropriate academic vocabulary and grammatical features that suit personal statement genre.

One of the significant projects that can be used as a guidebook and handouts for teachers and ELT professional for integrating corpus-based activities in classroom, is

Aljuhani's, H. (2016). She tried to offer designed corpus-based models, lesson plans and instructional procedures for integrating corpora in a lexicalized stylistics approach to teaching English as a foreign language literature. This project has provided detailed guidelines and samples about how to apply the Lexical Stylistics Corpus-based Approach can improve EFL learners' linguistic skills. The proposed Concordancers combined with poetic and lexical stylistics can provide a wide range of literary tools accessible for EFL learners. Furthermore, it offered these tools, along with detailed ready-made sample lesson plans, to EFL teachers and learners.

2.12.4 Conclusion and Summary for Previous Studies

The aforementioned reviewed studies have shown the effectiveness of the use of corpus-based approach in language teaching and learning, especially vocabulary learning, with some concerns about the challenges and difficulties that have to be met while using this approach. They also revealed the recency and vitality of this discipline as a whole, whereas there are common aspects can be summarized in the following points: First, there is a consensus clearly stated in most studies on the importance and effectiveness of corpus linguistics as a recent rising discipline in language on language analysis, i.e. studying linguistic features for various genres of writing or speech. Second, it's notable that the surging advancements in technology have motivated studying and utilizing corpora in different aspects of language study and language teaching. Third, there are relatively low relevant studies on corpora generally, and specifically using corpora in EFL classrooms in Saudi Arabia, even in most Arabic countries. This necessitate the urgent need for investigating corpora's efficacy in EFL Arabic contexts. Fourth, methodologies used in most studies varied from quantitative methods like teachers' questionnaires and learners' questionnaires, or qualitative like professional interviews, reflections sheets, and activity logs. Fifth, few attempts were made to investigate its effectiveness on both learning and retention of vocabulary knowledge in EFL classroom, and none was made on Saudi EFL context.

At this juncture, there are similar studies conducted in other contexts, but the assigned variables, methods employed, application context and population are different. For

illustration, Alruwaili (2018) investigated the effect of training learners on corpora skills on the application of corpus-based approach, while Almutairi (2016) investigated the effectiveness of using specialized corpus in teaching academic writing. Consequently, the researcher believed that none of these studies, conducted on the Saudi context, adopted a classroom-based action research to investigate the effect size of using corpus-based approach on vocabulary receptive and productive knowledge and retention rates in Saudi EFL classroom. In addition, none of them empirically investigated the potential challenges in using corpus-based approach in ELT for both teachers and learners.

To fill in this gap, the current study utilized experimental classroom-based research to investigate the effectiveness of using corpus linguistics in vocabulary learning and retention in Saudi EFL context at tertiary levels, and examine students' attitudes towards and teachers' awareness of corpora employment in EFL classrooms. In addition, it intended to figure out applicable models, designs, and activities for effectively integrating corpora in vocabulary classes. Therefore, this study could offer applicable implications and practical recommendations on using Corpora in vocabulary classes to EFL teachers, professional, EFL/ ESL learners, language researchers and linguists. Findings could contribute in improving the application of corpus-based approach in teaching English as a second language in the Saudi context.

2.13 Conclusion and Summary for Literature Review

Previewing the previous essential relevant studies and discussions, the researcher has come to conclusions that can be recapped in the following elements: (1) most EFL/ EFL learners, including Saudi learners, consider vocabulary as a challenging essential part in language learning and mastery. (2) It is quite obvious that actual practices of teaching or learning vocabulary in EFL classrooms merely focus on specific limited items of vocabulary learning, i.e. the meaning and the form of single words while the other essential items of a word are completely or partially neglected, i.e. semantic prosody, collocations, colligation, context, registers, and clusters or the formulaic language. (3) These neglected items of word study, especially collocations and co-occurrences, are of great importance for learner's fluency and mastery of a language. (4) Even learners of advanced levels have

revealed little knowledge and use of this formulaic language, natural word clusters, which in turn hinders them from attaining near native level of language proficiency. (6) there are some discrepancies between corpus-based research recommendations and implications and the on-ground teaching practices in EFL classrooms. (7) The great and rapid advancement in technology is not substantially exploited in EFL classroom, especially in regard to the available, user-friendly and free online language corpora. (8) There is a scarcity of experimental research in the field of corpus-based approach in the Saudi EFL classrooms. Based upon the aforementioned conclusions, the researcher intended to help EFL learners and teachers adopt attested viable approach in teaching or learning vocabulary which considered one of the crucial elements of language mastery, through shedding light on the gap between recent research recommendations and teachers/ learners actual perceptions and practices, and testing the efficacy of adopting corpus-based approach in vocabulary learning/ teaching.

CHAPTER III

RESEARCH METHODOLOGY

3.0 Introduction

The current study has followed an experimental method that is designed to offer answers to the main research questions and provide more information related to the potential vocabulary gains and retention rate after applying corpus-based approach. This type of research, quasi-experimental or experimental research design essentially aim at investigating the casual relationship between two variables in the treatment, independent and dependent variables. For more specifics, the independent variable is the influential variable which has a direct impact on the dependent one (Loewen & Plonsky, 2016). Thus, adopting the independent variable will inevitably cause a change or variation in the dependent variable. In this study, the independent variable was considered the use of corpus-based approach which was assumed to have direct impact on the dependent variable, experimental group's learners' post-tests scores.

On the other hand, there are some embedded moderating variables that are defined as variables that modify the relationship between an independent variable and a dependent variable. In this case, the moderating variable is the experimental group's learners' English proficiency levels which may have an impact on the casual relation between the independent variable, using corpus-based approach, and its dependent ones, vocabulary learning and retention scores.

Therefore, the design has included a placement test, a pretest, a posttest, and a delayed posttest in which all learners of both groups had to do. The purpose of both the placement test and the pretest is to ascertain and prove that the two randomly assigned groups were comparable before the treatment, while the immediate posttest measures the immediate vocabulary gains of the treatment, a six week of corpus-based or conventional teaching. In addition to these tests, a two-week delayed posttest has been included to examine the effects of the treatment over the longer term. The researcher has substantially employed quantitative analysis on the collected data from tests scores and questionnaires.

The research questions are: 1. What is the effect of using corpus-based approach on improving vocabulary learning in Saudi EFL classroom? 2. How far is corpus-based approach effective in improving vocabulary retention in Saudi EFL classroom? 3. What is the effect size of learner's proficiency level (as a moderator variable) in the effectiveness of using corpus-based approach in vocabulary learning? 4. What are the attitudes of the experimental group students towards using corpus-based approach in vocabulary learning classes? 5. To what extent are EFL teachers in Saudi Arabia aware of using corpus-based approach in EFL classrooms?

3.1 Research Procedures

The researcher has designed the procedures of this study in a way that provide reliable and validated data for testing the study hypotheses and answering its questions. These procedures can be summarized in the following steps. First, the researcher has randomly assigned two groups as the participants of the study; the experimental group, and the control group. The experimental group has been taught vocabulary lessons through using corpus-based approach and utilizing COCA Corpus to enhance their learning experience, whereas the control group has been taught the same target vocabulary via traditional approach. Second, the researcher has designed and applied Cambridge General Placement Test, target vocabulary pre-tests, post-tests, and two-week delayed post-tests on all learners of experimental and control groups to assess their vocabulary knowledge and retention before and after the treatment. Third, the researcher has designed and administered a questionnaire to examine the attitudes of the experimental group's students towards employing corpus-based approach and corpora related activities in vocabulary learning. This questionnaire has been administered at the end of the application process. Fourth, A follow- up teacher's questionnaire has been used with a randomly selected group of university English teachers in Qassim Region, Saudi Arabia, to find out their perceptions on and awareness of corpus-based approach in EFL classrooms. Lastly, the quantitative data, taken from tests results and questionnaires responses, has been statistically analysed using Statistical Program for Social Sciences Package (V.25).

Each tool has been intended to investigate or measure a question of the study. For illustration, with the classroom-based experiment in which the researcher has applied conventional approach vs corpus-based approach on the control group and experimental group respectively, the researcher has used tests scores as a tool to measure the effectiveness of corpora integration approach on learners' learning and retention of vocabulary. Second, learners' questionnaire as a quantitative tool has been intended to investigate learners' perceptions of and attitudes towards corpora integration in vocabulary classes and reinforce results taken from the application of classroom-based study on the experimental group. Lastly, teachers' questionnaires to investigate teachers' levels of awareness of the use of corpus-based approach in vocabulary teaching, especially those of tertiary or university levels. (See Figure 3.1 an Overall Research Design)

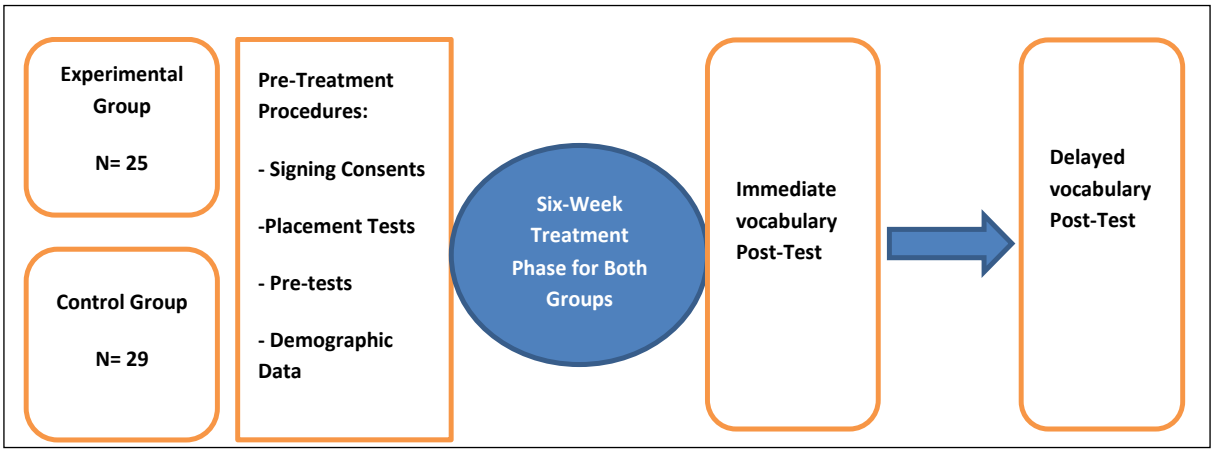


Figure 3.1: The Overall Research Design and Data Collection Procedures.

3.2 Study Population and sampling

The researcher has investigated the first-year students of Foundation Program for Humanities and Administrative Sciences College at Buraydah Colleges in Qassim, Saudi Arabia. They have to study a general English course of 20 hours a week for 16 weeks, 5 hours of them is devoted to reading comprehension and vocabulary learning through using a coursebook titled *Interactions 1 Reading (Sixth Edition) by Elaine Kirn and Pamela Hartmann, published by McGraw Hill Incorporation* (See Appendix A Reading Course Book Content and Objectives).

Oftentimes, one instructor teaches Eng. 140 course, reading and writing sections, for one or two sections of students in a semester. Each section or group of students normally ranges from 20 to 40 students. In this course, ENG. 140 they study four parts of general English, Reading, Writing, Listening and Speaking, for 20 hours in a weekly basis. Specifically, Reading and vocabulary section is taught for at least 5 hours weekly to cover five chapters of reading and vocabulary coursebook, *Interactions 1 Reading by Elaine Kirn and Pamela Hartmann (2007).*

The researcher has already made all required consents and arrangements for conducting this experiment at the end of the first semester of the academic year 1440/1441H. (See Appendix B. Buraydah Colleges Official Consent). All students in the experimental group have been notified about the study and its procedures and had accepted to be part of the experiment and they had given the right to withdraw from the experimental group to the control one at any time. (See Appendix C. Learner's Consent Form).

3.2.1 Participants

The study sample consists of 54 students studying foundation courses at Humanities and Administrative Sciences College, at Buraydah Colleges in Saudi Arabia. They are mostly 18 to 22 years enrolled in Foundation Courses to pursue their studies at one of four available programs, Accounting, Law, Human Resources Management, and Business Administration at Buraydah Colleges. Participants have been randomly placed into the control and the experimental condition following a chance procedure (Gravetter & Forzano, 2018; Hatch & Lazaraton, 1991; Kirk, 2009; Loewen & Plonsky, 2016; Nunan, 1992). One class has become 29 students and has been assigned as the control group (traditional approach for teaching reading and vocabulary), and the other one ($n=25$) has been assigned as the experimental group that received a different treatment (corpus-based teaching for reading and vocabulary). Both groups have been taught by the researcher who has been teaching general English course, reading and writing sections, for 10 hours a week for each, at least 5 hours of them has been devoted for teaching vocabulary and reading skills, with much focus on vocabulary knowledge and use in the first three chapters of the reading course book. (See Appendix D. Detailed Plan for the classroom Treatment). Students

enrolled in the experimental group have been informed about the research and signed a consent form to participate in the experiment. In the meantime, they have been told that they have the right to participate in this experimental research or to withdraw from the experiment if they feel uncomfortable at any time without any consequences.

3.2.3 Participants Demographic Data

Earlier, the researcher collected the personal data of participants and quantitatively analyzed it. Then, he concluded that the participants of both groups had many things in common such as language background, technological skills, age and the availability of a personal computer and internet connection. Both groups have very close mean responses and percentages regarding these points. The main findings for this part can be summarized in the following bar charts in figure (3.2).

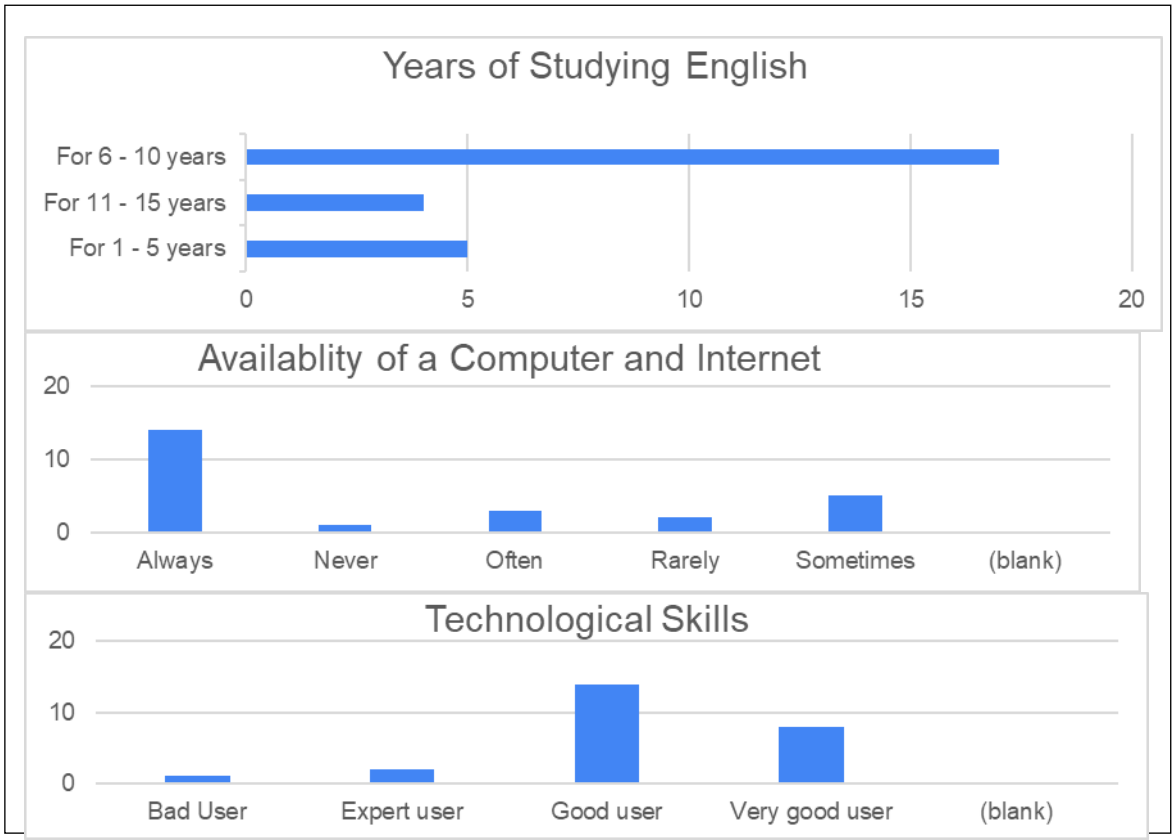


Figure 3.2: Participants Demographic Data

For more details, 70.8% and 20.8% studied English for time periods of (6-10 years) and (11- 16) respectively in their pre-university stages. Moreover, 56.5% and 30.4% were very good and good technology users respectively. Lastly, 52.2% and 13% of participants were always or often having access to a computer and internet connection while only 3 students out of 24 (12.6%) were rarely or never having available computers and internet connection. It can be concluded from the results presented above that most of participants have solid and reliable language and technology backgrounds that qualify them to be part of the experiment which requested using computers and online language corpora in some tasks and activities, especially for experimental group, in their vocabulary classes.

3.3 Data Collection Tools

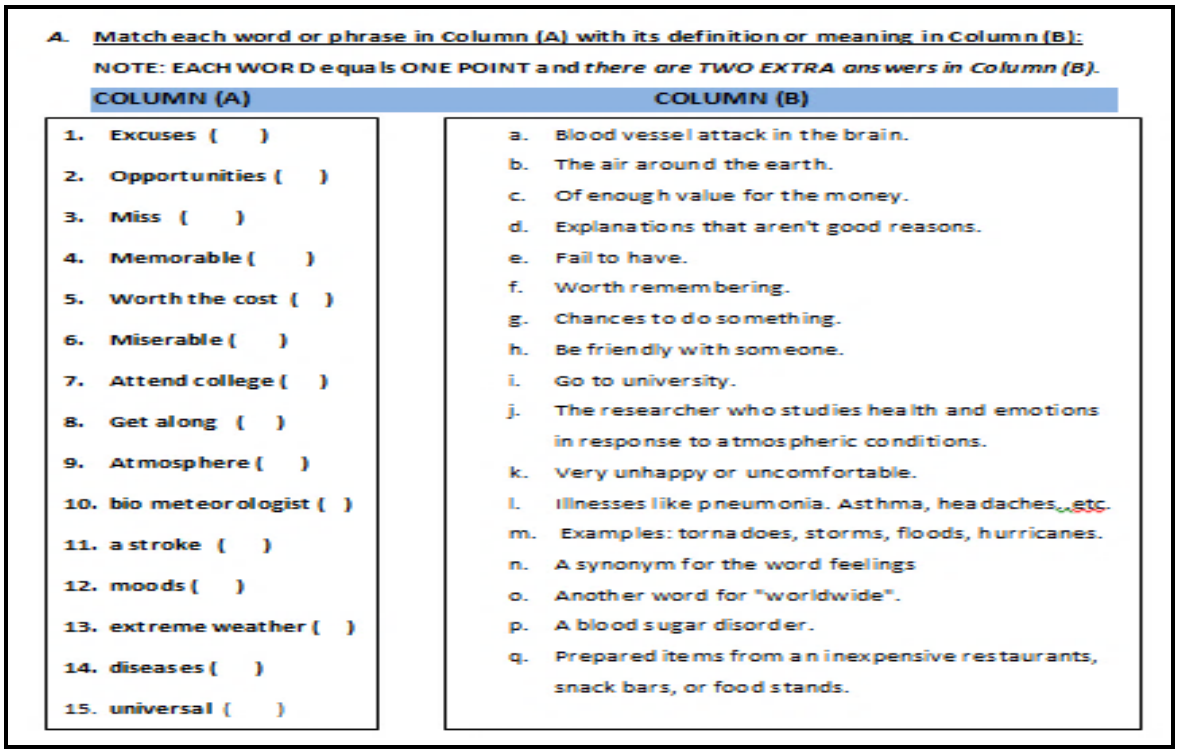
Three essential tools have been used to collect the study data to serve the purpose of answering the intended research questions. According to the aforementioned methodology used in this study, the researcher has designed and applied a variety of tools to achieve the integrity and reliability of results and interpretations. In addition, he has employed a variety of statistical tools to analyze the data obtained from tests and questionnaires. The next paragraphs will show detailed descriptions for each tool and its application procedures, reliability and validity.

3.3.1 Tests

The substantial assigned tool to measure the effectiveness of using corpus-based approach vs traditional approach on vocabulary learning and retention is vocabulary achievement tests. The researcher has applied four tests on participants of both groups, 54 students. At the early beginning of the study, the researcher has applied Cambridge General Placement Test (CPT) on students of both groups to figure out the differences in language proficiency level between both groups' students (See Appendix E. for Cambridge General Placement Test). This placement test is fundamental for the study feasibility as it has intended to find any statistically significant differences in language levels between the two assigned groups. Then, at the beginning of the first week came the vocabulary pretest to measure students' potential pre-existing vocabulary knowledge (See Appendix F. for Pre-

Test). After six weeks of study for both groups, experimental and control, an immediate post-test has been applied to measure their gains after six-week treatment (See Appendix F. for Post-Test). The main aim for this test is identify whether or not there are statistically significant differences in gains between the two groups after going through six –week treatment and having been exposed to different teaching approaches, traditional approach with the control group vs corpus-based approach with the experimental group. After a two-week interval, another delayed post-test has been applied on both groups to measure the differences, if any, in retention rate for both groups (See Appendix F. for Delayed Post-Test).

The tests were organized into two main equal categories, receptive or definitional knowledge and productive or transferrable knowledge; each part has 25 points out of 50. Both tests were identical, except that the items in the post-test and delayed post-test were re-arranged to dismiss the effect of memorizing the answers. It is noteworthy here that the pre-and post- tests are equal in terms of number of questions, target vocabulary items, marks distribution and question types. However, students of both groupps have not been told that the pre-test will be taken a second time as a posttest and a delayed post-test. Tests are mainly based on the assigned target vocabulary in the three chapters in *Interactions 1 Reading Book*. The main themes of these tests were assessing students' perceptive and productive knowledge of the target vocabulary. All Tests, Pre or post or delayed post-tests, are composed of 50 English vocabulary questions (25 receptive questions testing for word meaning recognition, synonyms, and spelling and 25 productive sentence or authentic concordances lines completion questions) have been used to measure Saudi EFL learners' target vocabulary knowledge of the 116 target words, as in the following examples (Figure 3.3 and Figure 3.4):

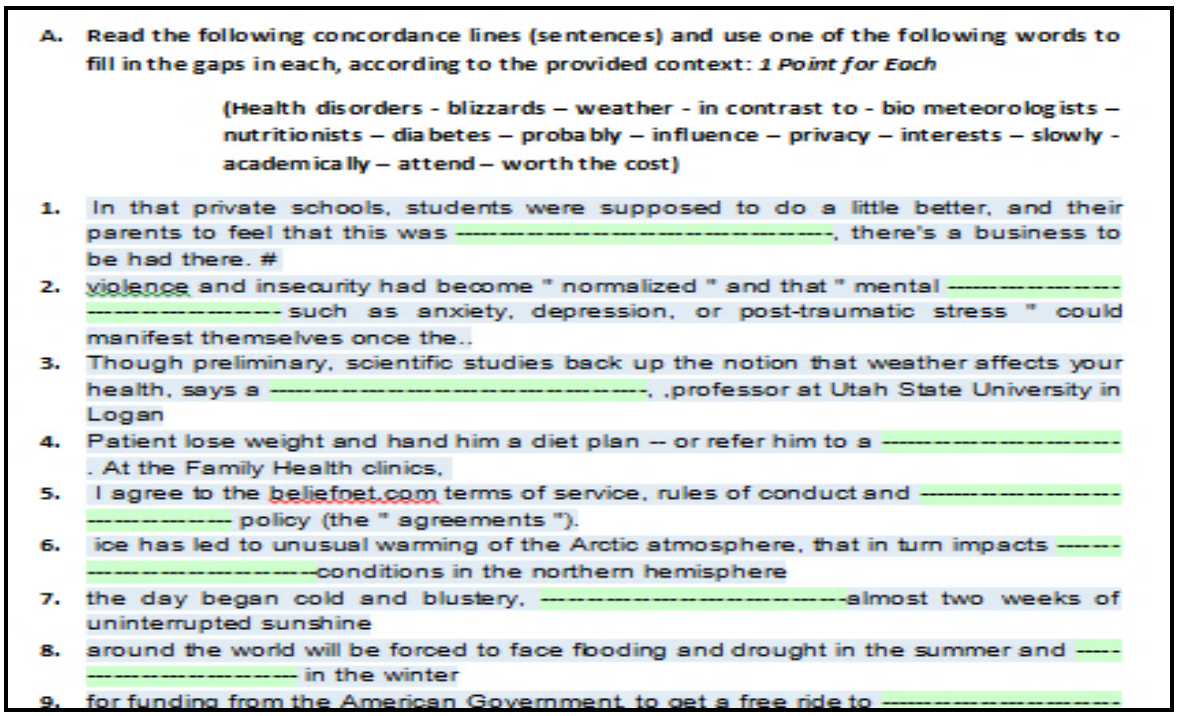


A. Match each word or phrase in Column (A) with its definition or meaning in Column (B):

NOTE: EACH WORD equals ONE POINT and *there are TWO EXTRA answers in Column (B).*

COLUMN (A)	COLUMN (B)
1. Excuses ( )	a. Blood vessel attack in the brain.
2. Opportunities ( )	b. The air around the earth.
3. Miss ( )	c. Of enough value for the money.
4. Memorable ( )	d. Explanations that aren't good reasons.
5. Worth the cost ( )	e. Fail to have.
6. Miserable ( )	f. Worth remembering.
7. Attend college ( )	g. Chances to do something.
8. Get along ( )	h. Be friendly with someone.
9. Atmosphere ( )	i. Go to university.
10. bio meteorologist ( )	j. The researcher who studies health and emotions in response to atmospheric conditions.
11. a stroke ( )	k. Very unhappy or uncomfortable.
12. moods ( )	l. Illnesses like pneumonia. Asthma, headaches, etc.
13. extreme weather ( )	m. Examples: tornadoes, storms, floods, hurricanes.
14. diseases ( )	n. A synonym for the word feelings
15. universal ( )	o. Another word for "worldwide".
	p. A blood sugar disorder.
	q. Prepared items from an inexpensive restaurants, snack bars, or food stands.

Figure 3.3: Samples of Receptive Vocabulary Question.



A. Read the following concordance lines (sentences) and use one of the following words to fill in the gaps in each, according to the provided context: *1 Point for Each*

(Health disorders - blizzards – weather - in contrast to - bio meteorologists – nutritionists – diabetes – probably – influence – privacy – interests – slowly - academically – attend – worth the cost)

1. In that private schools, students were supposed to do a little better, and their parents to feel that this was ----------, there's a business to be had there. #
2. violence and insecurity had become " normalized " and that " mental ---------- such as anxiety, depression, or post-traumatic stress " could manifest themselves once the..
3. Though preliminary, scientific studies back up the notion that weather affects your health, says a ----------, ,professor at Utah State University in Logan
4. Patient lose weight and hand him a diet plan -- or refer him to a ---------- . At the Family Health clinics,
5. I agree to the beliefnet.com terms of service, rules of conduct and ---------- policy (the " agreements ").
6. ice has led to unusual warming of the Arctic atmosphere, that in turn impacts ----------conditions in the northern hemisphere
7. the day began cold and blustery, ----------almost two weeks of uninterrupted sunshine
8. around the world will be forced to face flooding and drought in the summer and ---------- in the winter
9. for funding from the American Government to get a free ride to ----------

Figure 3.4: Samples of Productive Vocabulary Questions.

The topics and the relative weight of the questions in the vocabulary achievement tests have been equally selected from the three chapters. Collocations, contexts, and language chunks have been received much focus. The reliability of the vocabulary achievement test has been tested through applying the test on another sample of 15 male and female students early before the treatment. Test-retest method has been applied on this sample then reliability coefficient factor has been calculated. According to the tests and retests scores of the pilot sample, the reliability coefficient was 0.828; which indicates a good level of reliability. In addition, the researcher found that the Pearson inter-rater reliability was relatively high, 0.906.

After collecting all tests scores for learners SPSS Program was employed to run T-test in order to compare the means of the two groups in both pre-and post-tests and delayed posttest. For more details, the researcher has run a paired T-test to compare the means of pre- and post-tests in one sample (i.e., for each group). Then, he has used the independent samples T-test (also called the Two-Samples T-test) to determine whether there is a statistically significant difference between the means scores of the two groups. Furthermore, for the third question which required to measure the impact of the moderator variable, learners' proficiency levels, on the dependent variable learners' vocabulary gains and retention levels when using corpus-based approach, the researcher used descriptive statistical analysis for the mean scores of the four levels and Two-Ways-Repeated Test to measure the significance of these variances between the means.

3.3.2 Questionnaires

In order to investigate Saudi EFL learners' attitudes towards and views on integrating corpus-based approach in their vocabulary classrooms, questionnaires (See Appendix G for Learners' Questionnaire) have been conducted immediately after the delayed posttests, at the commencement of the ninth week of the treatment. It was composed of two parts, 4 personal questions and 12 five-point Likert questions. This main section covered three main themes: 1. benefits of COCA corpus use in vocabulary learning, 2. difficulties and challenges in using corpus, and 3. their overall perceptions of corpus use in vocabulary classes. The adopted scale for the questions was five points as follows

“strongly disagree” = 1, “disagree” = 2, “Neutral” = 3 “agree” = 4, “strongly agree” = 5. After negatively keyed items were reverse scored, Cronbach’s alpha was calculated, and the scale was found to have a high level of internal consistency, R=0.86. The questionnaire was partially adapted from (Yoon & Hirvela, 2004).

Another questionnaire for teachers was designed and administered to investigate university teachers' awareness of corpus-based approach in vocabulary teaching (See Appendix H Teachers’ Questionnaire). It was composed of two parts, 4 personal questions and 10 Likert- type questions. The adopted scale for the questions was five points as follows “strongly disagree” = 1, “disagree” = 2, “Neutral” = 3 “agree” = 4, “strongly agree” = 5. After negatively keyed items were reverse scored, Cronbach’s alpha was calculated, and the scale was found to have a high level of internal consistency, R= 0.936.

To make sure that both questionnaires are valid and reliable for collecting the data and measuring the assigned target, it has been sent to the jury, 7 experts from various related disciplines, including university and teacher training lecturers in the field of educational technology, methodology and applied linguistics (See Appendix I for A Jury Consent Form and the Jurors' List).

3.4 Data Analysis

At the end of the study the researcher surveyed EFL university teachers’ awareness of using corpus-based approach in teaching vocabulary classes using an electronic questionnaire designed and distributed through Google Forms Application. The results of this questionnaire have been quantitatively analyzed to come up to solid facts about the level of teachers’ awareness of the use of Corpora in teaching vocabulary. Simultaneously, another questionnaire has been also administered and done through Google Forms for the experimental group’s learners to measure their attitudes towards using corpus-based approach in vocabulary learning and retention and whether they will use it later on their own. Collected data of both surveys have been quantitatively analyzed by using SPSS software, through descriptive statistics analysis features to answer questions four about learners’ attitudes and five about teachers’ perceptions and awareness.

Lastly and most importantly, at the end of the treatment, in the ninth week, the researcher has analyzed the tests scores for pre-, post-, and delayed post-tests for both groups, experimental and control to test the main hypothesis of the study. To get a detailed and authentic answer for the first and main study's question: First the researcher measured the difference between pre-test and post-test scores for each individual group and ran a t-test for each group separately, One Sample T-test, to measure the gains of each group after the six-week experiment. Second, the researcher ran Independent Samples T-test to compare the mean differences of corpus-based group's post-test scores to the conventional group's scores to test if there were significant differences in the final gains between them. For the third question of the study, the researcher employed descriptive statistics analysis on the mean scores of pre and post-tests of the four sub-groups (elementary, pre-intermediate, intermediate, and upper-intermediate) to find the variances among them; then the research has used Two Ways-Repeated ANOVA Test to measure the significance of these variances among sub-groups.

3.5 Instructional Procedures

For experimental group, the instructor has used corpus-based instruction and corpora-based materials to present target vocabulary in each assigned reading passage and to reinforce previous learnt ones. Teaching has been happening in English Lab to enable learners to instantly access the assigned online corpus and be participating in corpora activities prepared by the researcher, whereas for the other control group, teaching has been happening in the traditional classroom using a whiteboard and sometimes slideshows projection. All computers in the English Lab are equipped with internet connection and wireless network for smartphones and personal laptops to help learners' instant access to online corpora. In addition, the researcher has employed Edmodo online platform for homework assignments, class announcements, quizzes and teachers' feedback on group work assignments submissions. Both Classroom instruction or Edmodo feedback and follow-up have been available equally and fairly for participants of both groups on a daily basis regardless of the form and nature of required activities and assignments (See Tables 3.1 and 3.2 for more details).

Early orientation sessions have been held for participants of the experimental group at the early beginning of the experiment to be familiarized with using COCA Corpus tools and searching options and be introduced to the meaning of collocations and formulaic language rather than single separated words. In the study, participants not only received training on the technical use of the corpus but were also given guidelines on how to read and analyze the concordance lines, following Sinclair's (*Reading Concordances*) model. This enhanced their performance in the answering activities. As seen in previous research, when participants receive technical guidelines without training on how to read and analyze the concordance lines, or if they undergo a short training period, the result is either insignificant performance or less positive attitudes (e.g., Pérez-Paredes et al. 244; Gaskell and Cobb 311-17). Sinclair's framework informed the design for the teaching materials used in this experiment, and the activities were divided according to the framework which consists of seven steps: initiate, interpret, consolidate, report, recycle, result and repeat (Sinclair, Reading Concordances xvi-xvii). Sinclair's reading concordance model, adopted here, is both comprehensive and detailed. I chose this framework because it encourages three major cognitive processes in learning vocabulary: noticing, retrieval and creative use. Nation (98- 100) argued that these three cognitive processes are significant conditions for improving vocabulary learning as well as in evaluating vocabulary-teaching tasks.

Substantially, at the end of the course, participants of both groups have to acquire the same target vocabulary for each chapter whatever the employed teaching approach is (See Appendix D. for Detailed Plan for the Classroom Treatment).

Table 3.1: An Overview of the Treatment Plan.

Domain	Course Learning Outcomes (CLOs)	Teaching Strategies and Learning Resources	
		Control Group	Experimental Group
1. Knowledge	1.1 Identify the meaning, definitions, and synonyms for the target words. 1.2 Identify the appropriate collocates for some target vocabulary in real contexts. 1.3 Recognize the correct spelling, pronunciation, part of speech, and use of the target vocabulary in each chapter	• Using the Traditional Classroom with its normal facilities to deliver lessons. • Traditional lecturing approach and other context and definitions strategies will be used for words teaching	• Using the Department's Computer Lab to deliver all the vocabulary classes using available PCs and internet connection. • Discovery learning and autonomous problem-solving approaches will be used to learn target words.
2. Cognitive	2.1 Utilize new learnt vocabulary in authentic contexts and personal situations. 2.2 Demonstrate high level of awareness for words frequency in different genres. 2.3 Utilize word clusters rather than single isolated words. 2.4 Employ words collocation for producing spoken or written utterances.	• Using the assigned Course book, Interactions 1, Reading Section, by *E. Hirn and Hartmann*, to teach the target vocabulary in the first three chapters. • Traditional lecturing approach and other context and definitions strategies will be used for words teaching	• Using the assigned course book for teaching reading strategies only, and using COCA Corpus to learn the target vocabulary in each chapter. • Using printouts and hands-on mainly derived from COCA Corpus Concordance lines and queries data. • Discovery learning and autonomous problem-solving approaches will be used to learn target words.
3. Interpersonal Skills and Responsibility	3.1 Students are able to accomplish the assigned tasks and appropriately fulfill their role in their team, contribute to others' work, and solve their team problems.	• Using work sheets, slides and activities mainly based on the course book activities and exercises. • Students are given pair work and group works tasks based on the course book activities and exercises.	• Using printouts and hands-on mainly derived from COCA Corpus Concordance lines and queries data. • Students are given pair work and group works tasks based on the online use of corpora searching tools and concordancers.
4. **IT and Communication Skills**	4.1 Students are able to use different forms of IT tools appropriately to assist learning and researching, and to give presentations. 4.2 Students are able to use online dictionaries and online	• The teachers will be using the available projector and personal lab top in the class to give them presentations and demos in using online dictionaries.	• All students will be using the available PCs or their own smart phones to get direct access to online corpora use and doing queries about the target vocab.

All students have received a six-week treatment for learning 116 target vocabulary items distributed on three chapters. For experimental group, COCA Corpus will be used for the corpus-based vocabulary instruction and the construction of related exercises. COCA Corpus refers to Corpus of Contemporary American English (COCA). It has been essentially utilized by both teachers and learners of the experimental group to learn or teach the target vocabulary. It can be accessed via the following link: www.english-corpora.org/COCA/. On the other hand, the control group has been taught the same target vocabulary using mainly the assigned course book exercises and activities, *Interactions 1 Reading Course Book.* For both groups, the class has recurred three times a week for two-hour sessions dedicated to instruction of target vocabulary items, about 116 items, over a six-week period.

Taking into consideration ethical issues concerning experimental research, including that there is a group of students may gain additional benefits and another does not, the participants have been given the freedom to choose which class to attend and given the freedom to withdraw to the control group at any point of the experiment.

3.5.1 Courseware Design and Implementation

The reading content for the first level of Foundation Year has been designed to be taught and learnt through corpus-based activities using mainly Corpus of Contemporary American English (COCA). A new course specification for this subject, Reading 1, has been developed by the researcher to be followed in teaching these three chapters of the reading course, Interaction 1 Reading by Kirn, E. &Hartmann, P. (2012), to learners of experimental group while the other control group will be taught the same three chapter following the traditional departmental course specifications which is mainly based on the assigned course book structure and exercises (See Appendix D Detailed Plan for the Classroom Treatment). Each group will be taught by the same instructor, the researcher, to ensure other potential variables related to teachers' different styles. Classes for each group will be happening three times a week, two hours per each class. This will be happening for 6 consequent weeks which means a total 40 hours of teaching for each group, as shown in the following table (3.2).

Table 3.2: Covered Chapters and Target Vocabulary in Each.

No.	Topics to be Covered	Number of Assigned Vocabulary/ per week	Week No.	Number of Hours
1.	Orientation for the Study, Placement Tests, and Pretests	---	1st	4 hr.
2.	Chapter 1: Academic Life around the World.	20	1st	4 hr.
3.	Chapter 1: Academic Life around the World.	41	2nd	6 hr.
4.	Chapter 2 Experiencing Nature.	16	3rd	6 hr.
5.	Chapter 2 Experiencing Nature.	20	4th	6 hr.
6.	Chapter 3: Living to Eat, or Eating to Live.	17	5th	6 hr.
7.	Chapter 3: Living to Eat, or Eating to Live.	12	6th	4 hr.
8.	**IMMEDIATE POSTTEST**	----	6th	2 hr.
9.	**INTERVAL**	---	7th	--
10	**INTERVAL**	---	8th	--
11	**DELAYED POSTTEST**	---	9th	2 hr.
12	**Total Weeks and the Teaching Actual Hours**	**116 words**	**6 weeks**	**40 hours**

The researcher adopted a gradual approach for getting students, of experimental group, engaged in corpus-based learning, starting with detailed workshops on using corpora in vocabulary learning, specifically COCA corpus, then moving a step forward through using worksheets, hands-on, and materials based mainly on concordances lines from COCA on the target vocabulary, and the last step is learners' direct access to COCA. The first stage has been designed to be in the first introductory week of the study in which students are introduced to the course content, learning objectives, assessment methods, and most importantly teaching strategies. They are introduced to the basics and rationale of using corpora in learning vocabulary; a detailed description of the whole experiment processes and their expected roles will be negotiated and compromised with them (See appendix J COCA Corpus Handout, for more details and examples). In the second stage, which will be starting in the commencement of the second week, the teacher will be the main source of guided activities and hands-on which prepared before each class to cover the predetermined target vocabulary for each class. All these activities are based on and substantially derived from concordances lines from COCA corpus. The researcher has to prepare worksheets for teaching and assessment for each target vocabulary through modifying the raw data taken from searches on COCA whereas with the control group he will be teaching the assigned

exercises and activities on their course books with some supporting worksheets for assessment, besides the availability of electronic dictionaries on their cell phones. For much more details about the study plan and course description, see Appendix D (Detailed Plan for the Classroom Treatment) and Appendix K (Printouts and Exercises for Corpus-Based Experimental Group).

The teacher will be the key in planning and designing the individualized classroom tasks and the appropriate mode of corpora integration, basically depending upon the target objectives and content of each chapter and in the light of his students' language capacity to handle these objectives. As the researcher mentioned in the previous paragraphs that it will be a simplified gradual process in which all learners will be smoothly introduced to the corpora, COCA Corpus, through three stages process. These stages can be broadly named introduction and orientation, Teacher's corpora-based activities, and lastly learners' direct access to corpora and individualized tasks. Simultaneously, the teachers can use mixed modes in his classroom; mode 1: teachers make all the teaching materials from concordance lines before the class and project them to students in the classroom. After that, they are asked in groups or in pairs or individually to find out the linguistic rules behind these lines. Lastly, under the teacher's guidance, they need to conclude the collocates, colligation, sematic prosody and context and co-occurrences of the target words. Mode 2: The teacher sets a group work for students to find out "co-occurrence in context" of the target words in the lesson through using COCA corpus to grasp the underlying rules and typicality of co-occurrences. Then he sets a whole class activity, each group shows its selected concordances lines to the whole class, then the class with little direction from the teacher's side summarizes the basic linguistics rules for each word. Mode 3: it's a type of guided exercises essentially derived from concordances lines, such as fill in the gaps, or identify the part of speech for the collocates before and after the key word, or identifying the meaning of words or using the target words in authentic novel texts taken from corpora too. Samples of these exercises and activities as shown in figures 3.5 and 3.6, for many examples of printouts and worksheets, see Appendix M (Printouts and Exercises for Corpus-Based Experimental Group).

A. ***Read the concordance lines below and answer the following questions.***

Attitudes:

approaches to difficult problems , focused specifically on	attitudes	about	crime	.	the findings can be used to understand the
you . It 's just couches ! " # On her	attitude	about	money	?	One of the big things I loved about growing
n't really make that much difference in terms of their own	attitudes	about	the	nominee	. MR-WARREN : It made a great deal of
' use of homophobic epithets . Group-level homophobic	attitudes	and	bullying	behavior	were included at Level 2 to predict the
. " We 've seen remarkable changes in the kids '	attitudes	and	confidence	levels	, " says Pam . " Our introvert ,
Quick to support . Fine range in zone coverage . Great	attitude	and	effort	!	Can get separated from a man-to-man coverage .
urge to relieve suffering toward revising the static , habitual	attitudes	and	norms	that	created it . Zen koan study , contingency "
knowledge of the harmful effects of noise actually affect the	attitudes	and	practices	!	Study design # This is a cross-sectional
mid-1980s , and some people assert that spouses fulfilled their	attitudes	and	responsibilities	for	decades before that . However , the

1. What does the word "attitude" mean?
2. What part of speech is it?
3. What are the prominent collocates in these sentences?
4. Tick the appropriate synonyms for the word attitudes? View, stance, opinion, feeling, richness, academic.
5. Write a sentence on your own using this word?

Figure 3.5: Samples of Printouts for Corpus-based Experimental Group

GROUP WORK TASKS – HOME ASSIGNMENTS - SAMPLE A

Complete the table below, following the stated instructions:

- Work in GROUPS, you and your PARTNERS together to fill in the table.
- Go to the website (www.english-corpora.org/COCA/
- Search for the words (low carb diet – nutritionists – diabetes – fast food - desserts)
- Use Key Word In Context (KWIC) Type of search.
- AS A LAST OPTION, you can use the available links for bilingual dictionaries on the website

Words/ elements	low carb diet	nutritionists	diabetes	Fast food	desserts
Part of speech					
meaning					
Collocations before					
Collocations after					
synonyms					

Figure 3.6: Samples of Hands-on/ Activities for Corpus-based Experimental Group.

It's quite obvious that learners of the control group have their assigned context for the target vocabulary in each chapter, two reading passages of about two pages each, besides their own dictionaries and exercises set by the course book designers who basically set one single topic for a set of words. On the other hand, learners of the experimental

group are introduced to each target word in different contexts and with a variety of frequent collocates in many disciplines which in fact enables them to be immersed in authentic overwhelming language input that in turn is expected to enhance their linguistic knowledge about vocabulary and increase their language competence besides affecting the retention rate of learnt vocabulary. (See Appendix K Printouts and Exercises for Corpus-based Experimental Group).)

A detailed two course descriptions for the specific content of reading course has been prepared by the researcher to be followed during the study, one for the experimental group and the other for the control group. Each description is completely aligned with the required learning objectives and content set by the department of Foundation Courses in Burayadah Colleges, whereas there are some differences in methods of teaching and learning tools (See Table 3.3 Similarities and Differences Between the Two Sample Groups).

Table 3.3: Similarities and Differences between the two Sample Groups

Similarities in the Experiment	**Variances in The Conditions of The Experiment**		
1. The Teacher 2. The Syllabus and Lessons 3. Target Vocabulary 4. Allocated Time 5. Comparable Numbers and Levels 6. Placement Test 7. Pretest/Posttest &Delayed test	**Features**	**Experimental group**	**Control group**
	Method	Corpus-based approach	Conventional Approach
	Classroom	Language laboratory	Traditional classroom
	Materials	Handouts with concordance lines	Worksheets based on coursebook activities
	Activities	Hands-on & paper-based	Paper-based only
	Teacher's roles	Supporter and facilitator Monitor Tasks-planner	Knower Lecturer Administrator Presenter

To be cautious and accurate, each group has taken a placement test to classify participants' proficiency levels in each group and be aware of level differences and measure resulted gains for each constituent level in both groups. This placement test has been adopted by the researcher from Cambridge University Placement Tests, and applied at the early beginning of the first week of the study (See Appendix E. Cambridge General Placement Test). Results for both groups have been analyzed and accordingly students of both groups have been categorized into one of four levels, Elementary, Pre-intermediate,

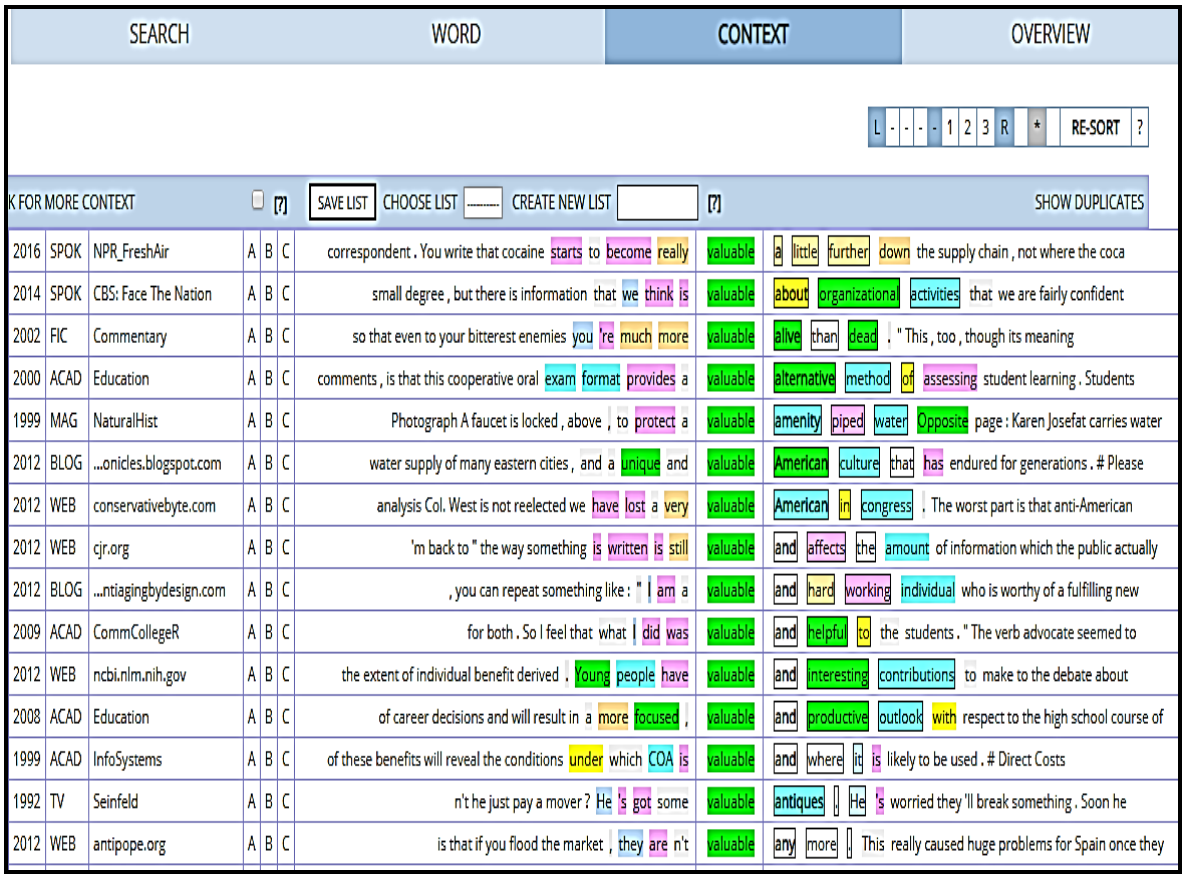
Intermediate, and Upper-intermediate. Levels distribution for learners of both groups has been shown in Table 3.3 (Learners' Proficiency Levels Distribution).

Table 3.4: Learners' Proficiency Levels Distribution

Proficiency Level	Experimental Group	Control Group	Total
Elementary	**5**	**4**	**9**
Pre-Intermediate	**5**	**5**	**10**
Intermediate	**5**	**10**	**15**
Upper Intermediate	**10**	**9**	**17**
Total	**25**	**29**	**54**

3.5.2 Instruction Materials

The coursebook: *Interaction 1 Reading* Course Student's *book (By Kirn, E. &Hartmann, P., 2012)* has been assigned as the main text book in the course. There is a vocabulary section at the beginning of each unit, and there are approximately 25 to 60 target words in each chapter. These sections included exercises like matching and fill-in-the-blanks types but most of them fabricated exercises and activities, i.e. non authentic or real-life examples. For the experimental group, the vocabulary tasks in the textbook were replaced with corpus-based vocabulary tasks; for more samples see Appendix K (Printouts and Exercises for Corpus-Based Experimental Group). Corpus of Contemporary American English (COCA) (Davies, 2012) has been assigned the essential resource to choose appropriate concordance data for the target vocabulary and prepare corpus-based vocabulary materials and tasks. The researcher has chosen COCA Corpus for its richness, nearly one billion words, free availability, and its user-friendly interface. It offers authentic examples from spoken language, fiction, magazines, newspapers, TV and movies, web and blogs, and academic texts. Corpus-based vocabulary activities included analyzing concordance lines and answering questions, matching activities, and fill-in-the blank exercises. This kind of vocabulary activities and group work tasks were completely new and unfamiliar to the participants as most of them had not experienced them before in their EFL classrooms. Figure 3.7 shows a screen shot for using COCA Corpus in searching for the word "*valuable*" through KIWC search type (Key Word in Context).



SEARCH | WORD | CONTEXT | OVERVIEW

L - - - - 1 2 3 R * RE-SORT ?

K FOR MORE CONTEXT [?] SAVE LIST CHOOSE LIST ---------- CREATE NEW LIST [?] SHOW DUPLICATES

2016	SPOK	NPR_FreshAir	A	B	C	correspondent . You write that cocaine starts to become really	valuable	a little further down the supply chain , not where the coca
2014	SPOK	CBS: Face The Nation	A	B	C	small degree , but there is information that we think is	valuable	about organizational activities that we are fairly confident
2002	FIC	Commentary	A	B	C	so that even to your bitterest enemies you 're much more	valuable	alive than dead . " This , too , though its meaning
2000	ACAD	Education	A	B	C	comments , is that this cooperative oral exam format provides a	valuable	alternative method of assessing student learning . Students
1999	MAG	NaturalHist	A	B	C	Photograph A faucet is locked , above , to protect a	valuable	amenity piped water Opposite page : Karen Josefat carries water
2012	BLOG	...onicles.blogspot.com	A	B	C	water supply of many eastern cities , and a unique and	valuable	American culture that has endured for generations . # Please
2012	WEB	conservativebyte.com	A	B	C	analysis Col. West is not reelected we have lost a very	valuable	American in congress . The worst part is that anti-American
2012	WEB	cjr.org	A	B	C	'm back to " the way something is written is still	valuable	and affects the amount of information which the public actually
2012	BLOG	...ntiagingbydesign.com	A	B	C	, you can repeat something like : " I am a	valuable	and hard working individual who is worthy of a fulfilling new
2009	ACAD	CommCollegeR	A	B	C	for both . So I feel that what I did was	valuable	and helpful to the students . " The verb advocate seemed to
2012	WEB	ncbi.nlm.nih.gov	A	B	C	the extent of individual benefit derived . Young people have	valuable	and interesting contributions to make to the debate about
2008	ACAD	Education	A	B	C	of career decisions and will result in a more focused ,	valuable	and productive outlook with respect to the high school course of
1999	ACAD	InfoSystems	A	B	C	of these benefits will reveal the conditions under which COA is	valuable	and where it is likely to be used . # Direct Costs
1992	TV	Seinfeld	A	B	C	n't he just pay a mover ? He 's got some	valuable	antiques . He 's worried they 'll break something . Soon he
2012	WEB	antipope.org	A	B	C	is that if you flood the market , they are n't	valuable	any more . This really caused huge problems for Spain once they

Figure 3.7: A sample for using KWIC Type of query in COCA Corpus for the adjective "VALUABLE":

3.5.3 Instruction Plans and Teaching Materials Validation

The research has consulted 7 experts in the fields of ELT, Linguistics, Curricula Design and Research Methodologies to check the validity of all the instructional materials, assessment procedures, learning objectives, and lessons plans before getting started in the application process. This validity check has included all the proposed plans, lessons objectives, assessment timetable, target vocabulary distribution, teaching methods for each group, proposed activities and tasks, the assigned coursebook and the assigned corpora, and finally the achievement vocabulary tests (See Appendix I for the Jury of Research Tools). First, the researcher received some valuable suggestions and feedback from the Jury and accordingly he did some changes to the tests and the activities. Second, another draft has been sent back to the jury for their final assessment and remarks. As it has been shown in the following table (Table 3.5), there was a high percentage of satisfaction among the jury

on the appropriateness of the proposed teaching tools, assessment tools, teaching plans, assigned coursebook and assigned corpus, and designed activities and tasks for both groups.

Table 3.5: Teaching Plans and Materials Validation Results

Study Tools and Criteria of Assessment	Minimum	Maximum	Mean	St. Deviation
1. Teaching Plans: Criteria	**-0.5**	**1**	**0.6428**	**0.5149**
A. Fit the Time Framework	-0.5	1	0.5714	0.612
B. Appropriate to the Levels	0	1	0.5	0.651
C. Equal Content for both Groups	0	1	0.5714	0.606
D. Measurable Objectives	0.5	1	0.9285	0.188
2. Assessment Plans: Criteria	**-0.5**	**1**	**0.7321**	**0.2672**
A. Convenient in Respect to the Time Table	-0.5	1	0.5	0.4082
B. Appropriateness of Assessment Tools	1	0.5	0.7142	0.2672
C. Representativeness for The Proposed Objectives	1	1	1	0
D. Equality between the Two Groups	0	1	0.7142	0.3933
3. Activities and Tasks: Criteria	**-0.5**	**1**	**0.75**	**0.2868**
A. Achieving the planned Objectives	0.5	1	0.8571	0.2439
B. Fit the Adopted Approach for Each Group	0.5	1	0.7857	0.2672
C. The Variety of Types and Forms	0.5	1	0.9285	0.1889
D. Promoting Students' Participation and Passion	-0.5	1	0.4285	0.447
4. Pre-tests and Post-tests: Criteria	**-0.5**	**1**	**0.7714**	**0.3296**
A. Diversity of Used Questions	1	1	0.7857	0.3933
B. Representativeness of the Target Vocabulary	-0.5	1	0.4285	0.6726
C. Suitable Length and Time for the Test	-0.5	1	0.7142	0.3933
D. Meeting Different Achievement Levels	0.5	1	0.9285	0.1889
E. Convenient Distribution of Marks.	1	1	1	0

N.B.: Strongly Agree = 1 *Neutral = 0* *Disagree = -0.5*
Agree = 0.5 *Strongly Disagree = -1*

The mean scores of the Jury assessment for the four illustrated elements were relatively high as it ranges from *Agree* (0.5) to *Strongly Agree* (1). The highest validity score (0.77) was given to the criteria of Pre-tests and Post-tests then immediately came the validity of designed activities and tasks (0.75), while assessment plans and teaching plans were (0.73) and (0.64) respectively. It is noteworthy that the majority of the jury (6 out of 7) highly valued the appropriateness of the planned objectives in the teaching lesson plans and the diversity and purposefulness of the proposed activities and tasks especially for corpus-based group. On the other hand, three of them negatively rated the appropriateness

of corpus-based activities in promoting student's passion and participation. Some of them justified that most of these activities were not familiar to the students and they might not be enthusiastic to try new things. Overall, the research has already applied some changes based on the initial feedback from the jurors in lesson plans, activities and tasks designs, and tests. Therefore, the overall ratings of the jury for the final drafts of instructional tools and plans were relatively high and positive.

3.5.3 Testing and Progress Assessment for Both Groups

Students were supposed to acquire about 116 words distributed over three chapters in a time period of six weeks, three classes per week. The researcher has been trying to assure and ensure the equality and consistency of teaching delivery, time, support and assessment as possible as he could in a strict and hard endeavor to exclude other interrelated factors that may positively or negatively affect the results of the study. Each group has been clearly introduced to the assigned course, *Reading 1 Interaction 1,* through a presentation and handouts that outline the course main components. However, for the experimental group much time and efforts have been done to introduce them to the unfamiliar teaching and learning approach, corpus-based approach, and the use of COCA corpus as previously mentioned. The assessment plan is the same for both groups, i.e. Pre-post and weekly tasks tested both experimental and control groups in both definitional knowledge as well as transfer of knowledge to a novel context. The table below is shown the scheduled assessment plan for learners of both groups (Table 3.6).

Table 3.6: Scheduled Assessment Plan for Learners of Both Groups, Experimental and Control.

Assessment Methods	Assigned Week	Intended Aim
Placement Test (Cambridge General Placement Test) (50 Marks)	The Beginning of the 1st week	Categorizing Students into their expected levels
Pre-test (50 Marks)	The second day of the 1st week	To measure students' knowledge about the target vocabulary
Post-test (50 Marks)	Thursday, at the end of the 5th week	To measure students gains after going through the study for 5 weeks
Delayed post-test (50 Marks)	At the end of the 8th week	To measure student's retention rate for the learnt vocabulary after an interval of 3 weeks.
Experimental Group Learners' Questionnaire (50 Marks)	At the beginning of the 9th week.	To investigate Experimental Group's learners' experience with Corpus-based approach in learning vocabulary

3.6 Ethical Considerations

Ethical considerations in research are critical, because it can affect the integrity of research. In the proposed study, I have followed the ethical research guidelines of Sudan University of Science and Technology. The first step is to obtain participants' consent and then ensure the confidentiality of all the participants by using pseudonyms for the participants. Each participant has been provided an informed consent to read and review. This consent confirms the participants' understanding of the study and their agreement to participate. In case the participants feel uncomfortable, they have been given an option to withdraw from the study. The researcher has been saved all the data on a password protected personal laptop and will share the data with supervisor. A backup file on an external Google Drive has also been made.

Early enough before the commencement of the study, the researcher explained the experiment's objectives and stages for the students and they have been assured that it is a volunteering job and there won't be further consequences for withdrawing or quitting at any stage of the study. They have been reassured that all quizzes and tests will not be counted for their final assessment or their grades for the course Eng. 140, Reading 1. The consent forms and all related materials have been safely kept in a lock in the researcher's house for 3 years and then will be torn up before disposal. All participants have been de-identified for privacy issues, so specified code for each participant has be used. All these procedures have been followed by the researcher to ensure the ethical code of the study and safeguard the participants' privacy.

3.7 Summary

To sum up, this chapter is intended to provide a detailed description of all study procedures, data collection tools, data analysis, and sampling of the population. The learning product of the two treatments was measured by changes that occurred in the gains of the learners' receptive and productive knowledge of the target 116 vocabulary items. The elicitation instruments were filling in the gaps, matching, multiple choice, and short answer questions. It was hoped that such a multi-treatment investigation would create a completer

and more comparative picture of the proposed pedagogical approach. The analysis and findings will follow in the next chapter. The following table is showing an overall summary for the research methods.

Table 3.7: Summary of Research Methods

Research Questions	Data Collection Tools	Data Analysis
1. What is the effect of using corpus-based approach on improving vocabulary learning in Saudi EFL classroom?	- Placement Test. - Pre-Test. - Post-Test.	- One Sample Independent Test. - Independent Samples T-Test.
2. How far is corpus-based approach effective in improving vocabulary retention in Saudi EFL classroom?	- Placement Test. - Pre-Test. - Post-Test. - Delayed Post Test.	- One Sample Independent Test. - Independent Samples T-Test.
3. What is the effect size of learner's proficiency level (as a moderator variable) in the effectiveness of using corpus-based approach in vocabulary learning and retention?	- Placement Test. - Pre-Test. - Post-Test. - Delayed Post Test	- One Paired Sample T-Test. - Two ways ANOVA Test
4. What are the attitudes of the experimental group students towards using corpus-based approach in vocabulary learning classes?	Learners' Questionnaire.	- Descriptive Statistical analysis
5. To what extent are EFL teachers in Saudi Arabia aware of using corpus-based approach in EFL classrooms?	Teachers' Questionnaire	- Descriptive Statistical analysis

CHAPTER IV

RESULTS ANALYSIS AND DISCUSSIONS

4.0 Introduction

The results of this study have been analyzed and interpreted in the light of the proposed questions and hypotheses. The first, second, and third questions were answered and their hypothesis were tested by employing T-tests and ANOVA tests on the tests scores of both groups, Experimental and Control. On the other hand, the fourth and fifth questions were answered through running descriptive statistical analysis on the responses of teachers' and learners' questionnaires. The next paragraphs will show and interpret results in the same order of the study questions starting with the first and main question which measures the effect size of using corpus-based approach on students' vocabulary learning. However, a fundamental hypothesis has to be tested first as an essential basis for measuring the gains of both groups after going through the treatment; it is the hypothesis of comparability of the assigned groups which will be handled in details in the following paragraphs.

4.1 The Comparability of the Assigned Groups

The researcher followed an experimental design, as he randomly selected and assigned participants into two groups (Experimental and Control), and for detecting the resulted differences in gains between these assigned groups a preliminary hypothesis has to be proven first. This is the hypothesis of comparability of the randomly assigned groups, experimental and control, which considered a baseline and foundation for any further measures or analysis or interpretation concerning the actual gains of vocabulary learning for learners. A detailed description for all tests and scores analysis that have been applied to get this hypothesis proven is shown in the following paragraphs.

As it has been detailed in the previous chapter, this study has adopted a non-equivalent control group design; then randomization as in true experimental design has not been attainable for the researcher. In addition, the use of intact classes in quasi-experimental design is presumably desirable in many educational research contexts as it doesn't impact or impair the existing school system (Porte, 2002). The researcher had given

access to two intact classes in Burayadah Colleges that had very similar educational backgrounds and learning circumstances. Initially, all these students have been allocated to the study. Then, students in the two classrooms were randomly assigned to experiment group 1 (Corpus-based method in teaching the target vocabulary, *N*= 25), and a control group (Conventional methods in teaching the target vocabulary, *N* = 29).

Prior to this assignment, it was crucial to make sure that any variation in research results between groups could not be attributed to variations in the participants' English proficiency levels or vocabulary knowledge backgrounds. To address these two issues, two commonly used and freely available tests were administered to the participants: Cambridge General Placement Test (CPT) and Vocabulary achievement pre-test prepared by the researcher based on the target vocabulary. It was thus necessary to have empirical evidence that the two assigned groups are comparable and there are not any statistically significant differences between the mean scores of their placement test and pre-test scores.

4.1.1 CPT Proficiency Exam

Generally speaking, all students in the first year of their university study have already studied English in their pre-university stages for at least 9 years. Consequently, the great bulk of the population or the sample of this study are expected to be at the intermediate level and above in English proficiency. In order to validate this claim, all participants in each of the two groups have been given a placement test CPT (Cambridge General Placement Test) at an early stage of the study. The scores of both groups, experimental and control, have revealed diversity in learners' proficiency levels. On one hand, 64.2% of control group's learners are categorized as intermediate, upper-intermediate and advanced levels and only 35.6% are categorized as elementary and pre-intermediate levels. On the other hand, 60% of experimental group students are categorized as intermediate, upper-intermediate and advanced levels and only 40% are categorized as elementary and pre-intermediate levels. The mean score for the experimental group is 28.2 out of 50 while it is 25.8 out of 50 for the control group as it is shown in details in table (4.1).

Table 4.1: Placement Test Scores Descriptive Statistics for Both Groups.

Test	Group	N. of Sts.	Mean	Std. Deviation	Std. Error Mean
PLACEMENT TEST	Control	25	25.79	8.441	1.688
	Experimental	28	28.20	7.997	1.511

Though there have been slight differences between the scores of both groups as it has been just shown, we don't know whether they accordingly affect the treatment or don't. Therefore, the researcher has run independent samples T-Test to prove whether these differences in proficiency levels, placement test scores, statistically significant, at the level P≤ 0.05, that may have direct impact on the treatment results or NOT. It has been shown in table (3.7) that P value is 0.29 which is much greater than the significance level ≤ 0.05. The above shown results proved that the slight differences in proficiency levels are not statistically significant and thus warranted the comparability of the randomly assigned two groups, the control and the experimental and supported dismissing the variable of groups' proficiency levels from affecting the scores of vocabulary achievement post-test at groups level. (See table 4.2 and Figure 4.1)

Table 4.2: T-Test for Placement Test Scores for Both Groups.

Independent Samples Test for Placement Test Results for Both Groups (Experimental. and Control)										
Parameter / Test		***Levene's Test for Equality of Variances***		***t-test for Equality of Means***						
		F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
									Lower	Upper
PLACEMENT TEST	Equal variances assumed	.586	.448	1.069	51	.290	2.414	2.259	-2.120-	6.949
	Equal variances not assumed			1.066	49.578	.292	2.414	2.266	-2.138-	6.966

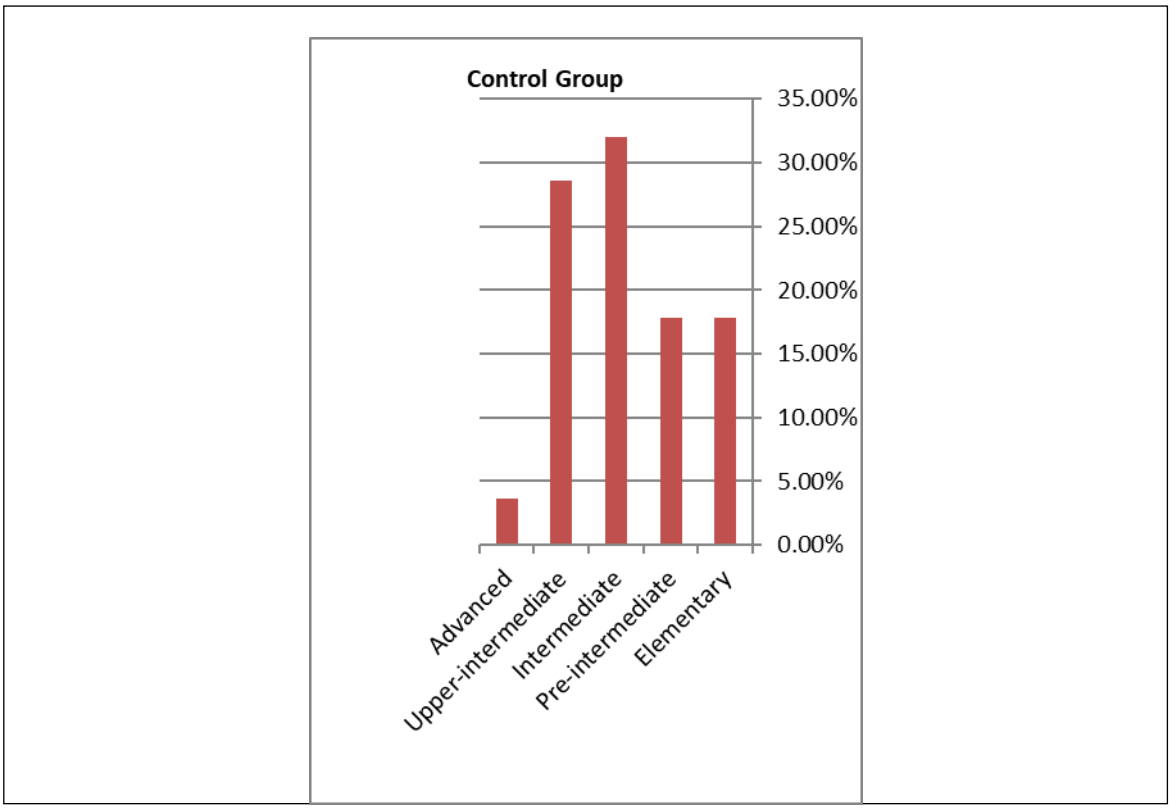


Figure 4.1: Learners Proficiency Categorization According to CPT Scores

4.1.2 Vocabulary Achievement Pre-Test

An achievement pre-test has been given to students of both groups, experimental and control groups, as a second tool to test the comparability of both groups besides being a baseline for measuring the resulting gains after going through the treatment. For the experimental group, the students achieved the mean score of 23.48 out of 50 whereas the control one achieved the mean score of 20.14 out of 50. Though they are not identical means, it has been proved to be NOT statistically significant through running independent samples T-Test. As It's shown in table (4.3) that P value is 0.17 which is greater than the significance level ≤ 0.05. (See Tables 4.3 and 4.4)

Table 4.3: Vocabulary Pre-Test Scores Descriptive Statistics

Group Statistics for Pre-test Scores					
TEST	**group**	**N**	**Mean**	**Std. Deviation**	**Std. Error Mean**
PRETEST	Experimental	25	23.48	8.377	1.675
	Control	28	20.14	9.272	1.752

Table 4.4: T-Test for Vocabulary Pre-Test Scores for Both Groups

Independent Samples Test for Pre-Test Scores for Both Groups (Experimental. and Control)										
Parameter / Test		***Levene's Test for Equality of Variances***		***t-test for Equality of Means***						
		F	**Sig.**	**t**	**df**	**Sig. (2-tailed)**	**Mean Difference**	**Std. Error Difference**	**95% Confidence Interval of the Difference**	
									Lower	Upper
PRETEST	Equal variances assumed	.270	.606	1.368	51	.177	3.337	2.439	-1.559-	8.233
	Equal variances not assumed			1.376	50.990	.175	3.337	2.424	-1.530-	8.204

4.2 QUESTION ONE: What is the effect of using corpus-based approach on vocabulary learning in Saudi EFL classroom?

The first question is considered the essential and main research question as it aims to compare the effect size of applying corpus-based approach to conventional teaching or learning approach in vocabulary classroom. The researcher hypothesized that there were statistically significant differences between experimental group's students mean scores in vocabulary posttests and their counterparts mean scores in the control group in favor of the experimental group by the virtue of corpus employing. To test this hypothesis, pre and posttests have been applied on both groups following a specific timeline as mentioned in the previous chapter. A set of steps will be done to thoroughly answer this question;

- First: comparing the mean scores of pretests between the groups to prove that they are comparable; this step has already been done and validated in the previous paragraphs.

- Second: measuring the differences between pre sand post-tests scores in each group separately to identify the gains of each individual group after going through the 6-week treatment.
- Third: measuring the differences between the mean scores of immediate posttests of both groups to identify the effect size of each teaching or learning approach (conventional approach and corpus-based approach) and test the effectiveness of corpus-based on students' vocabulary learning.

According to the means of the pre-test results, both groups performed at a very similar rate (Experimental pretest mean score = 23.48, Control pretest mean score= 20.14, and *P value=0.177*). These results were meaningful, as they revealed that there was no significant discrepancy between the groups pre-experimentally in terms of their pre-learning backgrounds. After confirming and proving the comparability of the two assigned groups and setting an equal baseline for any further measurements of vocabulary gains, the researcher has measured the gains of each individual group by analyzing the immediate posttests scores. Data collected from this test have been analyzed by running One Paired Sample T-Test on pre and immediate post-tests scores of each group separately. For the control group, Table 4.5 reveals that the pre-test mean score is 20.14 whereas the posttest mean score is 32.21 which indicates and average of 12.07 points of gains. To test the significance of these gains One Paired Samples T-test has been run and revealed that *P =0.00* which indicated a statistically significant difference at the level of ($P \leq 0.05$) between pre and post-tests scores (See Table 4.6).

Table 4.5: Control Group's Pre and Post Tests Scores Descriptive Analysis

TEST		Mean	Mean Gains	N	Std. Deviation	Std. Error Mean
Control Group	PRETEST	20.14	**12.07**	28	8.936	1.752
	POSTTEST	32.21		28	9.272	1.601

Table 4.6: Paired Samples T-Test for Control Group's Pre and Post Tests Scores

Parameters/ Test		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	**PRETEST - POSTTEST**	-14.302	5.128	.704	-15.715	-12.889	-20.306	27	.000

Similarly, For the experimental group, descriptive statistics in Table 4.7 reveals that the pre-test mean score is 23.48 whereas the posttest mean score is 40.28 which indicates and average of 16.8 points of gains. To test the significance of these gains One Paired Samples T-test has been run on pre and post -tests scores and revealed that $P = 0.00$ which indicated a statistically significant difference at the level of ($P \leq 0.05$) between pre and post-tests scores (See Table 4.8).

Table 4.7: Experimental Group's Pre and Post Tests Scores Descriptive Analysis

TEST		Mean	Mean Gains	N	Std. Deviation	Std. Error Mean
Experimental Group	PRETEST	23.48	**16.8**	25	8.377	1.675
	POSTTEST	40.28		25	6.374	1.275

Table 4.8: Paired Samples T-Test for Experimental Group's Pre and Post Tests Scores

Parameters/ Test		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 2	**PRETEST - POSTTEST**	-14.302	5.128	.704	-15.715	-12.889	-20.306	24	.000

To sum, these results proved that vocabulary learning gains of both teaching methods, corpus-based and conventional approaches, were significant and clear in learners' immediate posttests scores. Regarding the last and the main step of this analysis, which is

measuring the differences in gains between the two groups and identifying the effectiveness of using corpus-based approach in vocabulary learning with the experimental group learners. As it can be seen in table 4.9, the mean scores of immediate posttests are 40.28 and 32.21 for the experimental group and the control group respectively whereas the mean gains scores are 16.8 and 12.07 for the Experimental and the control groups respectively. Although these mean posttest scores and gains reveal differences of vocabulary learning gains in favor of the experimental group (corpus-based approach), we still need further variances analysis to identify their statistical significance. For this reason, the researcher has run an Independent Samples T-test on immediate posttests scores of both groups to figure out the significance of these differences at the level ($P = 0.05$). As it is shown in Table 4.10 below, the P value is 0.000 which indicates a high level of significance for the differences in vocabulary learning gains in favor of the experimental group (corpus-based approach). This result has answered the first and main question and met the first hypothesis of the study which stated *"there are statistically significant differences in vocabulary learning between the control group which received conventional vocabulary learning and the experimental group which received corpus-based vocabulary learning in favor of the experimental group."*

Table 4.9: Experimental and Control Groups Posttests Scores Descriptive Statistics Analysis

Tests	group	N	Mean	Std. Deviation	Std. Error Mean
POSTTEST	Experimental	25	40.28	6.374	1.275
	Control	28	32.21	8.469	1.601

Table 4.10: Experimental and Control Groups Posttests Scores Descriptive Statistics Analysis

Independent Samples Test										
TEST / PARAMETER		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
POSTTEST	Equal variances assumed	1.219	.275	3.879	51	.000	8.066	2.079	3.892	12.240
	Equal variances not assumed			3.942	49.645	.000	8.066	2.046	3.955	12.176

4.3 QUESTION TWO: How far is corpus-based approach effective in improving vocabulary retention in Saudi EFL classroom?

In answering the second question of the study the researcher has hypothesized that there are statistically significant differences, at the level $P \leq 0.05$, between the mean scores of delayed post-tests of control group that received conventional vocabulary instruction and those of the experimental group that received corpus-based instruction in favor of the experimental group. To test this hypothesis the researcher has analyzed the collected data of delayed posttests scores (that measured vocabulary retention of learners) for both groups. As it's shown in Table 4.11, the mean scores of delayed posttests are 32.82 and 41.52 for the control and experimental groups respectively with a standard deviation 10.883 for the control group and 10.564 for the experimental one. Even though there are clear differences in mean scores of delayed posttests between both group in favor of the experimental group, another variance analysis, Independent Sample T-test, has been run to prove this claim. Table 4.12 shows that *P value = 0.005* which indicates that the differences in mean scores

of delayed posttests are statistically significant in favor of the experimental group. Therefore, the second hypothesis proposed by the researcher has been met and proved that suggested the effectiveness of corpus-based approach over the conventional approach in enhancing vocabulary retention rate.

Table 4.11: Descriptive Statistics Analysis of Delayed Posttest Scores for Both Groups

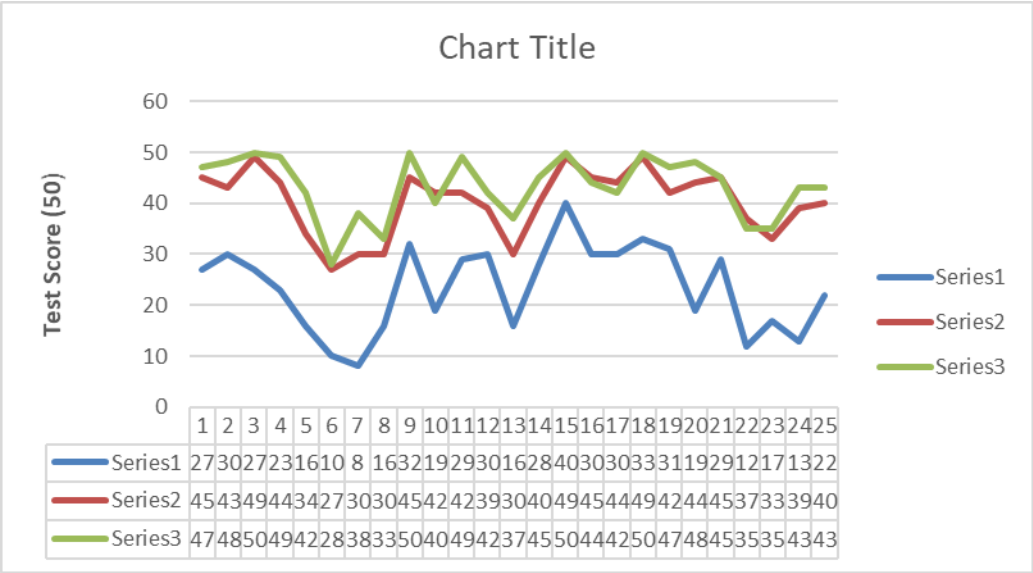
Tests	Group	N	Mean	Std. Deviation	Std. Error Mean
DELAYED POSTTEST	Experimental	25	41.52	10.564	2.113
	Control	28	32.82	10.883	2.057

Table 4.12: Comparing Variances of Delayed Posttest Mean Scores (T-Test)

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
TEST / PARAMETER		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
DELAYED POSTTEST	Equal variances assumed	.186	.668	2.945	51	.005	8.699	2.954	2.769	14.628
	Equal variances not assumed			2.950	50.627	.005	8.699	2.949	2.778	14.619

4.3.1 Learners' Scores in Definitional and Transfer Parts of Posttest

The researcher has tracked the overall learning development of learners of both groups throughout the various stages of the 6-week treatment through analyzing and comparing the means of tests scores as illustrated in Figures 4.2 and 4.3. Generally speaking, after going through the treatment most learners of both groups have shown progress in learning the target vocabulary in each chapter (approximately 118 words). However, these gains have been proved to be different in size in favor of the corpus-based group.



Note: Blue Line Refers to Pre-test Scores – Red Line Refers to Posttest Scores – Green Line Refers to Delayed Posttest Scores

Figure 4.2: Tracking Experimental Group's Learners progress

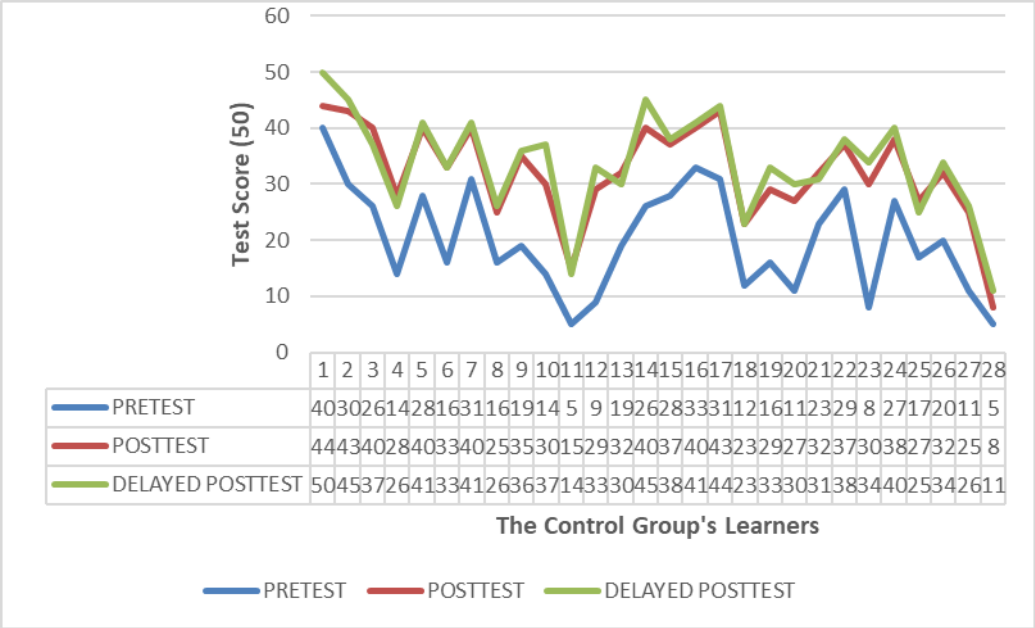


Figure 4.3: Tracking Control Group's Learners progress

The researcher has made further analysis on learners' scores of the two main parts of the post-test (Definitional Part and Transfer Part) to pinpoint strengths and weaknesses in

each learning approach (conventional vs corpus-based). As it's shown in Figures 4.4 and 4.5, learners of both groups have relatively high scores in the definitional part of the posttest while almost all of them have less scores in the transfer part which required using the target words and their clusters in novel contexts based mainly on authentic concordance lines. The results reveal slight differences in the scores of the definitional part (receptive knowledge) between learners of both groups whereas there are big differences in the transfer part (productive knowledge) in which learners use the target vocabulary or collocations in new contexts. The results revealed that the mean scores for the definitional part are 19 and 22.08 out of 25 for control and experimental groups respectively while the mean scores for the transfer part are 12.1 and 16.36 for control and experimental groups respectively. It can be concluded that the great effect size of using corpus-based approach in vocabulary learning is essentially in enhancing transfer or productive knowledge of vocabulary. This drawn conclusion has high consistency with a great bulk of corpora and DDL approach literature (Johns, 1991; Sinclair, 1992, Davies, 2008, Romer, 2010, et al).

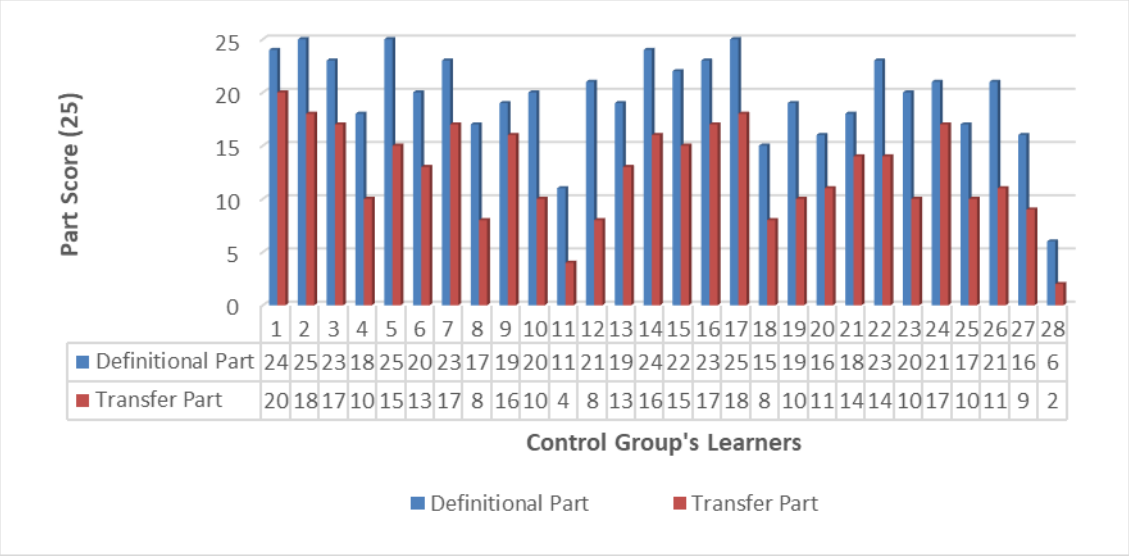


Figure 4.4: Comparing the scores of definitional and Transfer parts (Control Group)

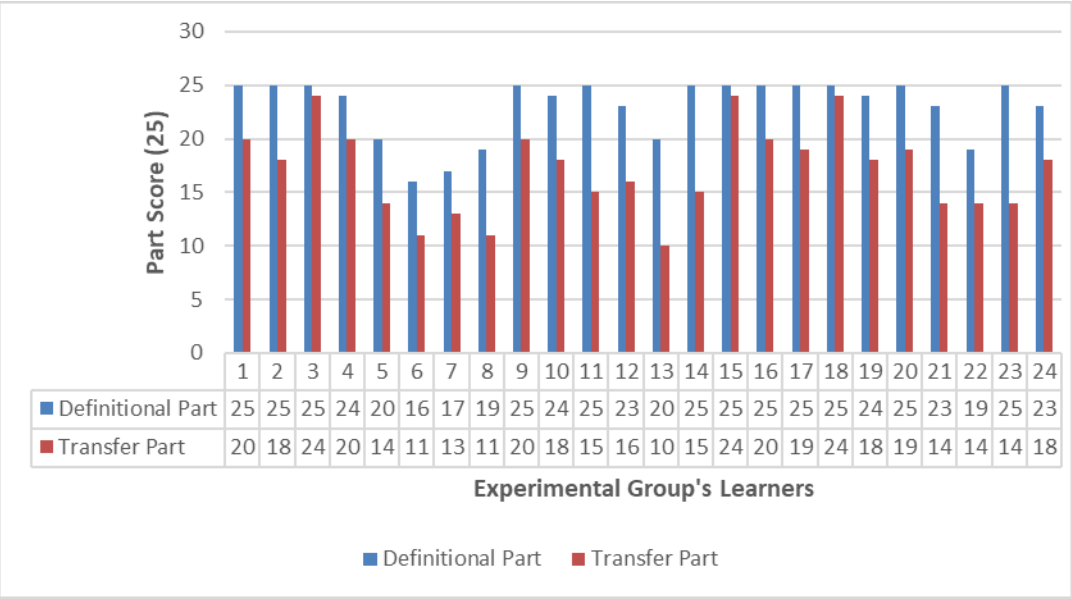


Figure 4.5: Comparing the scores of definitional and Transfer parts (Experimental Group)

To recap, the above shown results confirmed and proved that learners of the experimental group who received corpus-based instruction and learning activities outperformed their counterparts in the control group in vocabulary learning and retention rates. These results are aligned with many other recent studies (Alruwaili, 2018; Elsherbini, S., and Ali, A., 2017; Alharbi. R., 2017; Almutairi, N., 2016; Almahbashi, A. et al, 2015; Marza, 2014; Varley, 2008; and others) that proved the effectiveness of integrating corpora in enhancing vocabulary learning and retention.

4.4 QUESTION THREE: What is the effect size of learner's proficiency level (as a moderator variable) in the effectiveness of using corpus-based approach in vocabulary learning?

In answering the third question of the study the researcher has hypothesized that There are NO statistically significant differences, at the level $P \leq 0.05$, between the mean scores of delayed post-tests for learners of different proficiency levels in the experimental group that received corpus-based learning treatment. He has assumed that the effect size of proficiency level as a moderator variable alongside with using corpus-based approach as the main independent variable on learners' vocabulary posttest scores as a dependent

variable is null. To test this NULL hypothesis the researcher has run a TWO WAYS ANOVA test on experimental group's learners scores of the immediate post-test and pre-test to find out the significance of variances among the four proficiency levels (sub-groups) in the experimental group.

Data used to answer this question has come from the experimental group's students' scores of pretest and immediate post-test. As they have taken a pre-test and an immediate post-test based mainly on the target 118 words and their collocations of a total score of 50. Students responses have been cross-checked by another teacher at the Colleges, BPC, to ensure the validity and accuracy of ratings. Furthermore, both tests were set identical, except the arrangement of test items to exclude the shadowing effect. After getting the scores of both tests, the researcher has run Two-Way-Repeated-Measures-Anova Test in the SPSS program to identify whether there were improvements in the results of the four subgroups: upper-intermediate, intermediate, pre-intermediate, and elementary and if so, which subgroup has made the most improvement and which the least.

As it's shown in Figures 4.6, the four sub-groups (proficiency levels) have revealed some range of improvement regarding vocabulary learning gains which can be seen from comparing the scores of their pre-tests and immediate post-tests. To check the effect size in each subgroup the researcher has subtracted the mean scores of pre-tests from the mean score of post-tests for each subgroup. As it can be seen in Table 4.13, the mean scores of immediate posttests are 30.8, 38.2, 41.6, and 45.44 for elementary, pre-intermediate, intermediate, and upper-intermediate subgroups respectively. To be more specific, the mean gains have been calculated for each subgroup as it's shown in the same table 4.13; 18.4, 17.4, 15.4, and 16.44 for elementary, pre-intermediate, intermediate, and upper-intermediate subgroups respectively. These results indicate that the highest gains are in favor of the elementary level subgroup with the highest mean gains (18.4) followed by pre-intermediate level sub-group (17.4) and upper-intermediate (16.44) while interestingly the intermediate level subgroup came at the late with the least mean gains (15.4).

Table 4.13: Summary of Vocabulary Gains for the Four Proficiency Sub-groups

Sub-Groups	N	Pre-test Scores	Post-test scores	Mean Gains	St. Deviation
Upper-intermediate	10	29	45.44	16.44	2.716
Intermediate	5	26.2	41.6	15.4	2.074
Pre-intermediate	5	20.8	38.2	17.4	4.868
Elementary	5	12.4	30.8	18.4	3.701
Total	25	23.48	40.28	16.8	6.374

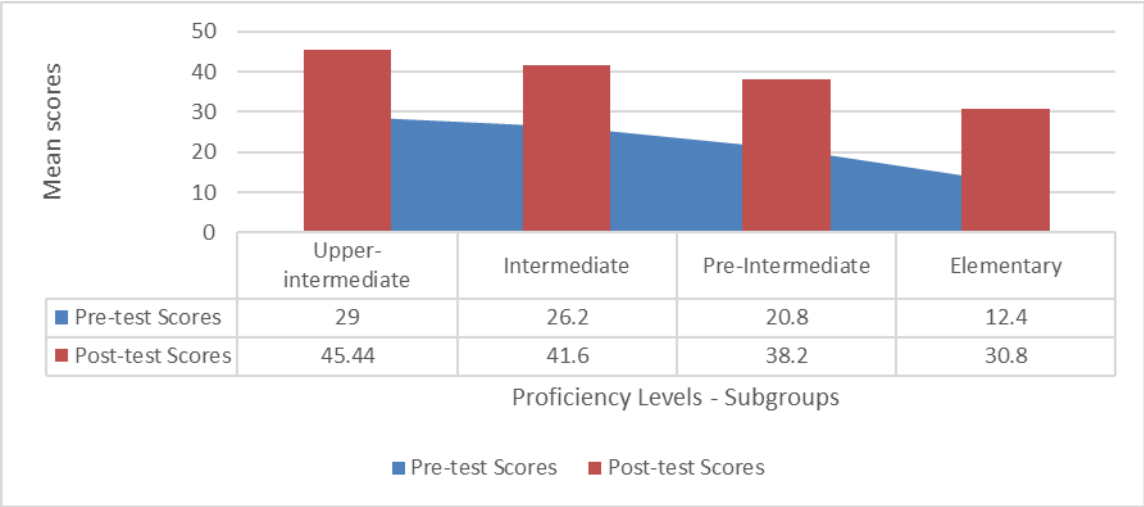


Figure 4.6: Vocabulary Gains for the Four Sub-groups

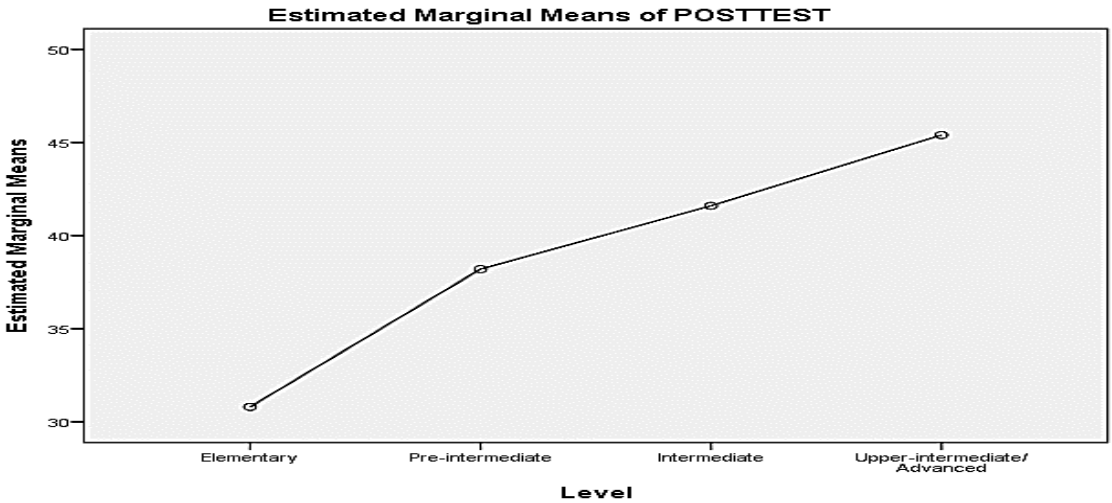


Figure 4.7: Mean Scores of Posttest for the Four Sub-groups

To check the significance of differences between the four subgroups in their mean scores of immediate post-tests, the researcher has run a Multivariate Analysis test in SPSS program. As it is shown in Table 4.14, there are statistically significant variances in subgroups mean gains at the level $P \leq 0.05$ as the P value is .000 and 001. The researcher concluded from these results that the null hypothesis is rejected in this question as there are statistically significant differences at the level $P \leq 0.05$ between the subgroups vocabulary learning gains based on language proficiency levels. Thus, we cannot exclude the effect of learners' proficiency level as a moderator variable while using corpus-based approach in learning vocabulary.

Table 4.14: Multivariate Analysis on Pre- and Post- tests Scores

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.993	1522.199[b]	2.000	20.000	.000	.993
	Wilks' Lambda	.007	1522.199[b]	2.000	20.000	.000	.993
	Hotelling's Trace	152.220	1522.199[b]	2.000	20.000	.000	.993
	Roy's Largest Root	152.220	1522.199[b]	2.000	20.000	.000	.993
Level	Pillai's Trace	.786	4.534	6.000	42.000	.001	.393
	Wilks' Lambda	.226	7.362[b]	6.000	40.000	.000	.525
	Hotelling's Trace	3.375	10.688	6.000	38.000	.000	.628
	Roy's Largest Root	3.359	23.515[c]	3.000	21.000	.000	.771

a. Design: Intercept + Level
b. Exact statistic
c. The statistic is an upper bound on F that yields a lower bound on the significance level.

Table 4.15: Comparison of Posttest Mean Scores among the Four Sub-groups

Multiple Comparisons

Dependent Variable	(I) Level	(J) Level	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
POSTTEST	Elementary	Pre-intermediate	-7.40*	2.108	.002	-11.78	-3.02
		Intermediate	-10.80*	2.108	.000	-15.18	-6.42
		Upper-intermediate/ Advanced	-14.60*	1.825	.000	-18.40	-10.80
	Pre-intermediate	Elementary	7.40*	2.108	.002	3.02	11.78
		Intermediate	-3.40	2.108	.122	-7.78	.98

		Upper-intermediate/ Advanced	-7.20*	1.825	.001	-11.00	-3.40
	Intermediate	Elementary	10.80*	2.108	.000	6.42	15.18
		Pre-intermediate	3.40	2.108	.122	-.98	7.78
		Upper-intermediate/ Advanced	-3.80*	1.825	.050	-7.60	.00
	Upper-intermediate/ Advanced	Elementary	14.60*	1.825	.000	10.80	18.40
		Pre-intermediate	7.20*	1.825	.001	3.40	11.00
		Intermediate	3.80*	1.825	.050	.00	7.60

Based on observed means.

The error term is Mean Square (Error) = 80.157.

*. The mean difference is significant at the .05 level.

As it has been explained above, the variances of vocabulary gains are statistically significant (at the level p ≤0.05) among the four sub-groups with the highest gains for elementary group and the lowest interestingly for intermediate. However, to get deeper interpretation for these differences, the researcher decided to thoroughly investigate the four subgroups scores and gains at an individual level and measure each student contribution in its sub-group's overall gains (See Table 4.15 Summary of students gains).

Table 4.16: Summary of Students Gains Arranged by Proficiency Sub-groups

PROFICIENCY SUB-GROUPS	STUDENTS	PRETEST	POSTTEST	GAINS
Elementary	X1	10	27	17
	X2	8	30	22
	X3	16	30	14
	X4	16	30	14
	X5	12	37	25
	Total 5	***62***	***154***	***92***
Pre-intermediate	X6	16	34	18
	X7	28	40	12
	X8	30	45	15
	X9	17	33	16
	X10	13	39	26
	Total 5	***104***	***191***	***87***
Intermediate	X11	30	43	13
	X12	19	42	23
	X13	30	39	9
	X14	30	44	14

	X15		22	40	18
	Total	***5***	***131***	***208***	***77***
Upper-intermediate	X16		27	45	18
	X17		27	49	22
	X18		23	44	21
	X19		32	45	13
	X20		29	42	13
	X21		40	49	9
	X22		33	49	16
	X23		31	42	11
	X24		19	44	25
	X25		29	45	16
	Total	10	290	454	164
Total	**N**	**25**	**587**	**827**	**420**

It has been proven that the upper-intermediate sub-group has outperformed by far the other three subgroups (elementary, pre-intermediate, and intermediate) in the scores of immediate post-tests by a mean score 45.44 out of 50. However, this outperformed subgroup (upper-intermediate) did not achieve the highest learning gains which measured by subtracting pre-test scores from post-test scores. This can be partially interpreted in the light of the idea of "ceiling level" which refers to the narrow range between the initial score and the end score; as the upper-intermediate students have already had quite high scores in the pre-test and thus they have a very little space to achieve the full mark. To be more specific, there are some students in the upper-intermediate sub-group achieved quite high pre-test scores; such as X21 earned 40 points (out of 50), X22 earned 33 points and X19 earned 32 points. They were quite close to the highest score they could achieve, and their scores were already at a high level. This phenomenon, ceiling level, thus can explain the low level of gains in the side of upper-intermediate sub-group which considered the superior one regarding posttests scores.

On the other hand, the elementary sub-group which came last in posttest scores (mean score 30.8), has achieved the highest level of gains. This also can be attributed to the aforementioned phenomenon "ceiling effect" as most its members achieved very low scores in their pretests which means a very long range of gains to reach the end score. For example, the students X1, X2, and X5 achieved in the pre-tests 17, 22, and 25 points respectively, whereas they achieved 27, 30, and 37 points in posttest. This means that they had jumped to

much higher results but they are still the lowest level except for student (X5) who achieved a relatively high score (37 out of 50).

Interestingly, the intermediate sub-group, the middle group of the four sub-groups, did not make high gains (total gains 77) compared to elementary and pre-intermediate sub-groups which made 92 and 87 total gains respectively. It can be partially explained by the relatively high pretest scores of 60% of students (i.e. students X11, X13, and X14 have achieved 30 points for each) and only 40% of students have achieved high gains (i.e. X12 and X15 have achieved 32 and 18 respectively). Eventually it made the overall gains of this sub-group the lowest among others.

4.5 Summary of Findings for Pre and post-tests Scores

According to the previous detailed results of pre and post-tests scores of all participants, it can be concluded that 100% of participants have achieved some improvement in vocabulary learning whether they studied in experimental group (corpus-based approach) or control group (conventional approach). However, there are statistically significant differences between the control group immediate post-test and delayed post-test scores and those of the experimental group in favor of the experimental. Thus, using corpus-based approach has been proved to have a positive impact on learners' vocabulary learning and retention. In addition, there are statistically significant differences at the level $p \leq 0.05$ among experimental group's learners' vocabulary gains depending on their language proficiency level. The highest level of gains had achieved by the elementary sub-group while the least achieved by the intermediate one. Despite the upper-intermediate sub-group have by far outperformed the other three groups in posttest scores, they have achieved low gains compared to elementary and pre-intermediate sub-groups. This has been attributed to the "ceiling effect" phenomenon which refers to the range between pre-test score and the end score. This range is very big in the side of elementary sub-group and very small in the upper-intermediate one.

To recap, corpus-based approach has been confirmed to be effective approach for teaching students of all proficiency levels with more benefits to higher levels. Second,

learners' proficiency level is concluded to be a moderator variable alongside with using corpus-based approach affect vocabulary learning gains and retention rates.

4.6 QUESTION FOUR: What are the attitudes of the experimental group students towards using corpus-based approach in vocabulary learning classes?

This question is an attempt to investigate learners' attitudes towards employing corpus-based approach in their vocabulary classes as a tool for learning and looking up target words rather than using conventional dictionaries. The researcher has hypothesized that the attitudes of the experimental group learners towards using corpus-based approach in vocabulary classrooms are relatively positive. To test this hypothesis, the researcher has used a questionnaire at the end of the treatment (in the ninth week) to collect data about their overall views on corpora use and faced challenges. The questionnaire has been divided into three main themes; the first theme included 4 questions about participants' demographic data which has been administered at an early stage of the treatment, the second theme consisted of 7 questions which essentially surveyed their overall evaluation of COCA corpus integration benefits, the third theme included 5 questions about difficulties and challenges of using COCA corpus use. Items from 1 to 4 were used early to collect students' demographic data like technological skills, language study period, availability of a personal computer and internet connection. Items from 5 to 11 were 5-point Likert scale questions used to survey their perceptions on corpora benefits and its overall effectiveness while items 12 to 16 were used to investigate challenges and difficulties.

The questionnaire data has been analyzed using descriptive statistics of SPSS program. Data analysis will be presented under the three themes of the questionnaire. The total number of students who responded to all the questions of the questionnaires is 24 out of 25 students in the experimental group. The findings for the three themes will be presented in the following paragraphs.

4.6.1 Participants' Demographic Data

This part of the questionnaire was administered in the first week of the study to collect participants' personal data and have a close understanding for their language and technology backgrounds. The main findings for this part can be summarized in the following pie charts in Figure 4.8.

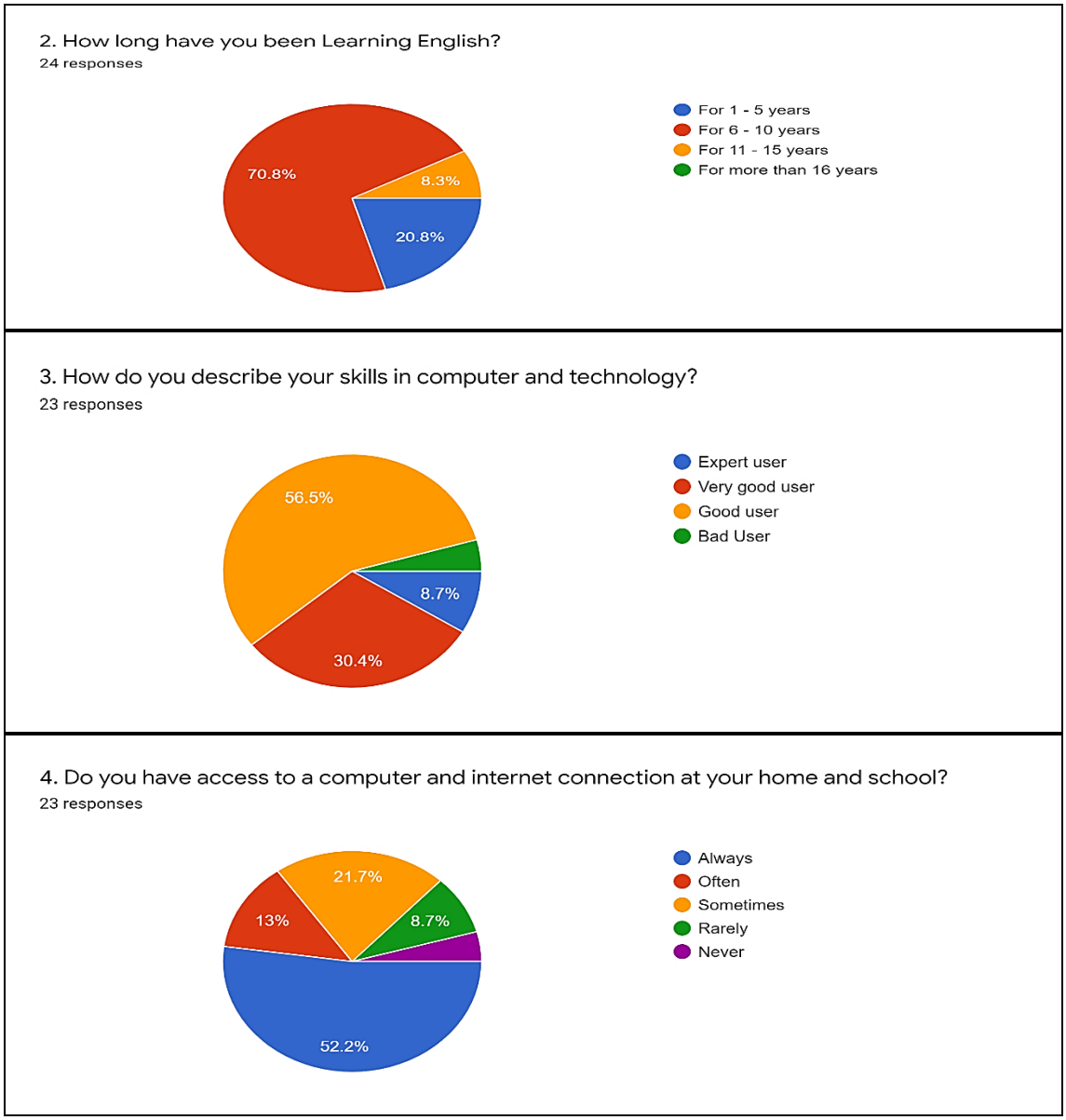


Figure 4.8: Descriptive Statistics for Participants' Language and Technology Backgrounds

For details, 70.8% and 20.8% studied English for time periods of (6-10 years) and (11- 16) respectively in their pre-university stages. Moreover, 56.5% and 30.4% were very good and good technology users respectively. Lastly, 52.2% and 13% of participants were always or often having access to a computer and internet connection while only 3 students out of 24 (12.6%) were rarely or never having available computers and internet connection. It can be concluded from the results presented above that most of participants have solid and reliable language and technology backgrounds that qualify them to be part of the experiment which requested using computers and online language corpora in some tasks and activities in their EFL classroom.

4.6.2 Students' Perceptions on Corpus-based Approach Benefits and Feasibility

The following table is presenting an overall overview for the three main themes of the questionnaire, overall perceptions, benefits, and difficulties. According to table 14.16, the mean score for participants' responses for items of the theme of corpora use benefits is 4.072 and for the overall perceptions is 3.081 which indicates high positive attitudes towards corpus-based approach. On the other hand, the mean score for their responses to the theme of difficulties and challenges is 3.613 which indicates some potential difficulties and problems in using corpora. Though the mean response to problems and challenges themes is relatively high (3.61), most participants believed that by the course of time they became successful in dealing with these challenges and got familiarized with the massive provided concordances data. This notion has been proved through analyzing their responses to question number 13 that states *"I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily".* The mean response for this question is 3.66 as 53.9% of participants agree and strongly agree that they have already overcome potential challenges over the course of time and 42.3% are still neutral and only 3.4% disagree.

Generally speaking, it can be obviously concluded that learners' attitudes towards employing corpus-based approach in vocabulary learning classes is positive and agreeable.

The means scores for each question will be discussed and analyzed according to the three main themes in the following paragraphs.

Table 4.17: Summary of Descriptive Statistics for the Main Themes

Themes	N. of Items	Minimum	Maximum	Mean	Std. Deviation
Benefits of COCA Corpus Use	4	2	5	4.072	0.915809
Overall Perceptions on Corpora Use	4	2	5	3.081	0.891025
Problems and Challenges of Corpus-based Approach.	4	1	5	3.613	0.900717
Total Valid Responses	Items = 12 items - N. of Participants = 24 out of 25				

The first theme of the questionnaire was about the potential benefits of employing corpus-based approach; this theme included the items 6, 7, 8, and 10. The following table (Table 4.17) shows the mean scores for participants response for each item. It can be seen that most students' responses to the benefits of corpora items are above the middle point, the point 3, in a five-point Likert scale of ratings. Thus, most of them realized an believed in the effectiveness of COCA corpus activities and tasks in their EFL classroom. For more details, the first item, item 6, 73.1% of the participants agreed on the effectiveness of corpora activities and printouts over the conventional dictionaries, while only 3.8% disagreed and 23.1% had neutral opinions. For the second item (item 7), 73.1% agreed and highly appreciate the authentic input of corpora in learning various linguistic aspect of words but only 7.7% disagreed and 19.2% had neutral attitudes. The third item, item 8, received 65.4% positive responses and only 7.7% negative responses and the other 26.9% participants were neutral. For the last item in this theme, item no. 10, 69.3% of respondents agreed on the importance of corpora in providing authentic contexts and co-occurrences of words while only 7.6% disagreed and 23.1% were neutral.

Overall, 70.23% of participants responded positively to the proposed benefits and uses of concordances lines and corpus-based activities while only 6.75% of them responded negatively and the remaining participants, about 23.08%, responded neutrally.

Table 4.18: Summary of Descriptive Statistics for the Items of the Benefits Theme

Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
6. I feel that using corpus-based activities /COCA based activities and printouts in learning vocabulary is more helpful and beneficial than using conventional methods like dictionaries and coursebooks reading texts?	24	0%	3.8%	23.1%	26.9%	46.2%	4.16	.916
7. I feel that using Corpora and being exposed to massive authentic input contributes positively in learning many aspects and linguistic elements of a word?	24	0%	7.7%	19.2%	26.9%	46.2%	4.16	1.01
8. I feel that being involved in corpora individual activities and group tasks increases my vocabulary retention?	24	0%	7.7%	26.9%	38.5%	26.9%	3.916	.880
10. After going through this study, I realized the importance of context and co-occurrences provided by the corpora queries to effectively learn new words, rather than using dictionaries.	24	3.8%	3.8%	23.1%	38.5%	30.8%	4.041	.858
Overall Evaluation for the Theme	24	1%	5.75%	23.08%	32.70%	37.53%	4.072	0.915809

Moving to the second theme which was essentially about participants' overall satisfaction about using corpus-based approach in learning vocabulary; this theme included the items 5, 9, 11, and 13. As it's shown in (Table 4.18) that the average score of participants responses for all items of this theme is 3.83 which is above the middle point (point 3) in the 5-point Likert scale of rating. It can be seen that most students' responses to the benefits of corpora items are above the middle point, the point 3, in a five-point Likert scale of ratings. Therefore, most of the respondents generally believed in the effectiveness of corpus-based approach in vocabulary learning and retention. For more details, the first item, item 5, 69.3% of the participants agreed on COCA corpus as a helpful tool in their classes whereas the percentage of disapproval was only 11.55 and the remaining portion, 19.2%, were neutral. The second item in this theme, item no. 9, was about being comfortable while being involved in the corpus-based group. 92.3% of participants showed 3-point and above satisfaction level in the 5-point scale while only 7.7% expressed their dissatisfaction. In the case of the third item, item no. 11, they showed their willingness to use corpora at their own in the future in a relatively high rate, about 73.1% of approval and only 7.7% of disapproval. For the last item of this theme, item no. 13, 64.45% of

participants showed their ability to come over some challenges and problems over the course of time and they got familiarized to the new approach.

To recap, this theme had four items that measured general satisfaction of participants about using corpus-based approach. All the items received high levels of approvals, about 70%, and very low levels of disapproval, about 7.75% and the remaining part, 27.88%, came in the middle (at the point 3 out of the 5-point scale). These results proved that there is high consistency between participants responses for the items of the two themes, the theme of corpora benefits and the theme of the overall satisfaction about corpus-based approach.

Table 4.19: Summary of Descriptive Statistics for the Items of Overall Perceptions Theme

Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
5. I feel that the use of corpora (COCA Corpus) was helpful for me in learning vocabulary in the course ENG.140?	24	0%	11.5%	19.2%	46.2%	23.1%	3.83	.936
9. I generally feel comfortable to use concordances lines to learn new words and to be involved in the experimental group for a 5-weeks period?	24	0%	7.7%	30.8%	50%	11.5%	3.70	.750
11. I think that I will use Corpora or COCA Corpus at my own in the future to learn new words or look up words?	24	7.7%	0%	19.2%	26.9%	46.2%	4.125	1.034
13. I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily.	24	0%	3.8%	42.3%	38.5%	15.4%	3.66	.8164
Overall Evaluation for the Theme	24	2%	5.75%	27.88%	40.40%	24.05%	3.833	0.8912

4.6.3 Potential Challenges in Using Corpus-based approach

Moving to the third and last theme that has essentially devoted to participants' challenges and difficulties throughout their experience with corpus-based activities and tasks in vocabulary classrooms. This theme has included the items 12, 14, 15, and 16. It can be seen in the following table (Table 4.19) that the average score of participants responses for all items of this theme is 3.61 which is above the middle point (point 3) in the

5-point Likert scale of rating. Generally, this result indicated that corpus-based approach was perceived to some extent as a quite challenging method by the participants of the experimental group as the majority of them (58.93%) responded positively to the challenges of corpora use while only 5.9% didn't perceive these challenges.

The potential four challenges that have been surveyed were the massive amount of data provided, the cut-off sentences of concordances lines, the unknown vocabulary in queries resulted data, and the language proficiency level of the participant. All these items have been rated positively by the respondents which indicated that there is some sort of challenge while using this approach. However, their highly positive responses to the item 13 that stated *" I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily"* could interpret this discrepancy between participants' high level of approval for using corpus-based approach and concurrently their high level of approval for perceiving challenges and difficulties in this approach. From this item which has been averagely rated at 3.66 level, it can be inferred that though the participants encountered some challenges at the beginning, they gradually improved and got familiarized to the use of concordances lines, cut-off sentences and selecting relevant lines from the available massive provided ones.

For more details, the first item, item 12, 57.7% of the participants agreed that the big amount of concordances lines provided could be overwhelming and cause frustration while 34.6% remained neutral and only 7.6% disagreed. For the second item (item 14), 72% of participants agreed on the importance of language proficiency as an essential factor for benefitting from this approach while only 8% disagreed and the remaining participants (20%) were neutral. The third item, item 15, has been rated as follows: 52% approval, 48% neutral and surprisingly 0% disapproval. These results highlighted the cut-off sentences in some concordance lines as a high challenging factor in employing corpora in the classroom. For the last item in this theme, item no. 16, 52% of respondents agreed on the difficulty of some unfamiliar vocabulary while learning new vocabulary through concordance lines.

To recap, this theme had four items that measured potential challenges and difficulties that could face participants of the experimental group throughout applying corpora activities and tasks in EFL classrooms Most of the items received average levels of approvals, the average score was 3.61 at a five-point scale of rating. This result could be seen contradicted to their relatively high approvals on the benefits of corpora and its overall effectiveness as a tool for leaning vocabulary. However, the researcher reached a compromise by realizing that though they appreciate COCA corpus as a valuable tool for learning vocabulary, they also encountered some challenges in the application stage especially at the early beginning as they were not familiarized with this brand-new tool. This idea could be statistically proved through their responses to question no. 13 *" I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily"* as only one participant disagreed while all the 23 students agreed with different levels of approval. Thus, the researcher's fourth hypothesis has been met and proved.

Table 4.20: Summary of Descriptive Statistics for the Items of the Challenges Theme

Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
12. I feel that the queries concordance lines or context examples are overwhelming and may cause frustration or distraction especially.	24	3.8%	3.8%	34.6%	50%	7.7%	3.583	.8805
14. I feel that my language proficiency level impacts my ability to effectively use corpus-based printouts and activities.	24	0%	8%	20%	52%	20%	3.791	.9315
15. I have some difficulty in using the COCA corpus due to cut-off sentences in some concordance outputs.	24	0%	0%	48%	36%	16%	3.66	.8164
16. I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance outputs.	24	0%	8%	40%	36%	16%	3.417	.9743
Overall Evaluation for the Theme	24	0.95%	4.95%	35.65%	44%	14.93%	3.61275	0.900675

4.7 QUESTION FIVE: To what extent are EFL teachers in Saudi Arabia aware about using corpus-based approach in EFL classrooms?

This research question is intended to survey a randomly chosen sample of EFL university teachers in Qassim Region, in Saudi Arabia. The researcher has decided to investigate teachers' awareness and perceptions towards corpus-based approach believing in the central and essential role of teachers in initiating and adopting brand-new approaches in their classrooms. Therefore, the researcher believed that a thorough and authentic investigation of the efficacy of an instructional approach have to include teachers' perceptions and attitudes into investigation to get a comprehensive viewpoint and applicable implications. For this question, the researcher, depending on his long experience in language teaching in different colleges in Saudi Arabia, has hypothesized that Saudi university EFL teachers have relatively low level of awareness towards using corpus-based approach in EFL classroom.

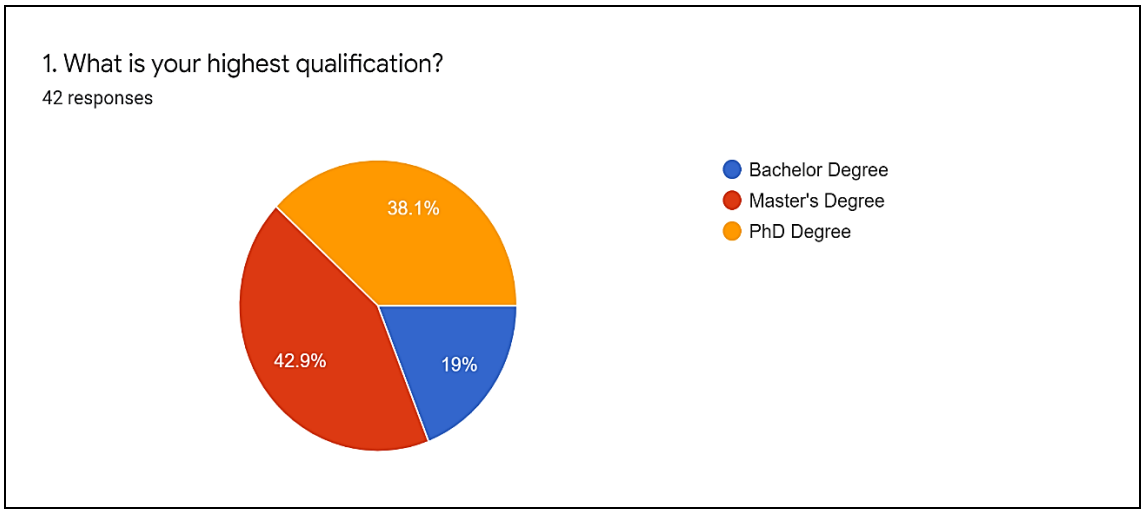
To test this hypothesis, the researcher has administered a questionnaire at the end of the treatment (in the ninth week) to collect data about teachers' awareness on and overall attitudes towards corpus-based approach. The questionnaire has been divided into three main themes; the first theme included 4 questions about participants' demographic data which has been focused on teaching experience, qualifications, technological skills, and gender (an optional question) , the second theme consisted of 5 questions which essentially surveyed their overall perceptions on or knowledge about corpus integration in EFL classroom, the third theme included 5 questions about views on factors of implementing or adopting corpus-based approach in EFL classrooms. Items from 1 to 4 were used mainly to collect respondent teachers' demographic data like technological skills, teaching experience, the highest earned qualification, Items from 5 to 9 were 5-point Likert scale questions used to survey their overall knowledge about corpus-based approach as a brand-new approach in ELT discipline while items 10 to 14 were used to their views on requirements of adopting or activating corpus-based

The questionnaire data has been analyzed using descriptive statistics of SPSS program. Data analysis will be presented under the three themes of the questionnaire. The

total number of the teachers who responded to all the questions of the questionnaires is 42 EFL university teachers in Qassim Region, in Saudi Arabia. The findings for the three themes will be presented in the following paragraphs.

4.7.1 Respondent Teachers' Demographic Data

The researcher has used Google Forms to design and administer this questionnaire to a randomly chosen sample of Saudi EFL university teachers to investigate their knowledge about and views on corpus- based approach. EXCEL and SPSS programs have been used to analyze collected data as illustrated in the following tables and figures. For the first them, demographic data, data can be seen in the following figure (figure 4.9)



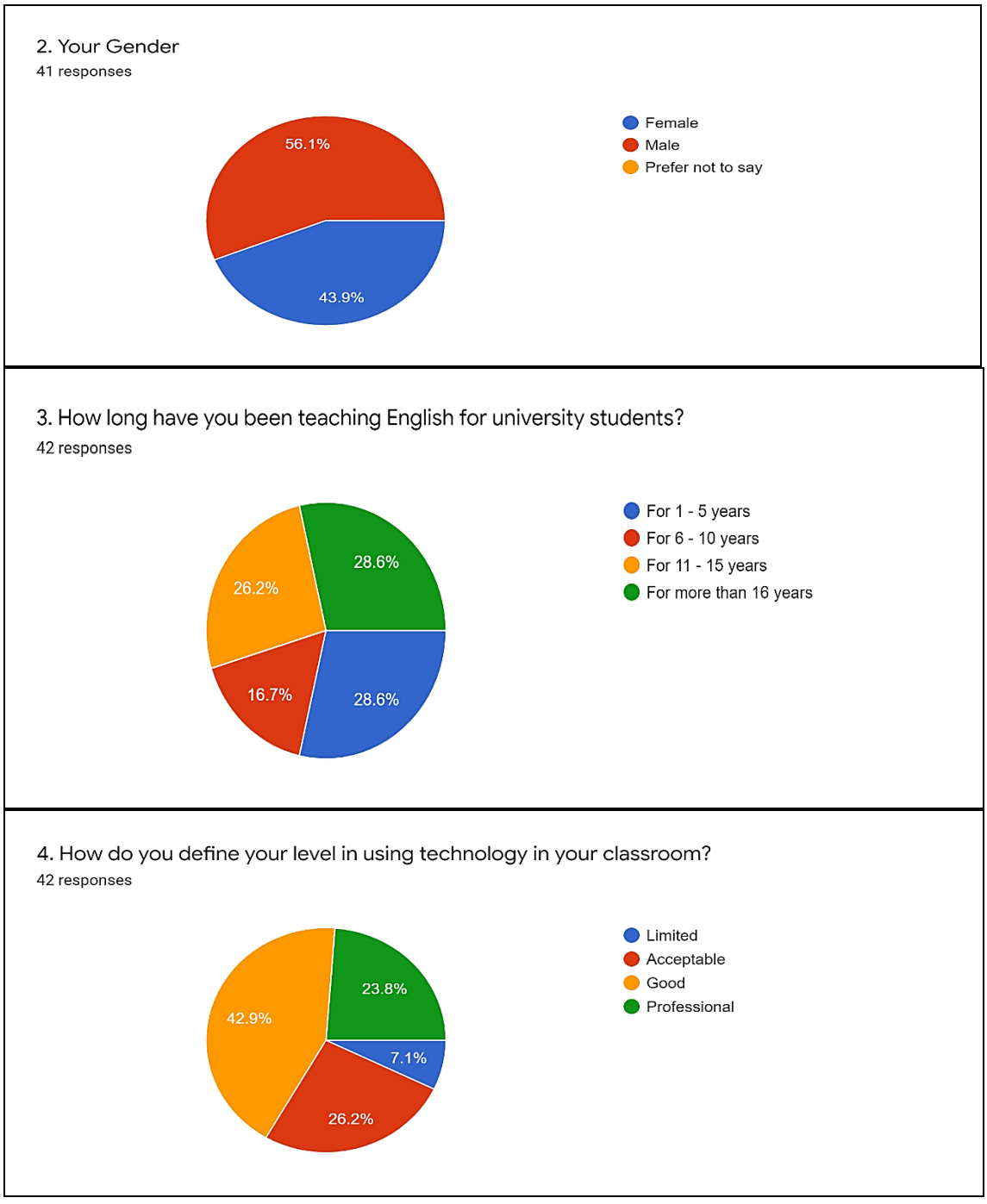


Figure 4.9: Descriptive Statistics for Respondent Teachers' Demographic Data

For details, 38.1% of respondent teachers have PhD Degree English language discipline while M.A degree holders represent 42.9% and in the last place comes B.A degree holders, only 19%. For participants' gender, 56.1% were male teachers while 43.9%

were female teachers which could give some kind of gender balance in the study sample. The third item was about teaching experience; data revealed that both 1-5 years and 16 and above years of teaching experience teachers were 28.6% followed by 11-15 years teachers and then came last teachers of 6-10 years teaching experience. The last item in this theme was about technological skills as an essential part for integrating corpus-based approach and using concordances. Data revealed that 42.9% of respondents were good technology users, 23.8% were professional users, and 26.2% had acceptable skills while only 3 teachers (7.1%) had limited technological skills. These results ascertained that the researcher had chosen a balanced sample of teachers regarding teaching backgrounds, gender, technological skills and qualifications. This balance could result in having a comprehensive representative chosen sample for the population of the study.

4.7.2 Teachers' awareness about Corpus-based Approach

The following table is presenting an overall overview for the two main themes of the questionnaire, teachers' overall awareness or perceptions on corpus-based approach as a brand-new approach, and their views on employing this approach in EFL classrooms. According to table 14.20, the mean score for participants' responses for items of the theme of teachers' awareness on corpus-based approach is 2.15 and for their views on its utilization is 3.5 in a 5-point Likert scale. These results indicated a relatively low level of teachers' awareness towards this approach as most of their responses ranges from *never* to *rarely*. On the other hand, the average rating for the third theme, their views on employing corpus-based approach and its essential requirements, was 3.5 in a five-point Likert scale which ranges from *sometimes* to *often*.

Generally speaking, it can be obviously concluded that respondent teachers' awareness towards employing corpus-based approach in vocabulary learning classes is relatively low, whereas they have highly positive perceptions and views on employing and adopting corpus-based approach in EFL classrooms. The mean scores for each question will be discussed and analyzed according to the two main themes in the following paragraphs.

Table 4.21: Summary of Descriptive Statistics for the Themes of Teachers' Questionnaires

Themes	N. of Items	Minimum	Maximum	Mean	Std. Deviation
Awareness Towards Corpus-based Approach	5	1	4	2.158	0.835809
Their Views on Employing Corpus-based Approach	5	2	5	3.489	1.0125
Total Valid Responses	Items = 10 items - N. of Participants = 42 out of 42				

Starting with the main theme of the questionnaire which was basically intended to investigate teachers' overall awareness and knowledge about corpus-based approach; this theme included the items 5, 6, 7, 8, and 9. As it's shown in (Table 4.21) that the average score of teachers' responses for all the five items of this theme is 2.19 which is below the middle point (point 3) in the 5-point Likert scale of rating. Consequently, it can be concluded that the respondent teachers had a surprisingly low awareness and knowledge level about the proposed approach and the majority of them (approximately 60%) received no or very little training about it whether at teacher preparation colleges or in-service training and workshops programs.

For some more specifics, 69.3% of the informants showed that they *never* or *rarely* heard or read about any of terminology of corpus-linguistics discipline and 30% of them *sometimes* heard or read it. Thus, two thirds (approximately 72%) of the participants have no or very little or moderate knowledge about corpus-linguistics. The second item in this theme, item no. 6, was about whether or not the respondents received training on corpus-based approach during their university study. They responded negatively to indicating that the majority (62.8%) in Saudi Arabia were *never* or *rarely* receiving any corpora training at their universities, while the minority (9.3%) were *often* or *always* receiving training. In the case of the third question, item no. 7, 62.7% of informants were *never* or *rarely* attending workshops or online courses about corpora use whereas 30% came in the middle and a minority (11.6%) were *often* or *always* attending workshops and online courses about corpora use. Moreover, 60.4% expressed that they *never* or *rarely* used any corpus-based tool in their classes, while only 11.6% said that they *often* or *always* use it. For the last item

in this theme, item no. 9, 74.4% of participants responded negatively to whether or not their institutions have subscriptions on corpora websites or have access to concordancing programs like Wordsmith or AntConc, while only one teacher (2.3%) responded positively and the rest (23.3%) remained neutral in this point.

Table 4.22: Summary of Descriptive Statistics for the Items of Overall Perceptions Theme

Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
5. Have you ever heard or read or learnt about the term CORPUS LINGUISTICS or CORPUS-BASED or ONLINE CORPORA or any other Corpus-linguistics terms?	42	23.2%	18.6%	30.2%	18.6%	9.3%	2.45	1.063
6.How often have you ever been trained during your study at your university on using corpus linguistics approach in teaching or research purposes?	42	37.2%	25.6%	27.9%	7%	2.3%	2.14	1.072
7. How often have you ever attended a workshop or a training program on using online corpora in classrooms?	42	48.7%	14%	30.2%	4.7%	2.3%	2.00	1.104
8. How often have your used corpus-based approach in your classroom or for research purposes?	42	30.2%	30.2%	27.9%	9.3%	2.3%	2.26	1.060
9.Are your school or university interested in having accounts on online corpora or purchasing concordance software like Wordsmith or Antconc?	42	48.8%	25.6%	23.3%	0%	2.3%	2.10	1.056
Overall Evaluation for the Theme	42	37.62%	22.80%	27.90%	7.92%	3.70%	2.19	1.071

In summary, this theme had five items that measured teachers' overall knowledge and awareness about corpus-based approach and its tools. It can be concluded that the respondents had a relatively low level of awareness and knowledge about this trendy approach though they may be using many corpus-based teaching materials in their classrooms such as some corpus-designed coursebooks and dictionaries or testing materials or parallel translation corpora websites and Applications. All items of this theme were negatively responded to with a mean score of 2.19 in a 5-point scale. As a result, the last hypothesis has been met through proving the low level of knowledge and awareness of teachers towards corpus-based approach.

This finding is consistent with what the one of the prominent ELT (McCarthy, 2008) said "In my travels across several continents, my common experience is that teachers have

heard of corpora, but they are not quite sure what they are…" (McCarthy, 2008, p. 563). McCarthy (2008) elaborated that much of studies done on utilizing corpus linguistics in language teaching and learning have been implemented by researchers, scholars, who suggested that this is what teachers could need. Thus, He ascertained that corpus-based tools will not become known and prevail till the teacher becomes the informed consumer, knowing how to effectively employ corpora tools and methods.

4.7.3 Teachers' Views on Some Requirements of Corpus-based Approach

The second subordinate theme in teachers' questionnaire was about surveying their views on the main requirements for adopting or encouraging corpora employment in their EFL classrooms. This theme included 5 questions tried to elicit responses to the following thing: a. The effectiveness of integrating corpus-based activities, b. Integrating corpus use training in teacher training programs and colleges, c. the impact of lack of technological equipment in schools on corpora adoption, d. the impact of low technological skills of learners on adopting corpus-based approach, and lastly e. the suitable language levels for employing corpora.

As it can be seen in the following table (Table 4.22), the mean score for the teachers' responses for this theme is above the middle point, point 3, in a 5-point Likert scale which indicates a good level of approval for adopting and encouraging corpora use in EFL classrooms and corpus training in both in-service and pre-service training programs. For the last item which elicited their views on the appropriate language proficiency level to start using corpus-based approach directly, the majority (more than 70%) thought that intermediate and above levels can have direct access to corpora use in their classes, while the minority (less than 30%) believed that it can be used by low intermediate and beginners levels too. (See Figure 4.10 for more details)

Table 4.23: Statistics for the Items of Teachers' Views on Corpus-based Utilization Theme

Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
10. Do you think it is beneficial to integrate corpus-based activities in EFL classrooms?	42	0%	9.3%	44.2%	20.9%	25.6%	3.64	0.983
11. Do you think that Corpus-Linguistics has to be integrated in teacher training programs in your country?	42	0%	4.7%	30.2%	37.2%	27.9%	3.95	0.893

12. Do you think that lack of technological equipment in your school/ university prevents you from using online corpora in your classes?	42	20.9%	23.3%	39.5%	11.6%	4.7%	2.47	1.109
13. Do you believe that students' low technological skills may hinder integrating corpus-based activities in the classroom?	42	34.9%	18.6%	20.9%	23.3%	2.3%	2.31	1.206
Overall Evaluation for the Theme	42	14%	13.98%	33.70%	23.25%	15.13%	3.0925	1.04775

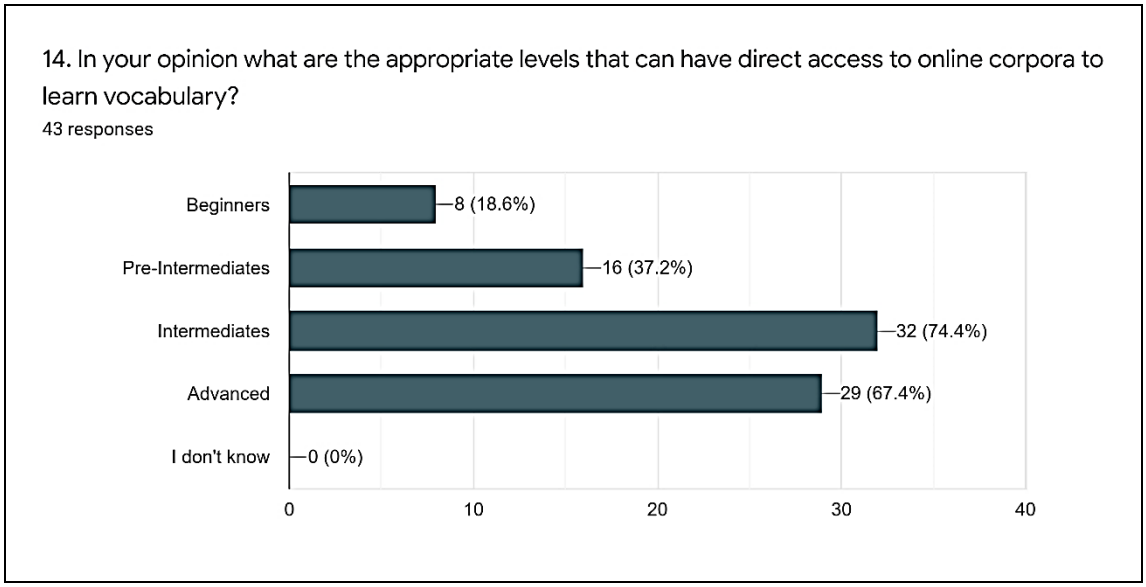


Figure 4.10: Teachers Views on the appropriate Levels for Direct Corpora Use

For much more specifics, regarding the first question (item no. 10), the mean score for their responses to the effectiveness for employing corpus-based activities in EFL classrooms is 3.64 of a 5-point scale rating. This result indicates high level of teachers' approval to employ activities based on corpora. For the second question in this theme (item no. 11), a very high level of approval (3.95 out of 5) for integrating corpora training in pre-service and in-service training programs. It can be seen from the previous table that only 2 teachers (4.2%) disapproved integrating corpora training in teachers training programs while about 40 teachers (95.8%) approved with different levels ranging from *sometimes* to *always*. Responses to the third and the fourth questions (items no. 12 and 13) revealed that the majority of teachers did not believe that lack of technological equipment or learners' low technological skills were preventing teachers from applying corpus-based activities in their classrooms.

Depending on these results, the researcher has concluded that neither lack of computers and internet connection nor learners and teachers' lack of technological knowledge hinder the application of corpus-based approach, but other essential factors related to both learners and teachers lack of knowledge, training, or even fear of trying brand-new approaches.

4.8 Summary

The first point that the researcher has proved as a starting off and basic point before measuring the vocabulary scores differences between the assigned two groups, was the comparability of the two groups, experimental and control. This point has been proved through using two diagnostic tests, CPT test and vocabulary pre-test; and though both groups showed some differences in scores of both tests, the variances have been proved to be nonsignificant as the *p value* was (0.29). At this point, the researcher could easily dismiss the variable of learners' language level or vocabulary backgrounds as an interfering factor that could affect the final scores of both groups.

The main question of this study which included the main variables of the study; the independent variable (using corpus-based approach in vocabulary classrooms) and the dependent variable (learners vocabulary gains in learning and retention). The findings indicated that there were statistically significant differences at the level ($p \leq 0.05$) between the mean scores of posttests for both groups in favor of the experimental group (corpus-based group). Another worth mentioning point is that these differences were substantially in the scores of the productive or transfer part of the posttest which ascertains the importance of novel contexts and authentic data provided by corpora rather than using limited fabricated contexts as in conventional teaching materials. This finding is in alignment with the majority of corpora literature which proved the effectiveness of corpora utilization in vocabulary learning and retention.

A subordinate question was devoted to measuring the differences in performance among learners of the four constituent sub-groups for the experimental group (corpus-based approach group). By this question, the researcher intended to measure the effect size of the moderator variable (learners' proficiency level) on the effectiveness of the independent

variable (using corpus-based approach) on the dependent variable (learners vocabulary gains or their posttest scores). Results revealed that there were statistically significant differences among the learners' gains scores of the four sub-groups in favor of the elementary sub-group with a mean score of 18.4 points and then came pre-intermediate one (17.4 points). The researcher has interpreted these results by the notion of *ceiling effect*; as the pre-test scores of these subgroups were very low and their posttest scores were average, while the upper-intermediate subgroup has relatively high pre-test scores and high posttest scores too. Therefore, in regard to the post test scores the upper intermediates achieved the highest (45.44 out of 50) and them came intermediates (41.6), pre-intermediates (38.2), and in the last place elementary (30.8 out of 50). On the other hand, elementary sub-group achieved the highest gains scores (18.4) then came pre-intermediates (17.4), upper intermediates (16.44), and surprisingly at last intermediates (14.13).

Regarding learners' attitudes, the questionnaire has intended to elicit their attitudes towards and views on three things: first, their overall perceptions on corpora use in vocabulary learning, second, their views on the benefits of COCA corpus use, and third, their views on the potential challenges or difficulties of corpus integration in vocabulary learning. For the first theme, the results revealed learners' positive attitudes and overall perception of corpus-based approach in vocabulary learning with a mean score above the middle point in a 5-point Likert scale rating. For the second theme, they responded positively to the benefits of concordance lines, authentic data, activities and tasks based on COCA corpus with a mean response 4.07 in a 5-point scale. Regarding the last theme, they also responded positively to some potential challenges in using corpus-based activities and tasks with a mean response 3.61 in a 5-point scale. At this conjuncture, this apparent discrepancy in learners' responses had to be justified; though they highly rated the effectiveness and benefits of corpora use, they also highly rated some potential challenges. The researcher has interpreted this point by analyzing their responses to question no. 13 *"I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily"*, as they highly approved this statement with a mean response (3.7) in a 5-point scale. Then, it can be concluded that they encountered some

challenges or difficulties at the beginning stage of use a completely new approach like corpus-based one, but over the course of time they have been trained more and developed some skills to overcome these challenges. This finding ascertains the importance of learners training on corpora use and the gradual integration of corpus in the classroom.

Lastly, the researcher analyzed teachers' awareness of and views on the use of corpus-based approach in EFL classrooms. The data revealed the reliability of the chosen sample as the proportions of distribution of gender, qualifications, teaching experience, and technological skills were balanced and achieved variety. The questionnaire has two themes, the main theme included 5 questions about their overall awareness of corpus-based approach. This theme received a relatively low rating (2.15 in a 5-point Likert scale) which indicated low level of teachers' awareness of employing corpus-based approach in their EFL classrooms. This result can interpret the big gap between corpus literature and de facto utilization of corpus-based approach in EFL classroom.

CHAPTER V

DISCUSSION AND CONCLUSION

5.1 Introduction

In this chapter the researcher will first discuss the main findings and conclusions of the study's five questions with much emphasis on the similarities and contradictions with previous studies and theories. The second part will be devoted to implications and suggestions for Adopting corpus-based approach in the Saudi context. The third part will include study limitations and recommendations for further studies.

Substantially, the study tried to investigate the feasibility or the effectiveness of incorporating language corpora in teaching and learning English vocabulary, that's why the researcher has indexed the study's five questions into three main themes or topics, i.e. the effect of corpus-based approach on vocabulary gains and retention, learners attitudes towards corpora employment, and teachers awareness of and perceptions on corpus-based approach in vocabulary learning and teaching classes. Accordingly, Findings discussions and implications will be demonstrated and categorized under these themes. The following table (Table 5.1) demonstrates the study's questions and themes.

Table 5.1: Study's Questions and Main Themes.

Study Questions	Study's Themes
1. What is the effect of using corpus-based approach on improving vocabulary learning in Saudi EFL classroom?	1. The effect of corpus-based approach on vocabulary gains and retention.
2. How far is corpus-based approach effective in improving vocabulary retention in Saudi EFL classroom?	
3. What is the effect size of learner's proficiency level (as a moderator variable) in the effectiveness of using corpus-based approach in vocabulary learning?	
4. What are the attitudes of the experimental group's students towards using corpus-based approach in vocabulary learning classes?	2. Learners attitudes towards corpora employment.
5. To what extent are EFL teachers in Saudi Arabia aware of using corpus-based approach in EFL classrooms?	3. Teachers awareness of and perceptions on corpus-based approach.

5.2 Summary of the Findings

As the researcher previously mentioned that the study's five questions have been catalogued into three themes, effect size of incorporating corpus-based approach in vocabulary learning gains and retention rate, learners' attitudes towards corpora utilization, and finally EFL teachers' awareness of and perceptions on corpus-based approach in the Saudi EFL context. For the first theme, learning gains, the findings strongly proved the effectiveness of corpora integration on learning gains and long-term retention. The average scores of posttests and delayed posttests for experimental group's learners (40.23 and 41.5 for posttests and delayed posttest respectively) were significantly higher than those of the control group (32.21 and 32.8) which indicated high differences in both vocabulary learning gains and retention rates.

The researcher has gone deeper in analyzing these variances in gains through pinpointing the areas of the experimental group's overperformance through analyzing each group's average score in both parts of the test separately. This analysis revealed that both groups did well in the definitional part of the test, while the experimental group by far overperformed the control one in the transferrable or productive part of the test. The researcher has concluded that the metalinguistic benefits provided by corpora use has enhanced the experimental group's learners' capacity in using the learnt words in novel contexts.

At the end of the experiment, leaners of the experimental group have revealed positive attitudes towards the use of corpus-based approach in their vocabulary leaning classes. The majority of them (about 68%) responded positively to the benefits and overall perceptions of corpora use and only 7.25% responded negatively and the remaining was neutral. At the same time, they highly rated the challenges and difficulties of using corpora, especially at early stages, which could be met and overcome over the time and increased training. In addition, the majority (approximately 73.1%) demonstrated their desire to use corpora at their own in the future and only 7.7% disagreed.

Regarding the last theme, teachers' awareness of and perceptions on corpus-based approach, results revealed quite low level of teachers' (mean score 2.15 in a 5-ponit scale)

awareness in the Saudi EFL context of Corpora use. However, the majority expressed positive perceptions towards integrating corpora in teacher training colleges and programs and encourage its use in EFL classrooms.

5.3 Discussions on the Study Findings

The results of the study will be compared to its broader context of corpus prominent studies and significant theories in order to identify similarities and differences in respect to the actual results. The following sections will handle deeper results discussions for the study's main topics.

5.3.1 Effects of Corpus-based Approach in Vocabulary gains and Retention Rate

The researcher has hypothesized that there are statistically significant differences in vocabulary gains and retentions rates between the corpus-based group and the conventional teaching one in favor of the corpus-based. These hypotheses have been met and proved through the given results from all tests scores and the resulting data from t-tests that ascertained the significance of these differences of tests scores. It is evidenced that the average scores for the experimental group is higher than the average scores of the control group for posttests and delayed posttests. The researcher concluded that using corpus-based approach can maximize learners' vocabulary acquisition and retention regardless of their language proficiencies. However, apart from the effect size on learning gains or retention rate, it was proved that the bigger effect size was on the transferrable or productive knowledge of vocabulary. The researcher found that both approaches, conventional and corpus-based approach, had a great effect size on the definitional knowledge of the target vocabulary, while corpus-based approach had greater effect on the transferrable or productive knowledge than the conventional approach. As the definitional part of a word simply requires word memorization which can be serviced by conventional methods such as dictionaries and paper-based coursebooks, while the transferrable or productive part of a part requires further linguistics and meta-linguistics knowledge of a word (Qian, 1999).

These conclusions could be elaborately justified in the light of Cobb's (1999) phenomenal study on corpora use. Cobb (1999) concluded that differences in the

definitional knowledge of vocabulary were marginal and nonsignificant, while the experimental groups' learners who received concordances-based hands-on and tasks showed higher ability in transferring this knowledge into new contexts and retaining it to much longer periods. On the other hand, students who used wordlists or dictionaries could have strong definitional knowledge of vocabulary items and even this knowledge was not well retained in a longer time span (Cobb, 1999, P.30).

Consistently, another prominent study was Nation's (2001) and Nagy's (2007), they concluded that the abundance of contexts and word encounters could facilitate and reinforce word's acquisition and retention. Consequently, vocabulary is best learnt and retained when being learnt through abundant authentic contexts with or without deliberate word-meaning teaching. This can be attributed to the notion of multifaceted nature of words learning that cannot be covered through wordlists or dictionaries. Knowing a word is not just identifying its meaning, it also includes many linguistic and metalinguistic aspects, such as, collocations, colligation, part of speech, synonyms, prosody, morphology, phonology, spelling, frequency, register, genre, and semantics or connotation. Most of these facets are less likely to be manifested and grasped through conventional tools, though using rich authentic concordance data and a variety of query tools and options could help in recognizing and understanding these aspects inductively or deductively. Thus, learning vocabulary through corpora integration enables learners to explore different linguistic features of a word through the provided variety of linguistic examples in different meaningful contexts embedded in the resulting concordances lines. In addition, this abundance of meaningful data enables learners to reencounter the target word in different and occurring contexts which significantly enhance both definitional and productive knowledge of words and long-term memorization.

Cobb (1999) ascertained that solely new words in conventional reading text books could be unnoticed by leaners with its limited chances of reoccurring and restricted contexts. Though there could be good leaning gains, these gains remain less transferrable and less retainable in a long-term than that of corpus-based.

Going beyond the benefits of providing abundant authentic data and recurring words encounters, some researchers highlighted the benefit of learner's autonomy when using corpus-based approach. This approach encourages learner's active participation and student-centered learning style. Chen (2004) claimed that a learning process in which learners have to learn knowledge prepared only by teachers, is considered a passive way of learning. Conversely, corpus linguistics enables learners to take control of their own learning to achieve the required tasks through identifying query types, choosing relevant concordances, noticing words in their contexts, identifying frequencies, word collocations, clusters, inferring patterns, inferring grammatical rules ……etc. Then learners have the chances to observe obtained concordances lines, identify word's collocations and parts of speech, interpret the queries data, and infer recurring patterns which essentially making them active and autonomous learners (Nation, 2001 & Chen, 2004). Hadley and Charles (2017) proved in his study that learners were persistently active while being involved in "a content decision making situation" throughout noticing and exploring obtained corpus data. This type of active interaction with language input made their classes much more meaningful, active, and learner-centered. This finding is aligned with Cobb's (1999) statement that "knowledge encoded from data by learners themselves will be more flexible, transferable and useful than knowledge encoded for them by experts and transmitted to them by an instructor or other delivery agents" (P.15).

To recap, the greater vocabulary gains and the longer retention of the experimental group's learners have been proved to be as a result of the use of corpus-based approach. This effect size of corpora utilization could be interpreted by its features of providing abundant authentic contexts, plenty of word encounters, enhancing learner's autonomy and active participation in content selection and learning processes, and manifesting linguistic and meta-linguistic aspects of a word. All these features could by far consolidate vocabulary knowledge acquisition and retention more than that's of the conventional methods.

5.3.2 Learners' attitudes towards Corpus-based Approach

The second theme of this study is about learners' attitudes towards employing corpus-based approach in their vocabulary learning classes. The research has found that the majority of participants (approximately 70%) had positive attitudes towards the use of corpora in their classroom. This finding is aligned with the results of many previous studies about corpora (Chan and Liou, 2005; Kaur and Hegelheimer, 2005; Hadley, 2002 and 2017; Sirphicharn, 2003; Horst et all, 2005; Cobb, 1999; and Cobb, 2007).

Generally speaking, they highly rated the statements about benefits of using corpora and their overall evaluation for corpus-based hands-on and tasks (mean responses 4.01 and 3.83 respectively). Concurrently, they highly rated potential challenges and difficulties throughout their experiment. The researcher has interpreted this discrepancy in the light of their high approval rating (mean response 3.66) of item 13 which stated that over time they got used to corpora use and managed to overcome these challenges and difficulties. This finding highlighted the importance of training learners and teachers before adopting corpus-based approach to ensure having the required skills of dealing with the abundance of provided data, search query types and options, noticing language patterns, determining the relevant concordance lines, and reading and comprehending cut-off sentences.

Though, the majority of them highly approved the benefits of using corpora and highly rated corpus-based approach as a beneficial tool in vocabulary learning and retention, they believed that it is a challenging approach which requires training over time and much support from teachers' side especially at the first stage. For this reason, the researcher has adopted a gradual approach for corpora integration, starting with simple paper-based activities and more teacher-centered activities and ending in autonomous learning activities and tasks with direct access to COCA Corpus and with a little teacher's role as a monitor to achieve these tasks. Hardly (2001) and Aston (2002) pointed out the significance of the teacher's role in this initial stage of application to mitigate the effect of potential confusion or frustration caused by large number of obtained concordance lines in each search. Therefore, they suggested in their studies that the teacher, at the beginning

stage, have to screen out difficult or irrelevant data, select the relevant easy ones, and organize them in paper-based activities to be easily manageable and comprehensible.

From another perspective Fox (1998) interpreted learners' positive attitudes towards corpus-based approach by the preference and excitement of independent learning through discovering linguistic patterns and doing things themselves. The idea of learning autonomy is believed to be highly motivating for learning and engaging in classroom activities.

5.3.3 Teachers' Views Toward Using Corpus Approach

The third theme in this study is related to teachers' awareness of corpus-based approach in the Saudi EFL context. Though there are prominent studies and sharply increasing interest in corpus use in the academic context among ELT scholars and linguistics researchers, the de facto teaching practices in EFL classrooms never reflect this trend or adopt these recommendations. This observation or finding was the strong motivation for the researcher to investigate teachers' awareness of and overall perceptions on corpus use as they are the corner stone of the educational context. McCarthy (2008) said in justification of this discrepancy, that *"In my travels across several continents, my common experience is that teachers have heard of corpora, but they are not quite sure what they are…" (p. 563).* He resumed in his elaboration that corpora studies and recommendations were proposed by scholars and researchers who decided that that is what teachers need to do in their classrooms. He stressed on the importance of involving teachers in these studies to be the informed customer who can use and evaluate corpora effectively. This idea of teachers' awareness was stressed through many other relevant studies (Boulton, 2010; Hunston, 2002; McCarthy, 2008; Romer, 2009). Consequently, there is an agreement among scholars and professionals that any endeavors for enhancing corpora use in EFL classroom have to start with teacher' awareness raising, training, and effective involvement in studies and experiments.

In this study, the chosen sample of Saudi EFL teachers has demonstrated relatively low level of awareness of corpus-based approach as approximately 41.8% said that they *never* or *rarely* heard of corpus linguistics or any of its terminologies, corpora, corpus-based approach, Data Driven Learning, Concordances lines, …etc. In addition, the majority

(62.8%) *never* or *rarely* studied corpus-linguistics in their universities, teacher training or preparation programs and the majority (62.7%) *never* or *rarely* received any corpora training throughout the years of their teaching profession. Moreover, the majority (74.4%) responded negatively to the statement about their universities or schools' interest in corpora use or online corpora subscription or concordancing software purchasing. All these results revealed the low levels of teachers' awareness and training of corpus linguistics which altogether considered the main hindrance of adopting corpus-based approach in the Saudi EFL context. Some researchers like Boultan (2008) and (2010), Hunston (2002) and McCarthy (2008), and Romer (2009) researchers, though, listed other supplementary reasons like teachers' lack of understanding, unwillingness to try new ways, fears of technology use, fears of not being able to interpret concordances outputs, lack of corpus-based teaching materials that meets lessons objectives, and lastly learners' fears and pressure. Flowerdew (2009) argued that besides all the aforementioned reasons there is still the argument on the limitations and challenges of corpus-based approach that may cause teachers' concerns in case of using it.

For the second theme of teachers' questionnaire, the majority (46.5%) responded positively to the integration of corpus-based approach in EFL context and only 9.3% responded negatively. Though most of them (approximately 60.4%) *never* or *rarely* experience the approach, they have positive attitudes and perceptions towards using corpora and integrating corpora training in teacher's preparation programs and in-service training programs. Generally, most of the, (approximately 53.5%) *never* or *rarely* agree that lack of learners' technological skills or lack of schools' technological equipment hinder the use of corpora in EFL classrooms. Lastly, regarding their views on the appropriate proficiency level for using corpora, the majority believed that corpora can be effectively used with intermediates and above proficiency levels.

Overall, EFL teachers in the Saudi context seemed to be less aware of corpora use in their classroom and had no or very little training on corpus tools whether in their preservice time period or in-service. At the same time, they had positive attitudes and enthusiasm towards being trained on corpora employment and integrating corpus training in teacher preparation programs. This lack of awareness or understanding or training has to

be met if we want to encourage corpora use in the Saudi context. The researcher found few studies that handled this essential part of corpus study related teachers' awareness, attitudes, roles, and training, that why further studies have to be done in this area of research.

5.4 Implications for Corpora Integration in EFL Classrooms

The benefits of employing corpora in teaching or learning vocabulary have been highly approved and empirically proved by numerous studies, though, the researcher assumed these benefits would not be attainable unless learners have sufficient corpus use training and experience in reading concordances lines, inferring patterns, using query tools and so on. This requirement substantially entails teachers' good understanding of corpus integration to be able to set applicable gradual corpus-based lesson plans, instructions, activities and tasks that ensures learner's autonomy and engagement. Therefore, teachers' thorough understanding of corpus use, learners' enough training and simplified well prepared materials are indispensable in initial stages of corpus application. Given the results of this study, the researcher has formulated the following implication for an effective corpus use in the classroom:

1. The results of these studies showed that both learners and teachers agreed on the effect of language proficiency level on corpus-based approach benefits. They believed that low levels may sink in the massive amounts of query obtained data and the numerous unfamiliar words in concordances which eventually effect their learning gains and cause frustration. Consequently, as I aforementioned, the researcher adopted a flexible mode of corpora utilization which ensured essential role for the teacher in supporting students and preparing paper-based activities, instructional guidelines, hands-on, and so on. Thus, the lower level the learners are, the more time and training are required to get them ready for this new way of learning. At this point, teachers are required to bridge the gap, especially with low levels learners and at initial stages of application, through employing ready-made teaching materials, scaffolding activities, simple instructional guidelines, and offering constant support during students' use of corpora.

2. Although the majority of learners highly rated corpus as an effective learning tool, some of them had experienced some difficulty in understanding the context because of unfamiliar words. Accordingly, the researcher, in this study, has chosen a middle-way approach; he never employed pure DDL or Corpus-based approach which requires learners' direct access to corpora with so much autonomy, rather he highlighted the role of the teacher as a lesson planner, activity designer, technical support provider, and sometimes translator for unfamiliar key words especially at initial stages and with low levels. Students could be encouraged to use other supplementary tools with concordances lines to get clarifications and overcome potential problems such as asking a partner or the teacher.

3. Krashen (2009) stressed on the providing comprehensible input as an essential condition for effective learning. However, it occasionally happens with learners, especially low levels, to encounter unfamiliar contexts or incomprehensible outputs that may cause frustration and distract learning process. Thus, the researcher suggested that teachers could prepare paper-based hands-on and materials including comprehensible inputs and simplified samples that fit learners' proficiency level and enhance learning the target vocabulary. This comprehensible and proficiency-fit input are suggested to be one level higher than the learner's level. In traditional strict corpus-based method or DDL, teachers' role is restricted to monitoring and planning, though, in this study the researcher has adopted a more flexible mode which ensured teacher's essential role in facilitating the learning process of students and making sure that they are able to effectively use the corpus and infer language patterns from obtained concordance lines before letting them work alone and have direct access to corpora. In this proposed model, teacher's involvement in every stage and every procedure is fundamental till students gain the required skills for corpus use and concordances reading and be able to screen out irrelevant contexts.

4. Though learners of corpus-based approach could have good vocabulary gains, they still need activities and exercises to ensure their recycling of the learnt vocabulary items in new contexts and situations. Accordingly, teachers are highly encouraged to prepare follow-up activities and exercises at the end of each chapter

or unit to let students reuse learnt words in new contexts. One of the most recommended activities is writing activity which ensure students productive use of vocabulary.

5. The researcher assumed that corpus-based approach can be used solely or in conjunction with other conventional methods according to the intended learning outcomes for the lesson. Consequently. A model that combine both concordances-based and conventional approach in an integrated way could be the best fit for the Saudi EFL context and more applicable for various learning objectives. Adopting such method could maximize corpus benefits, fit a variety of lesson types, and meet potential challenges in one-way corpora integration.

6. It's evidenced that learners' sufficient and effective training on corpus use and concordances reading is fundamental and indispensable for a successful corpus-based application. The researcher has noticed, in this treatment, that acquiring the technical skills of online corpora use happened easily and quickly, while the more sophisticated and complex concordancing skills of noticing, determining, grouping, analyzing, interpreting, and inferring took much longer time, carefully planned training, and adequate practice. Accordingly, learners' corpus training has to be carefully prepared in a gradual enriched way to cover both technical skills and the most importantly concordancing skills.

7. The cornerstone of the effective use of corpus-based approach is a well-trained teacher that has the required skills for introducing and implementing such brand-new approach in his classroom. Thus, equipping teachers with adequate training on all aspects of corpus use is inevitable for effective adoption of this approach in EFL context. The proposed model of training, whether pre-service or in-service, would include technical skills of online corpora use, pedagogical skills of dealing with obtained concordances lines, pedagogical skills of corpus-based lesson plans, pedagogical skills of designing corpus-based hands-on and designing corpus-based teaching materials and tests. Concurrently, EFL professionals could prepare corpus-based teaching materials, lesson plans, activities, guiding booklets, teaching

manuals, and video demo lessons to be easily accessed and ready for use for teachers.

5.5 Limitations of the Study

The study is limited to identifying effectiveness of using corpora in vocabulary learning in tertiary EFL classroom in the Saudi context. It's also restricted to the foundation level at a private college in Saudi Arabia, Burayadah Private Colleges in Qassim Region. It has been conducted on two groups of (54) male students and (42) EFL male and female teachers at different universities and colleges (Qassim University, Burayadah Private Colleges, Alrajhi University, and Taif University) in Saudi Arabia. The motivation in choosing this group stems from its distinctive features. It has a collection of teachers who have a diversity of cultural, academic and educational backgrounds. Consequently, the results of this study could be more reliable and meaningful.

The overall findings of this study showed encouraging improvements in the experimental group scores which indicated a remarkable effect size of using corpus-based approach in vocabulary learning. To make sure that this effect has attributed to corpus linguistics integration, the researcher has taken some procedures to minimize other potential confounding variables that may affect the results as follows:

1. The researcher tried to reduce learners' anxiety of exams and make them be viewed as classroom activities rather than personal assessment. In addition, the researcher tried not to stress on the significance of tests.
2. When learners were doing their pre-tests, they were not told that they would be given a posttest later. Thus, they were not motivating to write down or remember their answers in the pretest.
3. After doing the pretest, learners were not given the right answers for the items of the test to reduce the shadowing effect of remembering right answers when doing the posttest.

4. Despite, the items of pre-tests and post-test were identical, the researcher has reordered test items in post-test and delayed post-test.

On the other hand, there were some conditions that haven't been controlled and could have some effect on the scores, as follows:

1. The researcher noticed that some students in both groups, experimental and control groups, lose their interest or motivation in doing their best when answering tests as they felt that tests were insignificant and they didn't count for their assessment in the course. This was quite clear in a few tests' papers as students left a whole question blank.

2. Learners were told days earlier about the proficiency test CPT which made a few students studied the model answers for CPT before taking the exam; and this incident could affect to some extent the overall categorizations of proficiency sub-groups.

3. Learners were given the delayed post-test online using EDMODO platform for quizzes and assignments as attending to the Colleges was suspended at that time. Thus, they didn't receive the same supervision, feedback or monitoring as on campus.

Consequently, the pure causality of corpus-based approach adoption and the experimental group's improvement in vocabulary learning and retention could not be claimed unless all the three previously mentioned confounding items screened out.

5.6 Recommendations for Further Study

A highly recommended corpus study could be for investigating the changing roles of teachers in a corpus-based classroom compared to the conventional roles in traditional classrooms. This study could specifically identify teachers' roles and responsibilities in each stage of a corpus-based classroom. Another aspect of corpora that can be probed is designing a proposed model for learners training on both technical skills on online corpora use and concordancing skills required for analyzing and interpreting obtained query data.

This study would aim at designing training plans, teaching materials, instructional guidelines, corpus-based hands-on, and practice exercise. The third recommended is designing a curriculum for teacher' training that achieve all embedded skills in corpus-based approach application. This study could propose a framework for designing a teacher training curriculum that fill in this great gap of teachers' lack of training in EFL context. A fourth suggested study can investigate the effect of using a mixed approach of both concordances-based and conventional dictionaries like GOOGLE TRANSLATE and MERRIAM WEBSTER, ONLINE OXFORD DICTIONARY and others to check unfamiliar words in concordances lines. In the same context, another proposed study could investigate the effect of using a parallel online corpus on improving writing skills. Lastly, A further study is needed to investigate the effect size of corpus-based approach on reading skills or the level of reading proficiency as in this study the researcher measured its effect on vocabulary gains and retention rates.

REFERENCES

Abbs, B., et al., 2008. "Is overt repetition critical to expressive word learning? The role of overt repetition in word learning with and without semantics." *Applied Psycholinguistics 29(4):* pp. 627-667.

Aijmer, K. ed., 2009. *Corpora and language teaching (Vol. 33*). John Benjamins Publishing.

Alharbi, R.M.S., 2017. *Acquisition of lexical collocations: a corpus-assisted contrastive analysis and translation approach (*Doctoral dissertation, Newcastle University).

Aljuhani, H.S., 2016. Using corpora in a lexicalized stylistics approach to teaching English-as-a-foreign-language literature.

Al-Mahbashi, A., Noor, N.M. and Amir, Z., 2015. The Effect of Data Driven Learning on Receptive Vocabulary Knowledge of Yemeni University Learners. *3L: Language, Linguistics, Literature®, 21(3).*

Almutairi, N.D., 2016. The Effectiveness of Corpus-Based Approach to Language Description in Creating Corpus-Based Exercises to Teach Writing Personal Statements. *English Language Teaching,* 9(7), pp.103-111.

Al-Nasser, A.S., 2015. Problems of English language acquisition in Saudi Arabia: An exploratory-cum-remedial study. *Theory and Practice in Language Studies, 5(8),* pp.1612-1619.

Alqarni, A., 2019. Formulaic Sequences: An Investigation of an EFL Textbook in Saudi Arabia.

Alruwaili, A., 2018. *Integrating corpus linguistics in second language vocabulary acquisition* (Doctoral dissertation, University of Nottingham).

Alruwaili, A.K., 2020. Data-driven Learning Approach for Teaching Verb-Noun Collocations in an English Foreign Language Context. Asiatic: *IIUM Journal of English Language and Literature, 14(1),* pp.138-159.

Anthony, L., 2013. A critical look at software tools in corpus linguistics. *Linguistic Research, 30(2),* pp.141-161.

Aston, G., 1995. Corpora in language pedagogy: Matching theory and practice. *Principle and practice in applied linguistics: Studies in honor of HG Widdowson,* pp.257-270.

Aston, G., 1997. Enriching the learning environment: Corpora in ELT. In *Teaching and language corpora (*pp. 51-64).

Aston, G., 2002. The learner as corpus designer. In *Teaching and learning by doing corpus analysis* (pp. 7-25). Brill Rodopi.

Baleghizadeh, S. and Shahry, M.N.N., 2011. The effect of three consecutive context sentences on EFL vocabulary-learning. *TESL Canada Journal*, pp.74-74.

Barbieri, F. and Eckhardt, S.E., 2007. Applying corpus-based findings to form-focused instruction: The case of reported speech. *Language Teaching Research, 11(3),* pp.319-346.

Barcroft, J., 2007. Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning, 57(1)*, pp.35-56.

Barlow, M., 1996. Corpora for theory and practice. *International journal of corpus linguistics, 1(1),* pp.1-37.

Bennett, G.R., 2010. *Using corpora in the language learning classroom: Corpus linguistics for teachers. University* of Michigan Press.

Bernardini, S., 2002. Exploring new directions for discovery learning. In *Teaching and learning by doing corpus analysis* (pp. 165-182). Brill Rodopi.

Biber, D., Douglas, B., Conrad, S. and Reppen, R., 1998. *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Boontam, P. and Phoocharoensil, S., 2018. Effectiveness of English Preposition Learning through Data-Driven Learning (DDL). *3L: Language, Linguistics, Literature®, 24(3).*

Boulton, A., 2008, July. DDL: Reaching the parts other teaching can't reach?

Boulton, A., 2010. Learning outcomes from corpus consultation.

Bracewell, R., Laferriere, T. and Gregoire, R., 1996. The contribution of new technologies to learning and teaching in elementary and secondary schools. *Retrieved September, 2,* p.2003.

Carter, R., McCarthy, M., Channell, J. and McCarthy, M., 1988. *Vocabulary and language teaching* (pp. 68-75). London: Longman.

Celik, S. and Elkatmis, M., 2013. The Effect of Corpus Assisted Language Teaching on the Learners' Proper Use of Punctuation Marks. Educational Sciences: *Theory and Practice, 13(2),* pp.1090-1094.

Chambers, A., 2007. Popularising corpus consultation by language learners and teachers. In *Corpora in the foreign language classroom (*pp. 1-16). Brill Rodopi.

Chambers, A. and O'Sullivan, Í., 2004. Corpus consultation and advanced learners' writing skills in French.

Chan, T.P. and Liou, H.C., 2005. Online verb-noun collocation instruction with the support of a bilingual concordancer. In *Selected papers from the fourteenth international symposium on English teaching* (pp. 270-281).

Chang, W.L. and Sun, Y.C., 2009. Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning, 22(4),* pp.283-302.

Chen, J., Chau, R. and Yeh, C.H., 2004, January. Discovering parallel text from the World Wide Web. In *Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalization-Volume 32* (pp. 157-161).

Coady, J., 1997. L2 vocabulary acquisition: A synthesis of the research' in J. Coady and T. Huckin eds): Second Language Vocabulary Acquisition.

Coady, J., 1997. L2 vocabulary acquisition. *Second language vocabulary acquisition*, pp.273-290.

Cobb, T., 1997. Is there any measurable learning from hands-on concordancing? *System, 25(3),* pp.301-315.

Cobb, T., 1999. Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12(4),* pp.345-360.

Cobb, T., 2003. Do corpus-based electronic dictionaries replace concordancers. *Directions in CALL: experience, experiments, evaluation. Hong Kong: Polytechnic University,* pp.179-206.

Cobb, T., 2007. Computing the vocabulary demands of L2 reading. *Language Learning & Technology, 11(3*), pp.38-63.

Coniam, D., 1994. Using corpus word frequency data in the automatic generation of English language cloze test. *Entering text,* pp.29-44.

Conrad, S., 2005. Corpus linguistics and L2 teaching. In *Handbook of research in second language teaching and learning* (pp. 417-434). Routledge.

Cotos, E., 2014. Enhancing Writing Pedagogy with Learner Corpus Data. *ReCALL, 26(2),* p.202.

Coxhead, A., 2002. The academic word list: A corpus-based word list for academic purposes. In *Teaching and learning by doing corpus analysis* (pp. 72-89). Brill Rodopi.

Crystal, D., 1994. *An encyclopedic dictionary of language and languages*. Penguin.

Cullen, R. and KUO, I.C.V., 2007. Spoken grammar and ELT course materials: A missing link? *TESOL Quarterly, 41(2),* pp.361-386.

Dash, N.S., 2008. *Corpus linguistics: An introduction*. Pearson Education India.

Davies, M., 2009. The 385+ million word Corpus of Contemporary American English (1990–2008+): *Design, architecture, and linguistic insights. International journal of corpus linguistics, 14(2),* pp.159-190.

Davies, M., 2010. The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and linguistic computing, 25(4),* pp.447-464.

Davies, R.S. and West, R.E., 2014. Technology integration in schools. In *Handbook of research on educational communications and technology* (pp. 841-853). Springer, New York, NY.

De Groot, A. M. B., 2006. Effects of Stimulus Characteristics and Background Music on Foreign Language Vocabulary Learning and Forgetting. *Language Learning, Vol. 56(3),* PP 463–506.

Edo-Marzá, N., 2014. "A practical corpus-based approach to teaching English for tourism."

Elsherbini, S.A.H. and Ali, A.D., 2017. The Effects of Corpus-based Activities on EFL University Students' Grammar and Vocabulary and Their Attitudes toward Corpus. *Journal of Research in Curriculum Instruction and Educational Technology, 3(1),* pp.133-161.

Fareh, S., 2010. Challenges of teaching English in the Arab world: Why can't EFL programs deliver as expected?. *Procedia-Social and Behavioral Sciences, 2(2),* pp.3600-3604.

Fehr, C.N., Davison, M.L., Graves, M.F., Sales, G.C., Seipel, B. and Sekhran-Sharma, S., 2012. The effects of individualized, online vocabulary instruction on picture vocabulary scores: An efficacy study. *Computer Assisted Language Learning, 25(1),* pp.87-102.

Feng-ran, C.H.E.N., 2011. A corpus-based study of English vowel acquisition of Chinese college students [J]. *Journal of Hebei Normal University (Educational Science Edition),* 5.

Flowerdew, J., 1996. Concordancing in language learning. *The power of CALL,* pp.97-113.

Fox, G., 1998. Using corpus data in the classroom. Materials development in language teaching, ed. by Brian Tomlinson, 25–43.

Gass, S., Loewen, S. and Plonsky, L., 2020. Coming of age: the past, present, and future of quantitative SLA research. *Language Teaching*, pp.1-14.

Gavioli, L., 1997. Exploring texts through the concordancer: Guiding the learner. In *Teaching and language corpora (*pp. 83-99).

Gavioli, L., 2001. The learner as researcher: introducting corpus cocordancing in the classroom. *The learner as researcher,* pp.1000-1030.

Gavioli, L. and Aston, G., 2001. Enriching reality: language corpora in language pedagogy. *ELT journal, 55(3),* pp.238-246.

Geluso, J. and Yamaguchi, A., 2014. Discovering formulaic language through data-driven learning: Student attitudes and efficacy. *ReCALL: the Journal of EUROCALL, 26(2),* p.225.

Götz, S. and Mukherjee, J., 2006. Evaluation of data-driven learning in university teaching: A project report. *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods. Francfort: Peter Lang*, pp.49-67.

Gravetter, F.J., Wallnau, L.B., Forzano, L.A.B. and Witnauer, J.E., 2020. *Essentials of statistics for the behavioral sciences*. Cengage Learning.

Hadley, G., 2001. *Concordancing in Japanese TEFL: Unlocking the power of data-driven learning.* na.

Hadley, G., 2002. An introduction to data-driven learning.

Hadley, G. and Charles, M., 2017. Enhancing extensive reading with data-driven learning. *Language Learning & Technology, 21(3),* pp.131-152.

Hatch, E.M. and Lazaraton, A., 1991. *The research manual: Design and statistics for applied linguistics.* New York, NY: Newbury House Publishers.

Horst, M., Cobb, T. and Nicolae, I., 2005. Expanding academic vocabulary with an interactive on-line database. *Language learning & technology, 9(2),* pp.90-110.

Huang, C.R., Lin, J., Jiang, M. and Xu, H., 2014, August. Corpus-based study and identification of Mandarin Chinese light verb variations. In *Proceedings of the first workshop on applying NLP tools to similar languages, varieties and dialects* (pp. 1-10).

Huang, L.S., 2011. Corpus-aided language learning. *ELT journal, 65(4),* pp.481-484.

Hulstijn, J.H., 2007. The shaky ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal, 91(4),* pp.663-667.

Hunston, S., 2002. *Corpora in applied linguistics.* Ernst Klett Sprachen.

Johansson, C. and Geisler, C., 2009. The Uppsala Learner English Corpus: A new corpus of Swedish high school students' writing. *of Linguistics*, p.181.

Johns, T., 1991. Form printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In *T. Johns & P. King (Eds.),* Classroom Concordancing, *ELR Journal,* 4 (pp.27-45).

Johns, T., 1991. Should you be persuaded: Two samples of data-driven learning materials.

Johns, T., 2002. Data-driven learning: The perpetual challenge. In *Teaching and learning by doing corpus analysis (*pp. 105-117). Brill Rodopi.

Kaur, J. and Hegelheimer, V., 2005. ESL students' use of concordance in the transfer of academic word knowledge: An exploratory study. *Computer assisted language learning, 18(4),* pp.287-310.

Kennedy, G., 1991. Between and through: The company they keep and the functions they serve. *English corpus linguistics,* pp.95-110.

Kettemann, B., 1995. On the use of concordancing in ELT. AAA: *Arbeiten aus Anglistik und Amerikanistik,* pp.29-41.

Krashen, S., 2009. Does intensive decoding instruction contribute to reading comprehension? *Knowledge Quest*, 37(4), p.72.

Krishnamurthy, R., 2001. Learning and teaching through context-A data-driven approach. *TESOL Spain Newsletter, 24*, p.2001.

Laufer, B. (2001). Reading word-focused activities and incidental vocabulary acquisition in a second language. Prospect,16,44-54.

Laufer, B. and Rozovski-Roitblat, B., 2011. Incidental vocabulary acquisition: The effects of task type, word occurrence and their combination. Language Teaching Research, 15(4), pp.391-411.

Leech, G., 1997. Teaching and language corpora: A convergence. *In Teaching and language corpora* (Vol. 123).

Levchenko, V., 2017. Use of Corpus-based classroom activities in developing academic awareness in doctoral students. *The New Educational Review, 48(2),* pp.28-40.

Li, J., 2010. Learning vocabulary via computer-assisted scaffolding for text processing. *Computer Assisted Language Learning, 23(3*), pp.253-275.

Liu, D. and Jiang, P., 2009. Using a corpus‐based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal, 93(1*), pp.61-78.

Ma, Q., 2013. Computer assisted vocabulary learning: Framework and tracking users' data. *Calico Journal, 30,* pp.230-243.

MacArthur, F. and Littlemore, J., 2008. A discovery approach to figurative language learning with the use of corpora. *APPLICATIONS OF COGNITIVE LINGUISTICS, 6*, p.159.

Mahboob, A. and Elyas, T., 2014. English in the kingdom of Saudi Arabia. *World Englishes, 33(1),* pp.128-142.

Mackey, A. and Marsden, E., 2015. The IRIS Repository: Advancing Research Practice and Methodology. *In Advancing Methodology and Practice* (pp. 13-34). Routledge.

McCarthy, M., 2008. Accessing and interpreting corpus information in the teacher education context. *Language Teaching, 41(4),* p.563.

Meunier, F., 2002. The pedagogical value 0f native and learner corpora in EFL grammar teaching. *Computer learner corpora, second language acquisition and foreign language teaching,* 6, p.119.

Meyer, C.F., 2009. In the profession: The "empirical tradition" in linguistics. *Journal of English Linguistics, 37(2),* pp.208-213.

Meyer, C.F., 2002. *English corpus linguistics: An introduction.* Cambridge University Press.

Mukherjee, J., 2004. Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In *Applied Corpus Linguistics* (pp. 239-250). Brill Rodopi.

Nagy, W., 2007. Metalinguistic awareness and the vocabulary-comprehension connection. Vocabulary acquisition: *Implications for reading comprehension,* pp.52-77.

Nation, I.S.P., 2001. Learning vocabulary in another language. 2003. *Cambridge. Cambridge.*

O'Dell, F., 1997. Incorporating vocabulary into the syllabus. *Vocabulary: Description, acquisition and pedagogy,* pp.258-278.

Pickard, V., 1994. Producing a concordance-based self-access vocabulary package: Some problems and solutions. *Entering text,* pp.215-226.

Porte, G.K., 2003. Selected research published in Spain 1999-2002. *Language Teaching, 36(2),* p.110.

Qian, D., 1999. Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian modern language review, 56(2),* pp.282-308.

Reppen, R., 2010. *Using corpora in the language classroom.* Cambridge University Press.

Römer, U., 2009. Corpus research and practice. *Corpora and Language Teaching. Amsterdam*: John Benjamins, pp.83-98.

Römer, U., 2010. Using general and specialized corpora in English language teaching: Past, present and future. *Corpus-based approaches to English language teaching*, pp.18-35.

Römer, U., 2011. Corpus research applications in second language teaching. *Annual review of applied linguistics, 31,* p.205.

Schmitt, N. and Schmitt, D., 2020. *Vocabulary in language teaching*. Cambridge university press.

Seidlhofer, B., 2002. Pedagogy and local learner corpora. *Computer learner corpora, second language acquisition and foreign language teaching, pp*.213-234.

Seward, D. and Boertien, V., 2012. All about us. *In Exploring Time and Place Through Play* (pp. 32-47). David Fulton Publishers.

Shillaw, J., 1994. Using a corpus to develop vocabulary tests. *Entering text,* pp.166-182.

Simpson-Vlach, R. and Ellis, N.C., 2010. An academic formulas list: New methods in phraseology research. *Applied linguistics, 31(4),* pp.487-512.

Sinclair, J., 2009. *Collins COBUILD advanced dictionary*. Heinle Cengage Learning.

Sinclair, J. M., 2003. *Reading concordances*. London: Pearson Education Limited.

Sinclair, J. M., 2004. *Corpus, Concordance, Collocation.* UK: Oxford University Press.

Slamecka, N.J. and Graf, P., 1978. The generation effect: Delineation of a phenomenon. *Journal of experimental Psychology: Human learning and Memory, 4*(6), p.592.

Sokmen (2014). "A Corpus-Based Study of Fine Arts English Vocabulary." UMT-POLY Journal: 6-10.

Sripicharn, P., 2002. *Evaluating data-driven learning: the use of classroom concordancing by Thai learners of English* (Doctoral dissertation, University of Birmingham).

Sripicharn, P., 2003. Evaluating classroom concordancing: the use of concordance-based materials by a group of Thai students. *Thammasat Review, 8(1),* pp.203-236.

Stevens, V., 1991. Classroom concordancing: Vocabulary materials derived from relevant, authentic text. *English for Specific Purposes, 10(1),* pp.35-46.

Sun, Y.C. and Wang, L.Y., 2003. Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning, 16(1),* pp.83-94.

Thurston, J. and Candlin, C.N., 1998. Concordancing and the teaching of the vocabulary of academic English. *English for specific purposes, 17(3*), pp.267-280.

Watson Todd, R., 2001. Building and using your own corpus and concordance. *ThaiTESOL buletin vol. 14 no, 2.*

Tozcu, A. and Coady, J., 2004. Successful learning of frequent vocabulary through CALL also benefits reading comprehension and speed. *Computer assisted language learning, 17(5),* pp.473-495.

Tribble, C. and Jones, G., 1997. *Concordances in the classroom: A resource guide for teachers*. Athelstan.

Tribble, C., 2001. Small corpora and teaching writing. *Small corpus studies and ELT: theory and practice, John Benjamins, Amsterdam,* pp.381-408.

Varley, S., 2009. I'll just look that up in the concordancer: integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning, 22(2),* pp.133-152.

Wang, J., Liang, S.L. and Ge, G.C., 2008. Establishment of a medical academic word list. *English for Specific Purposes, 27(4),* pp.442-458.

Webb, S., 2007. Learning word pairs and glossed sentences: The effects of a single context on vocabulary knowledge. *Language teaching research, 11(1),* pp.63-81.

Wong, J., 2002. Applying conversation analysis in applied linguistics: Evaluating dialogue in English as a second language textbooks. *International Review of Applied Linguistics in Language Teaching, 40(1),* pp.37-60.

Yoon, H. and Hirvela, A., 2004. ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing, 13(4),* pp.257-283.

Yoon, H., 2008. More than a linguistic reference: The influence of corpus technology on L2 academic writing. *Language Learning & Technology, 12(2),* pp.31-48.

Zhang, C., 2010. An overview of corpus-based studies of semantic prosody. *Asian Social Science, 6(6),* p.190.

Zhang, W., 2008. In search of English as a foreign language (EFL) teachers' knowledge of vocabulary instruction.

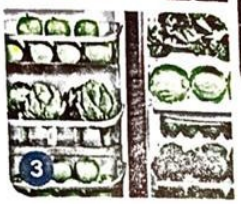
Zhen, F.C., 2005. Corpus-based data-driven foreign language learning: ideas, methods and techniques. *Foreign Language World, 4*, pp.19-27.

Zimmerman, C.B., 1997. Historical trends in second language vocabulary instruction. *Second language vocabulary acquisition, 6(1),* pp.5-19.

APPENDICES

APPENDIX A: READING TEXTBOOK CONTENT AND WORDLISTS

Table of Contents



Interactions 1 Reading, Diamond Edition
Published by McGraw-Hill ESL/ELT, a business unit of The McGraw-Hill Companies, Inc., 1221 Avenue of the Americas, New York, NY 10020.

ISBN: 978-0-07-714707-5

Senior Vice President, Products & Markets: Kurt L. Strand
Vice President, General Manager, Products & Markets: Michael J. Ryan
Vice President, Content Production & Technology Services: Kimberly Meriwether David
Director of Development: Valerie Kelemen
Director, Content Production: Sesha Bolisetty
Lead Project Manager: Rick Hecker
Senior Buyer: Michael R. McCormick
Designer: Page2, LLC
Interior Designer: Page2, LLC, Rob Richman
Senior Content Licensing Specialist: Keri Johnson
Manager, Digital Production: Janean A. Utley
Compositor: Page2, LLC, Rob Richman

www.mhhe.com

McGraw-Hill
ELT

Self-Assessment Log

Read the lists below. Check (✓) the strategies and vocabulary that you learned in this chapter. Look through the chapter or ask your instructor about the strategies and words that you do not understand.

Reading and Vocabulary-Building Strategies

- ☐ Identifying cause and effect
- ☐ Recognizing reading structure: titles and paragraph topics
- ☐ Recognizing topics and main ideas
- ☐ Summarizing a paragraph
- ☐ Getting meaning from context, parentheses, and words with similar meanings
- ☐ Focusing on high-frequency words

Target Vocabulary

Nouns

- asthma
- atmosphere
- biometeorologists
- blizzards
- carbon dioxide (CO_2)
- damage*
- desert*
- disease*
- disorder
- effects*
- floods*
- flu (influenza)
- headaches
- heart attacks*
- humidity
- hurricanes
- meteorologists
- moods
- pneumonia
- rain*
- scientists*
- storms*
- strokes
- temperatures*
- weather*

Verbs

- affect
- increase*
- influence*

Adverbs

- slowly*

Adjectives

- depressed
- extreme*
- forceful*
- moody
- nervous
- powerful*
- temperature*
- typical*
- worse*

* These words are among the 2,000 most frequently used words in English.

Self-Assessment Log

Read the lists below. Check (✓) the strategies and vocabulary that you learned in this chapter. Look through the chapter or ask your instructor about the strategies and words that you do not understand.

Reading and Vocabulary-Building Strategies

- ☐ Previewing the topic
- ☐ Previewing vocabulary
- ☐ Recognizing reading structure in a textbook
- ☐ Recognizing topics, main ideas, and supporting details
- ☐ Understanding reasons why not to do something, or negative advice
- ☐ Summarizing
- ☐ Getting meaning from context: definitions and italics
- ☐ Recognizing words with the same or similar meanings

Target Vocabulary

Nouns

- academic life
- advice
- attitudes
- challenges
- comfort
- community
- cultural background
- debt
- decision*
- excuses*
- expenses
- features
- habits
- interests
- lifestyles
- luxury
- major (area of study)
- opportunities*
- personalities
- privacy
- reasons*
- rules
- safety
- security
- strangers
- values*
- variety
- worth*

Verbs

- attend
- care
- cause*
- consider*
- get along
- get through
- miss
- stay out
- relate
- rent
- survive
- take chances
- waste

Adjectives

- afraid*
- common
- considerate
- crowded
- furnished
- homesick
- impossible
- insecure
- lonely
- miserable
- private
- shared
- similar
- terrible
- unfamiliar*
- unsafe
- valuable
- high-paying

Adverbs

- academically
- away from home

* These words are among the 2,000 most frequently used words in English.

Self-Assessment Log

Read the lists below. Check (✓) the strategies and vocabulary that you learned in this chapter. Look through the chapter or ask your instructor about the strategies and words that you do not understand.

Reading and Vocabulary-Building Strategies

- ☐ Recognizing reading structure: main-idea questions for paragraph topics
- ☐ Recognizing one- or two-sentence statements of the main idea
- ☐ Recognizing supporting details
- ☐ Matching paragraph titles with topics
- ☐ Choosing titles and recognizing topic sentences
- ☐ Summarizing
- ☐ Getting meaning from context: italics and punctuation clues
- ☐ Recognizing vocabulary categories

Target Vocabulary

Nouns

- ■ breakfast*
- ■ bugs
- ■ complex carbohydrates
- ■ customs*
- ■ dairy
- ■ desserts
- ■ diabetes
- ■ diet
- ■ discussion*
- ■ dishes*
- ■ elements
- ■ fast food
- ■ fats*
- ■ grains*
- ■ habits*
- ■ insects*
- ■ low-carb diet
- ■ minerals*
- ■ nutritionists
- ■ preferences*
- ■ soy products

Verbs

- ■ diet

Adjectives

- ■ frozen*
- ■ universal*
- ■ worse*

Adverbs

- ■ probably*

* These words are among the 2,000 most frequently used words in English.

APPENDIX B: BURAYADAH COLLEGES' CONSENT ON THE STUDY

المملكة العربية السعودية
التعليم
كليات بريدة الأهلية
قسم اللغة الإنجليزية والترجمة

كليات بريدة
Buraydah Colleges

الـرقـــــم: 5/م. خ/12
التاريـــــخ: 2019/12/30م
المشفوعات: بدون
الموضــوع: الموافقة على اجراء دراسة بحثية

APPROVAL LETTER

TO WHOM IT MAY CONCERN ***Date:*** *Mon. 30 December 2019*

This letter is an official approval for our Colleague, Mr. Ahmed Fathy Youssef, to conduct an experimental research entitled ***"Using Corpus-Linguistics in Vocabulary Learning and Retention in Saudi EFL Tertiary Classrooms",*** on Foundation Year Students, LEVEL ONE students. The study is planned to be on this semester for 5 weeks, starting exactly on Sunday 19th January 2020, 24th Jumada the First 1441H, to Thursday 5th March 2020, 10th Rajab 1441H. According to the researcher's proposal, the participants will be studying general English courses at their first year in the Colleges. Specifically, they are registered students at English 1 (ENG 101 and 100) Courses, Sections 1084 and 1089. The study will include placement tests, pre-tests and posttests, and questionnaires for participant students and English teachers.

This letter has been issued from the ***English Language and Translation Department*** at **Burayadah Colleges**, in **Qassim in Saudi Arabia**, upon the researcher's request on Wed. 25th Jan. 2020.

For any further enquiries do not hesitate to contact the Colleges Dean or the Head of English Language Department.

Head of English Department,

Dr. Fatimah Mohammed Al Naeem.

Email: Fatima.elnaeem@yahoo.com

Mob.: 00966553452678

APPENDIX C: LEARNERS' CONSENT FORM

Department of English College of Languages, SUDAN UNIVERSITY OF SCIENCE AND TECHNOLOY
P.O.B 407, Sudan, Khartoum, Phone: +249154870242

Fax: +249154934052

Email: mamood_1961@sustech.edu

Supervisor's Name & Title: Prof. Mahmoud Ali Omar

Participant Information and Consent Form

Name of The Study: **"Using Corpus-Linguistics in Vocabulary Learning and Retention in Saudi Tertiary EFL Classroom"**

My Dear Students,

I am, Ahmed Fathy Youssef a Ph. D. researcher at the College of Graduate Studies, Sudan University of Science and Technology, English Language Department, Email: concordafg@gmail.com. This research is being conducted to meet the requirements of PhD Degree of Research in the College of Languages under the supervision of Prof. Mahmoud Ali Ahmed Omar, Mobile number: +249918191919, Email Address: mamood_1961@sustech.edu. The research is entitled **"Using Corpus Linguistics in Vocabulary Learning and Retention in Saudi Tertiary EFL Classroom".**

The major objective for my study is to measure the effectiveness of using corpus-based approach in vocabulary learning and retention which has been planned to be tested through an experimental classroom study done by the researcher. In addition, there are other two objectives:1. to measure university teachers' awareness of Corpus-based approach, 2. to identify students' attitudes towards corpora integration in their vocabulary learning classes.

If you decide to participate, you will be asked to take part in the experimental (Corpus-based Learning) group or control (conventional learning) group to learn the target vocabulary in your Reading Course ENG.140. Before the study, you will be having a placement test to get an overall assessment for your language proficiency. Then, you, as participants of the study, will take part in a group you have chosen: the experimental or the control group. If you are a member in the experimental group, the assigned vocabulary will be taught through corpus-based activities, tasks and materials while the other control group will be taught using conventional methods of teaching for the same assigned course, ENG. 140 Reading Section. For the Experimental group, they will be taught mainly in computer lab using the available computers and internet connection to access corpora data for learning tasks and activities, besides using ready-made printed corpus-based materials to learn and practice the target vocabulary in each lesson. On the other hand, the experimental group will be taught using the traditional assigned classroom with a board, data

show, depending mainly on the assigned Coursebook exercises and activities and printed worksheets based mainly on the course book design, Interaction 1 Reading Book 1 by Elaine Kirn and Pamela Hartmann. The study will run for six weeks. At the early beginning of first week of the study, you will be having a pre-test and at the end of the sixth week you will be having a post-test to assess your gains from the study.

Any information or personal details gathered in the course of the study are confidential, except as required by law. No individual will be identified in any publication of the results. Only the researcher will access the data for the purpose of the study. Your participation is highly valued and will only be used for the purpose of the current research. A summary of the results of the study can be made available to you on request. You can contact the researcher on his email to have the result of this study.

Participation in this study is entirely voluntary: you are not obliged to participate and if you decide to participate, you are free to withdraw at any time without having to give a reason and without any consequences. No extra marks will be given for your participation in this study and all quizzes and tests will not be count for your assessment in the course ENG. 140, Reading Section.

My Regards,

The Researcher.

I, ……. have read and understand the information above and my questions and enquiries have been fully answered to my satisfaction. I have had an opportunity to ask questions, and agree to participate in this research study. I have given all the aforementioned details and the proposed course plan and I have been given a copy of this consent form to keep.

Student's Name:

Student's Signature: **Date**

APPENDIX D: DETAILED PLAN FOR THE CLASSROOM TREATMENT

Course Plan for The Classroom Treatment

A. Introduction

This course plan is mainly designed as an essential part of the experimental study conducted by ***Mr. Ahmed Youssef***, ***PhD Student at Sudan University of Science and Technology***, on two groups of students studying a general English course, ENG. 140. This plan is intended to be a detailed description for all intended learning outcomes, teaching strategies, assessment methods, and time table for all study procedures. All students in both groups have to get a copy from this plan at an early stage of the experiment.

B. Specific Dates and Information about the Study

Course Teacher	Ahmed Youssef	**College/Department**	Humanities and Administrative Sciences
Email	concordfg@gmail.com	**Course Code**	ENG. 140
Mobile number	0546511761	**Level**	1
Credit Hours	8 hr.	**Semester**	Second
Starting Date	24th Jumada 1st, 1441H 19 January, 2020	**End Date**	3rd Rajab, 1441 H. 27th February, 2020
Placement test and pretest Date	Placement T.: Sunday 19 January, 2020 Pretest: Monday, 20 January, 2020	**Post Test Date**	On Thursday, the 3rd of Rajab, 1441 H., 27th February, 2020
Delayed Post Test	On Thursday the 17th of Rajab, 1441 H. 12th March, 2020	**Students' Questionnaires and Interview**	On Sunday, the 20th of Rajab, 1441 H. 15th March, 2020
Experimental Group's Students	29 male students	**Control Group's Students**	35 ale students

C. The Researcher's Main Objectives

The researcher intended to:

1. Measure the effectiveness of using corpus-based approach for learning vocabulary.
2. Measure the effectiveness of using corpus-based approach for vocabulary retention.
3. Investigate students' perceptions on using corpus-based approach for vocabulary learning and retention.
4. Investigate the challenges faced by students through using corpora rather than traditional dictionaries for learning the target vocabulary.

D. Course Learning Outcomes (CLOs)

1. Knowledge
2. Cognitive
3. Interpersonal Skills and Responsibility.
4. IT and Communication Skills

At the end of this course, the student should be able to:

Domain	Course Learning Outcomes (CLOs)	Teaching Strategies and Learning Resources	
		Control Group	Experimental Group
4. Knowledge	4.1 Identify the meaning, definitions, and synonyms for the target words. 4.2 Identify the appropriate collocates for some target vocabulary in real contexts. 4.3 Recognize the correct spelling, pronunciation, part of speech, and use of the target vocabulary in each chapter	• Using the Traditional Classroom with its normal facilities to deliver lessons. • Traditional lecturing approach and other context and definitions strategies will be used for words teaching	• Using the Department's Computer Lab to deliver all the vocabulary classes using available PCs and internet connection. • Discovery learning and autonomous problem-solving approaches will be used to learn target words.
5. Cognitive	5.1 Utilize new learnt vocabulary in authentic contexts and personal situations. 5.2 Demonstrate high level of awareness for words frequency in different genres. 5.3 Utilize word clusters rather than single isolated words. 5.4 Employ words collocation for producing spoken or written utterances.	• Using the assigned Course book, Interactions 1, Reading Section, by *E. Hirn and Hartmann*, to teach the target vocabulary in the first three chapters. • Traditional lecturing approach and other context and definitions strategies will be used for words teaching	• Using the assigned course book for teaching reading strategies only, and using COCA Corpus to learn the target vocabulary in each chapter. • Using printouts and hands-on mainly derived from COCA Corpus Concordance lines and queries data. • Discovery learning and autonomous problem-solving approaches will be used to learn target words.
6. Interpersonal Skills and	3.1 Students are able to accomplish the assigned tasks and appropriately fulfill their	• Using work sheets, slides and activities mainly based on the	• Using printouts and hands-on mainly derived from COCA Corpus

Responsibility	role in their team, contribute to others' work, and solve their team problems.	course book activities and exercises. • Students are given pair work and group works tasks based on the course book activities and exercises.	Concordance lines and queries data. • Students are given pair work and group works tasks based on the online use of corpora searching tools and concordancers.
4. **IT and Communication Skills**	4.1 Students are able to use different forms of IT tools appropriately to assist learning and researching, and to give presentations. 4.2 Students are able to use online dictionaries and online corpora to learn target vocabulary.	• The teachers will be using the available projector and personal lab top in the class to give them presentations and demos in using online dictionaries.	• All students will be using the available PCs or their own smart phones to get direct access to online corpora use and doing queries about the target vocab.

E. Topics to be Covered

No.	Topics to be Covered	Number of Assigned Vocabulary/ per week	Week No.	Number of Hours
13	Orientation for the Study, Placement Tests, and Pretests	---	1st	4 hr.
14	Chapter 1: Academic Life around the World. (Vocabulary)	20	1st	4 hr.
15	Chapter 1: Academic Life around the World. (Vocabulary)	41	2nd	6 hr.
16	Chapter 2 Experiencing Nature. (Vocabulary)	16	3rd	6 hr.
17	Chapter 2 Experiencing Nature. (Vocabulary)	20	4th	6 hr.
18	Chapter 3: Living to Eat, or Eating to Live. (Vocabulary)	17	5th	6 hr.
19	Chapter 3: Living to Eat, or Eating to Live. (Vocabulary)	12	6th	4 hr.
20	**IMMEDIATE POSTTEST**	----	6th	2 hr.
21	**INTERVAL**	---	7th	--
22	**INTERVAL**	---	8th	--
23	**DELAYED POSTTEST**	---	9th	2 hr.
24	**Total Weeks and the Teaching Actual Hours**	***116 words***	**6 weeks**	**40 ours**

F. Study Plan's Summary

Week No.	Commencement Date	Topics	Targeted Objectives	Assigned Tests
1.	**هـ1441/5/24**	**Students' Orientation for the Study** **Chapter 1:** Academic Life around the World. (Reading)	• T. Getting students familiarized with all steps, stages, procedures and tests planned in the study. • T. Assessing their general proficiency level through standardized placement test. (Cambridge General Placement Test) • T. Assessing their previous knowledge about the target vocabulary. • Sts. getting meaning from context; definitions and italics. • Sts. identifying the words with the same or similar meanings. • Sts. Identifying parts of speech for the target words.	**Placement Tests - Pretests**
2.	**هـ1441/6/1**	**Chapter 1:** Academic Life around the World. (Reading)	• Focusing on high-frequency words. • Identifying differences between genres. • Using the learnt words in real situations and personal sentences.	**Worksheets and Assignments**
3.	**هـ1441/6/8**	**Chapter 2**: Experiencing Nature. (Reading)	• Finding vocabulary clusters. • Identifying synonyms or examples for the target vocabulary.	**Worksheets and Assignments**
4.	**هـ1441/6/15**	**Chapter 2**: Experiencing Nature. (Reading)	• Getting meaning from the context, parentheses or similar meanings. • Identifying the frequent collocates of some target vocabulary. • Identifying the different parts of speech for the	**Worksheets and Assignments**

			single word.	
5.	**1440/6/22**	**Chapter 3:** Living to Eat, or Eating to Live?	• Getting key words meanings from the text context. • Using the learnt words in real situations and personal sentences.	**Worksheets and Assignments**
6.	**1441/6/29**	**Chapter 3:** Living to Eat, or Eating to Live?	• Looking up words using monolingual resources. • Participating actively in personalized learning tasks to use the target vocab.	**Posttests**

G. Students' Assessment Plan

Students' Assessment Plan Overview			
No.	***Assessment Methods***	***Assigned Week***	***Intended Aim***
1.	**Placement Test (Cambridge General Placement Test) (50 Marks)**	The Beginning of the 1st week	Categorizing Students into their expected levels
2.	**Pretest (50 Marks)**	The second day of the 1st week	To measure students' knowledge about the target vocabulary
3.	**Posttest (50 Marks)**	Thursday, at the end of the 5th week	To measure students gains after going through the study for 5 weeks
4.	**Delayed posttest (50 Marks)**	At the end of the 8th week	To measure students' retention rate for the learnt vocabulary after an interval of 3 weeks.
5.	**Experimental Group Learners' Questionnaire (50 Marks)**	At the beginning of the 9th week.	To investigate Experimental Group's learners' experience with Corpus-based approach in learning vocabulary

References:

Kirn, E. &Hartmann, P. (2012): ***Interactions*** 1(Reading) Diamond Edition, McGraw-Hill Inc., New York, NY 10020.

Course Specifications for ENG. 140 (2018). Foundation Courses Unit. Quality Assurance Unit. Buraydah Private Colleges.

McCarthy, M.l and F. O'Dell, "English Vocabulary in Use (Elementary)", Cambridge University Press, New York, 2010.

COCA (2020). Corpus of Contemporary America English. Retrieval link (www.English-corpora.org/COCA/)

The Study's Teacher/ Researcher **Foundation Courses Coordinator**

Ahmed Youssef **Dr. Mohamed Abd El-Motaal**

Outcomes

Placement Test

Contents

INTRODUCTION

The ***Outcomes*** **Placement Test Package** has been developed to help course providers place students in the most appropriate level of the *Outcomes* course.

Prior to testing

In order to adapt coursework to the needs, expectations and skill levels of students, it is generally advantageous to find out as much as possible about each individual student during the initial assessment process. To this end, the **Personal Data Form** and **Learning Profile Survey** may be photocopied and used as they are, or translated and/or otherwise adapted in order to gather important information about students.

Each student should be asked to fill out a copy of one or both of the forms before taking the test. It is important to explain that this information is being collected both for administrative purposes and so that teachers can better adapt classroom materials and techniques to suit the learners' needs; students should also be assured that all information on the forms will be kept strictly confidential.

The Outcomes Placement Test Package

The *Outcomes* Placement Test Package includes three types of tests. In addition to the core **Placement Test** there are also separate **Oral** and **Writing Placement Tests**.

Placement Test

The **Placement Test** consists of 50 items testing grammar and vocabulary presented and practiced over the whole range of the *Outcomes* series.

An **Answer Key** to the test is provided.

Oral Placement Test

A separate **Oral Placement Test** is included for course providers who want to utilize this form of assessment as part of their placement process. This test consists of an interview based on conversational prompts to be presented by the interviewer, and is accompanied by **Oral Assessment Guidelines** to help interviewers determine the appropriate level of the student.

Writing Placement Test

A separate **Writing Placement Test** is included for course providers who want to incorporate writing into their placement process. Students are asked to select a writing task and complete the task using the prompts provided in the rubric. **Writing Assessment Guidelines** are provided to help test evaluators determine the appropriate level of the student.

TEST ADMINISTRATION, SCORING AND RESULTS

The *Outcomes* Placement Test Package allows course providers to consider a range of options in determining the appropriate level of each student, and it has been designed so that course providers may choose the method that best suits their needs.

The most objective test, and the most straightforward to administer, is the **Placement Test**. The **Oral** and **Writing Placement Tests** are open-ended and, therefore, more subjective in nature. Some course providers may wish to use only the Placement Test in determining the appropriate level for each student, while others may wish to use it in combination with the Oral and/or Writing Placement Tests.

The charts on the following pages show suggested placement levels related to each of the different tests in the *Outcomes* Placement Test Package. However, depending on the curricular objectives of the course, evaluators may give unequal weighting to scores on the various placement tests. For a course that emphasizes speaking skills, for example, they may decide to weight the score on the Oral Placement Test more heavily than that on the Writing Placement Test, while for a course that emphasizes writing skills, they might consider doing the opposite.

Placement Test

Administration

Each student should be provided with a photocopy of the **Placement Test**. Test administrators may prefer students to mark their answers on the photocopy of the test itself. Alternatively, each student may be provided with a photocopy of the **Student Answer Sheet** *(page 14)* on which to mark their responses.

The test administrator should read the instructions aloud and make sure that students know where and how to record their answers (for instance, whether to mark their answers on the question paper or to use the Student Answer Sheet). Students should then be given 30 minutes to complete the written portion (items 1-50) of the Placement Test.

Scoring

Students are awarded one point for each correct answer, according to the **Answer Key** provided *(page 16)*.

Results

The chart below suggests which level of the *Outcomes* course would be most appropriate for a particular student, based on that student's total score on the Placement Test.

Placement Test Results	
Placement Test score **(50 items: 1 point per item)**	**Recommended level of *Outcomes***
0 – 18 points	*Outcomes Elementary*
19 – 25 points	*Outcomes Pre-Intermediate*
26 – 32 points	*Outcomes Intermediate*
33 – 39 points	*Outcomes Upper Intermediate*
40 – 46 points	*Outcomes Advanced*
47 – 50 points	Higher level series recommended. Visit www.heinle.com for suggested titles.

Overall Placement Considerations

In cases where a student's scores on each of the various tests indicate similar levels, placement is relatively straightforward. However, in cases where the student scores significantly higher on one test and lower on another, the evaluator(s) will have to decide how to weight the scores on the various tests to determine the student's final placement level.

For example, if a student scores at the lower end of *Outcomes Intermediate* on the Placement Test and near the middle of *Outcomes Pre-Intermediate* on the Oral or Writing Placement Test, the evaluator(s) should take the objectives of the course into consideration in determining which level the student should be placed in. If the main objective is to develop fluency in conversational skills and the Oral Placement Test score is lower, it might be best to place the student in the lower level.

Ongoing and End-of-Term Assessment in *Outcomes*

In addition to the *Outcomes Placement Test Package*, the *Outcomes Assessment CD-ROM with Exam View*® contains banks of test items that teachers can use to create specialized exams for each course. For more information, see the *Outcomes Assessment CD-ROM with Exam View*®.

PERSONAL DATA FORM

Today's Date: _____

Name: _____
Last First Middle Initial

Birth Date: _____ Age at last birthday: _____
Month / Day / Year

School / Work: _____

Home Address: _____
Number Street Apartment Number

_____
City / Town State / Province / etc. Postal Code

Telephone: _____ _____
Home Mobile

For emergencies, call: _____ _____
Name Telephone Number

Signature of Student: _____

The student completed this form: (Circle one) Alone With assistance

Signature of person helping student provide information: _____

LEARNING PROFILE SURVEY

1 In your language, are reading and writing easy or difficult for you? _____

2 How did you learn to read and write? (Circle one.) school family tutor other

3 Years of education: (Circle last year completed.) 5 6 7 8 9 10 11 12 13 14 15 16
(elementary) (secondary) (university)

4 Other languages you speak: _____

5 Other languages you read or write: _____

6. What type of materials do you like to read in your language?
_____

7 What do you write in your language? (Circle one.)
emails stories list notes other _____

8 Have you studied English before? Yes / No Where? _____

9 How long have you studied English? _____

10 Describe your ability in English: (Circle the best description.)

Understanding is ...	always difficult	sometimes difficult	never difficult
Speaking is ...	always difficult	sometimes difficult	never difficult
Reading is ...	always difficult	sometimes difficult	never difficult
Writing is ...	always difficult	sometimes difficult	never difficult

11 Why do you want to learn English? _____

12 What is the most important thing you want to learn?
_____

13 What questions do you have about this course? _____

The student completed this form: (Circle one) Alone With assistance
Signature of person helping student provide information: _____

Name: _____ **Date:**_____

PLACEMENT TEST

Circle the correct letter.

1 I'm 18 and my brother is 20, so he's ..... me.
a the oldest of **b** older than
c as old as

2 Carl's very ..... . He's never late, and he never forgets to do things.
a reliable **b** patient
c strict

3 We stayed in a lovely villa ..... the sea.
a it overlooks **b** overlooked
c overlooking

4 Not until the 1980s ..... for the average person to own a computer. **a** it was possible
b was it possible
c was possible

5 Jan ..... her arm on a hot iron.
a broke **b** burned
c sprained

6 Tomorrow's a holiday, so we ..... go to work.
a have to **b** mustn't
c don't have to

7 I usually ..... swimming at least once a week.
a go **b** do **c** play

8 My friend Siena ..... to Russia last year.
a went **b** has gone **c** has been

9. This is ..... area, with a lot of factories and warehouses. **a** an agricultural **b** an industrial
c a residential

10. If I ..... well in my exams, I ..... to university.
a will do; will go **b** will do; go
c do; will go

11. She was so upset that she burst ..... tears.
a into **b** out **c** with

12. Where did you go ..... holiday last year?
a for **b** on **c** to

13. Ocean currents ..... play an important part in regulating global climate. **a** are known to **b** thought to
ci are believed that they

14. My cousin ..... getting a job in Bahrain. **a** would like **b** is planning
a. is thinking of

15. I can't ..... your hair, because I haven't got any scissors.
a brush **b** cut **c** wash

16. I wish I ..... have an exam tomorrow! **a** don't **b** didn't **c** won't

17. The government plans to ..... taxes on sales of luxury items. **a** increase **b** expand
a. go up

18. When I first moved to Hong Kong, life in a different country was very strange, but now I'm used ..... here.
a living **b** to live **c** to living

19. There ..... milk in the fridge.
a is some **b** are some **c** is a

20. Criminals are people who are guilty of ..... the law.
a breaking **b** cheating
a. committing

21. Why on earth isn't Josh here yet? ..... for him for over an hour! **a** I'm waiting **b** I've been waiting **c** I've waited

22. "It's pouring down, and it's freezing." What are the weather conditions? **a** high winds and snow **b** heavy rain and cold temperatures

a. thick cloud but quite warm

23. ..... feeling OK? You don't look very well.
a Do you **b** You are **c** Are you

24. Daniel's hair is getting far too long; he should ..... soon.
a cut it **b** have cut it **c** have it cut

25. Mandy works for a computer software company. She got ..... recently, and so now she's an area manager. **a** made redundant **b** promoted
a. a raise

26. I can't hear you – it's ..... noisy in here.
a too **b** too much **c** too many

27. Jamal has just sent me ..... to arrange plans for this weekend. **a** a blog **b** an email
a. a website

28. I promise I'll call you as soon as I ..... . **a** I arrive **b** I arrived
a. I'll arrive

29. Photographers and designers need to be very ..... .
a creative **b** fit **c** annoying

30. The global financial crisis, ... is forcing lots of small businesses to close, does not look set to end soon. **a** it **b** that **c** which

31. There ..... a terrible accident if the pilot hadn't reacted so quickly. **a** had been **b** was **c** would have been

32. "Are you ready to order?"
"Not yet – I'm still looking at the ..... ." **a** bill **b** menu **c**. service

33. "My job is never boring."
The speaker's job is always ..... . **a** interesting **b** popular
a. difficult

34. I've been working here ..... about the last two years. **a** during **b** for **c.** since

35. "It leaves from Platform 2 at 4.15."
The speaker is talking about ..... .
a an airline flight **b** a train **c** a taxi

36. I went to a lovely ..... last Saturday. The bride was my best friend when we were at school. **a** anniversary **b** marriage **c** wedding

37. "I've got a headache." "Maybe you ..... to take an aspirin." **a** should **b** ought **c** don't

38. The patient had an ..... to insert metal pins in his broken leg. **a** injection **b** operation
a. X-ray

39. She won a seat in parliament at the last ..... .
a general election **b** opinion poll
a. referendum

40. I'm surprised you didn't get upset. If someone said that to me, ... really angry. **a** I'm **b** I was **c** I'd be

41. This used to be ..... part of the city, but since the old buildings were renovated it's become a very fashionable area.
a an affluent **b** a run-down
a. a trendy

42. Cassie went to bed early because she was ..... .
a tired **b** stressed **c** relaxed

43. In the 1960s, computers were ..... expensive that ordinary people couldn't afford them.
a so **b** such
a. too

44. Do you want ..... the match tonight? **a** watching **b** watch

a. to watch

45. Researchers claim the new discovery is a major ..... in the fight against malaria.
a breakthrough **b** investigation
a. progress

46. The Maths problem was really difficult and I just couldn't ..... the answer.
a check in **b** set off **c** work out

47. When I was a child, I never ..... about the future.
a have worried **b** used to worry **c** was worrying

48. A local politician has ..... charges of corruption made by the opposition party. **a** accused **b** blamed **c** denied

49. ..... worries me about society today is how completely we have come to depend on technology.
a That **b** What
a. Which

50. Cats and dogs are usually kept as ..... . **a** farm animals **b** wild animals **c** pet

STUDENT ANSWER SHEET

PLACEMENT TEST

Circle the correct letter.

1	**a**	**b**	**c**	**11**	**a**	**b**	**c**	**21**	**a**	**b**	**c**	**31**	**a**	**b**	**c**	**41**	**a**	**b**	**c**
2	**a**	**b**	**c**	**12**	**a**	**b**	**c**	**22**	**a**	**b**	**c**	**32**	**a**	**b**	**c**	**42**	**a**	**b**	**c**
3	**a**	**b**	**c**	**13**	**a**	**b**	**c**	**23**	**a**	**b**	**c**	**33**	**a**	**b**	**c**	**43**	**a**	**b**	**c**
4	**a**	**b**	**c**	**14**	**a**	**b**	**c**	**24**	**a**	**b**	**c**	**34**	**a**	**b**	**c**	**44**	**a**	**b**	**c**
5	**a**	**b**	**c**	**15**	**a**	**b**	**c**	**25**	**a**	**b**	**c**	**35**	**a**	**b**	**c**	**45**	**a**	**b**	**c**
6	**a**	**b**	**c**	**16**	**a**	**b**	**c**	**26**	**a**	**b**	**c**	**36**	**a**	**b**	**c**	**46**	**a**	**b**	**c**
7	**a**	**b**	**c**	**17**	**a**	**b**	**c**	**27**	**a**	**b**	**c**	**37**	**a**	**b**	**c**	**47**	**a**	**b**	**c**
8	**a**	**b**	**c**	**18**	**a**	**b**	**c**	**28**	**a**	**b**	**c**	**38**	**a**	**b**	**c**	**48**	**a**	**b**	**c**
9	**a**	**b**	**c**	**19**	**a**	**b**	**c**	**29**	**a**	**b**	**c**	**39**	**a**	**b**	**c**	**49**	**a**	**b**	**c**
10	**a**	**b**	**c**	**20**	**a**	**b**	**c**	**30**	**a**	**b**	**c**	**40**	**a**	**b**	**c**	**50**	**a**	**b**	**c**

ANSWER KEY

1	b	11	a	21	b	31	c	41	b
2	a	12	b	22	b	32	b	42	a
3	c	13	a	23	c	33	a	43	a
4	b	14	c	24	c	34	b	44	c
5	b	15	b	25	b	35	b	45	a
6	c	16	b	26	a	36	c	46	c
7	a	17	a	27	b	37	b	47	b
8	a	18	c	28	a	38	b	48	c
9	b	19	a	29	a	39	a	49	b
10	c	20	a	30	c	40	c	50	c

APPENDIX F: PRE- AND POST- TESTS

Receptive and Productive Pretest

Name: ---------- **Group:** Experimental ☐ Control ☐

PART ONE: WORDS' ***DEFINITIONS, MEANINGS, SYNONYMS AND SPELLING. (Receptive or Definitional Vocabulary Knowledge)***

A. **Match each word or phrase in Column (A) with its definition or meaning in Column (B):**
NOTE: EACH WORD equals ONE POINT and ***there are TWO EXTRA answers in Column (B).***

COLUMN (A)	COLUMN (B)
1. Excuses ( )	a. Blood vessel attack in the brain.
2. Opportunities ( )	b. The air around the earth.
3. Miss ( )	c. Of enough value for the money.
4. Memorable ( )	d. Explanations that aren't good reasons.
5. Worth the cost ( )	e. Fail to have.
6. Miserable ( )	f. Worth remembering.
7. Attend college ( )	g. Chances to do something.
8. Get along ( )	h. Be friendly with someone.
9. Atmosphere ( )	i. Go to university.
10. bio meteorologist ( )	j. The researcher who studies health and emotions in response to atmospheric conditions.
11. a stroke ( )	k. Very unhappy or uncomfortable.
12. moods ( )	l. Illnesses like pneumonia. Asthma, headaches...etc.
13. extreme weather ( )	m. Examples: tornadoes, storms, floods, hurricanes.
14. diseases ( )	n. A synonym for the word feelings
15. **universal** ( )	o. Another word for "worldwide".
	p. A blood sugar disorder.
	q. Prepared items from an inexpensive restaurant, snack bars, or food stands.

B. Write the rest of these words for the explanations on the lines: *ONE mark for each*

1. Dorm……. is a kind of shared living and learning community on a college or university.

2. Higher **ed**……. is a kind of studying at a college or university.

3. Housing **coop**…… is a housing with shared work and decision making.

4. According to **meteo**…… the earth's climate is changing continuously.

5. A blood sugar disorder is a **dia**……

C. There are three words of similar or related meaning, CROSS the odd word OUT:

1. waste – lose – high paid – use up.
2. participation– problems – trouble – challenges.
3. academically - terrible – unpleasant – miserable.
4. have an effect on – condition – affect – influence.
5. fall – go down – get lower – increase.

TOTAL POINTS FOR THIS PART EQUAL 25.

End of PART ONE (definitional or receptive knowledge)

PART TWO: WORDS AUTHENTIC CONTEXTUALIZATION, VOCABULARY USE, AND COLLOCATIONAL KNOWLEDGE. (Productive Vocabulary or Transferring Knowledge).

A. Read the following concordance lines (sentences) and use one of the following words to fill in the gaps in each, according to the provided context: *1 Point for Each*

(Health disorders - blizzards – weather - in contrast to - bio meteorologists – nutritionists – diabetes – probably – influence – privacy – interests – slowly - academically – attend – worth the cost)

1. In that private schools, students were supposed to do a little better, and their parents to feel that this was ----------, there's a business to be had there. #
2. violence and insecurity had become " normalized " and that " mental ---------- such as anxiety, depression, or post-traumatic stress " could manifest themselves once the..
3. Though preliminary, scientific studies back up the notion that weather affects your health, says a ----------, ,professor at Utah State University in Logan
4. Patient lose weight and hand him a diet plan -- or refer him to a ----------. At the Family Health clinics,
5. I agree to the beliefnet.com terms of service, rules of conduct and ---------- policy (the " agreements ").
6. ice has led to unusual warming of the Arctic atmosphere, that in turn impacts ----------conditions in the northern hemisphere
7. the day began cold and blustery, ----------almost two weeks of uninterrupted sunshine
8. around the world will be forced to face flooding and drought in the summer and ---------- in the winter
9. for funding from the American Government, to get a free ride to ---------- College.
10. A shorter school year, and longer summer, means that more students will struggle ---------- and teachers will be forced to spend more time on catch-up and remedial..
11. as women die of diseases that could be largely prevented, like lung cancer and ----------, or treated effectively, like heart disease.
12. whether this reward may in a way ---------- the blogger to be biased in his or her post.
13. This is a very unusual and serious situation and we have to take the matter very ----------and carefully.
14. I feel as if they actually put the customer first and that everything they do is in our ----------
15. It ---------- indicates that he has little or no motivation elsewhere in life, like you assumed..

B. Read the following concordance lines (sentences) and choose ONE WORD that comes with of them: *1 point for each gap*

1. Georgia State University has four new ---------- arranged in a traditional quad south of Georgia Tech near the…

- In the spring, just before I moved from the ---------- into an apartment across town…

a. City/ies **b. college/s** **c. dormitory/ies** **d. place/s**

2. … you a lot more than you can fear them. Everyone has different values and ---------- (emotions) toward wolves.

YES! Wasn't there even a post on this blog about how ---------- about college are changing?

- There's a different ---------- to police over there.

a. Challenge/s **b. opportunity/ies** **c. interest/s** **d. attitude/s**

3. then no one will have the ---------- that they didn't know

biggest problem is I blame others for the way that I am, I make ---------- . Well, I'm that way because..

you've got to make it easy for yourself and get rid of all the ---------- if you want to keep going.

a. excuse/s **b. opportunity/ies** **c. function/s** **d. hope/s**

4. That study also found no reduction in ---------- among women who received the study drug.

(Blog twitter) had and was willing to leave their Job was truly a ---------- of luck.

a. asthma **b. opportunity/ies** **c. headache/s** **d. pneumonia**

TOTAL POINTS FOR THIS PART EQUAL 25.

End of PART TWO (Transfer or Productive knowledge of Vocabulary)

APPENDIX G: LEARNERS' QUESTIONNAIRE ON CORPUS-BASED APPROACH

Experimental Group Students' Questionnaire on Using Corpus-based Approach/ COCA Corpus for Vocabulary Learning and Retention

Dear My Students,

This survey aims to ascertain your assessment of your learning experience in the course, especially what you have learned from the use of corpus research, contextualization, and COCA Corpus. This questionnaire is used to collect data for research concerning the issue of employing online corpora like COCA or BNC or any other online corpora in vocabulary learning and retention in EFL classrooms. The information gained by this questionnaire will be ONLY used for this purpose.

Your cooperation is highly appreciated.

My Regards,
The Researcher.
Ahmed Youssef

YOU ARE SUPPOSED TO ANSWER THIS SURVEY QUESTIONS ACCORDING TO YOUR OWN EXPERIENCE WITH THE 6-WEEKS EXPERIMENT THAT YOU'VE GONETHROUGH FROM WEEK 1st TO WEEK 6th IN THE 2nd SEMESTER, ACADEMIC YEAR 1440-1441.

PART 1: Personal Language and Technology Backgrounds

1. Your Name (Optional) ……

2. How long have you been Learning English? *

Mark only one oval.

- ◯ For 1 - 5 years
- ◯ For 6 - 10 years
- ◯ For 11 - 15 years
- ◯ For more than 16 years

3. How do you describe your skills in computer and technology? *

- ◯ Expert user
- ◯ Very good user
- ◯ Good user
- ◯

Bad User

4. Do you have access to a computer and internet connection at your home and school?

Mark only one oval.

- Always
- Often
- Sometimes
- Rarely
- Never

PART 2: Read the following statements and choose the closest response to you, from Strongly Agree (5) to Strongly Disagree (1) according to your personal actual experiment in using COCA Corpus in Course ENG.140.

5. 5. I feel that the use of corpora (COCA Corpus) was helpful for me in learning vocabulary in the course ENG.140? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree).

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

6. 6. I feel that using corpus-based activities /COCA based activities and printouts in learning vocabulary is more helpful and beneficial than using conventional methods like dictionaries and coursebooks reading texts? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark one.

	1	2	3	4	5	
Strongly Disagree						Strongly Agree

7. 7. I feel that using Corpora and being exposed to massive authentic input contributes positively in learning many aspects and linguistic elements of a word? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

8. 8. I feel that being involved in corpora individual activities and group tasks increases my vocabulary retention? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

9. 9. I generally feel comfortable to use concordances lines to learn new words and to be involved in the experimental group for a 5-weeks period? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

10. 10. After going through this study, I realized the importance of context and co-occurrences provided by the corpora queries to effectively learn new words, rather than using dictionaries. *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

11. 11. I think that I will use Corpora or COCA Corpus at my own in the future to learn new words or look up words? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

Challenges or Difficulties in Using Corpora.

12. 12. I feel that the queries concordance lines or context examples are overwhelming and may cause frustration or distraction especially? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

13. 13. I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily? *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

1	2	3	4	5

Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

14. 14. I feel that my language proficiency level impacts my ability to effectively use corpus-based printouts and activities. *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

15. 15. I have some difficulty in using the COCA corpus due to cut-off sentences in some concordance outputs. *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

16. 16. I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance outputs. *

1(strongly disagree) - 2 (disagree) - 3 (neutral) - 4 (agree) - 5 (strongly agree)

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	◯	◯	◯	◯	◯	Strongly Agree

APPENDIX H: TEACHERS' QUESTIONNAIRE ON CORPUS-BASED APPROACH

University English Teachers' Questionnaire on Using Corpus-based Approach/ Online Corpora in Teaching Vocabulary

My Dear Colleagues,

This questionnaire is used to collect data for a research concerning the issue of employing online corpora like COCA or BNC or any other online corpora in teaching vocabulary in Saudi EFL classroom. The information gained by this questionnaire will be used for this purpose ONLY.

Your cooperation is highly appreciated.

My Regards,
A. Youssef

N.B. ALL TEACHERS ARE REQUIRED TO ANSWER THESE QUESTIONS ACCORDING TO THEIR OWN TEACHING, TRAINING, AND PROFESSIONAL EXPERIENCE.

PART 1: PROFESSIONAL AND TECHNOLOGICAL BACKGROUNDS

1. What is your highest qualification? *

- ◯ Bachelor Degree
- ◯ Master's Degree
- ◯ PhD Degree

2. Your Gender *

- ◯ Female
- ◯ Male
- ◯ Prefer not to say

3. How long have you been teaching English for university students? *

- ◯ For 1 - 5 years
- ◯ For 6 - 10 years
- ◯ For 11 - 15 years
- For more than 16 years

4. 4. How do you define your level in using technology in your classroom? *

Mark only one oval.

- ◯ Limited
- ◯ Acceptable
- ◯ Good
- ◯ Professional

PART TWO: READ THE FOLLOWING STATEMENTS AND CHOOSE THE CLOSEST RESPONSE TO YOUR EXEPRIENCE.

5. Have you ever heard or read or learnt about the term CORPUS LINGUISTICS or CORPUS-BASED or ONLINE CORPORA or any other Corpus-linguistics terms?

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

	1	2	3	4	5	
Never	◯	◯	◯	◯	◯	Always

6. How often have you ever been trained during your study at your university on using corpus linguistics approach in teaching or research purposes? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

	1	2	3	4	5	
Never	◯	◯	◯	◯	◯	Always

7. How often have you ever attended a workshop or a training program on using online corpora in classrooms? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

Mark only one oval.

	1	2	3	4	5	
Never	◯	◯	◯	◯	◯	Always

8. How often have your used corpus-based approach in your classroom or for research purposes? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

Mark only one oval.

Never	◯	◯	◯	◯	◯	Always

9. Are your school/university interested in having accounts on or premium access to an online corpus, or purchasing any concordancing software programs like Wordsmith or AntConc? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

Mark only one oval.

	1	2	3	4	5	
Never	◯	◯	◯	◯	◯	Always

Part Three: Read the following statements and choose the closet response to your views on using corpus linguistics

10. Do you think it is beneficial to integrate corpus-based activities in EFL classrooms? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

Mark only one oval.

	1	2	3	4	5	
Never	◯	◯	◯	◯	◯	Always

11. Do you think that Corpus-Linguistics has to be integrated in teacher training programs in your country? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

Mark only one oval.

1 2 3 4 5

Never	◯	◯	◯	◯	◯	Always

12. Do you think that lack of technological equipment in your school/ university prevents you from using online corpora in your classes? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

	1	2	3	4	5	
Never	◯	◯	◯	◯	◯	Always

13. Do you believe that students' low technological skills may hinder integrating corpus-based activities in the classroom? *

1(never)- 2 (rarely) - 3 (sometimes) - 4 (often) - 5 (always)

Mark only one oval.

1	2	3	4	5
◯	◯	◯	◯	◯

14. In your opinion what are the appropriate levels that can have direct access to online corpora to learn vocabulary? *

Check all that apply.

- ☐ Beginners
- ☐ Pre-Intermediates
- ☐ Intermediates
- ☐ Advanced
- ☐ I don't know

APPENDIX I: THE JURY OF THE RESEARCH DATA TOOLS

(A): CONSENT FORM

Respected Juror/ Expert,

I am, Ahmed Fathy Youssef a Ph. D. researcher at the College of Graduate Studies, University of Sudan for Science and Technology, English Language Department, Email: concordafg@gmail.com. This research is being conducted to meet the requirements of PhD Degree of Research in the College of Languages under the supervision of Prof. Mahmoud Ali Ahmed Omar, Mobile number: +249918191919, Email Address: mamood_1961@sustech.edu. The research is entitled ***"Using Corpus Linguistics in Vocabulary Learning and Retention in EFL Saudi Tertiary Classroom".***

The major objective for my study is to measure the effectiveness of using corpus-based approach in vocabulary learning and retention which has been planned to be tested through an experimental classroom study done by the researcher. In addition, there are other two objectives:1. to measure university teachers' awareness of Corpus-based approach, 2. to identify students' attitudes towards corpora integration in their vocabulary learning classes. The attached documents are two questionnaires, one for learners and the other for teachers, and a proposed plan for 6-weeks teaching for both groups.

If you are kindly accepted to participate as an expert or juror, Then you're kindly requested to: 1. Sign the consent from, 2. Check the validity of the items of the attached questionnaires, and write any further recommendations or suggestions to make sure that each questionnaire is valid and reliable in measuring what it is specifically set to measure, 3. Assess the appropriateness of the attached proposed 6-weeks study plan for both groups, 4. Assess the validity of the achievement vocabulary tests (pre- and post- tests).

Any information or personal details gathered in the course of the study are confidential, except as required by law. No individual will be identified in any publication of the results. Only the researcher will access the data for the purpose of the study. Your opinion is highly valued and will only be used for the purpose of the current research.

My Regards,
The Researcher.

I, ……. have read and understand the information above and my questions and enquiries have been fully answered to my satisfaction. I have had an opportunity to ask questions, and agree to participate in this research study. I have given all the aforementioned documents and I have been given a copy of this consent form to keep.

Juror's/ Expert's Name:

Juror's Signature: Date

APPENDIX I- (B): THE JURY OF THE STUDY DATA TOOLS VALIDATION

No.	Juror's Name	Position	University	Contact Info.
1	Dr. Mohamed AbdelMatal	Assistant Professor, Applied Linguistics	Buraydah Private Colleges	Ma_metaal@yahho.com +96654864576
2	Dr. Abdelmoneim Suliman	Assistant Professor, Linguistics	Qassim University	Abdel_soliman@yahoo.com +966535529204
3	Dr. Fatima Al Naeem	Assistant Professor, Education&Translation	Omdurman Islamic University	Fatima.elnaeem@yahoo.com +966553452678
4	Dr. Waleed Ridwan	Assistant Professor, Research Methodology	Taif University	Waleed200274@yahoo.com +966567991765
5	Dr. Altrimizi Ali	Assistant Professor, Translation	Buraydah Private Colleges	tirmizyali@gmail.com +966551806650
6	Dr. Nesreen Afify	Assistant Professor, Translation	Buraydah Private Colleges	nesrgoda@yahoo.com +966568327170
7	Dr. Menna Al Masry	Assistant Professor, Linguistics	Tanta University, Egypt	mennaelmasry@yahoo.com +201275750676

(C): THE TOTAL RATINGS OF THE JURY FOR TEACHERS' QUESTIONNAIRE

Questionnaire's Items No.	The Items Statement	Valid	Not sure	Invalid
Item No. 5	**5.** Have you ever heard or read or learnt about the term CORPUS LINGUISTICS or CORPUS-BASED or ONLINE CORPORA or any other Corpus-linguistics terms?	5	2	0
Item No. 6	6.How often have you ever been trained during your study at your university on using corpus linguistics approach in teaching or research purposes?	6	1	0
Item No. 7	7. How often have you ever attended a workshop or a training program on using online corpora in classrooms?	5	2	0
Item No. 8	8. How often have your used corpus-based approach in your classroom or for research purposes?	4	3	0
Item No. 9	9.Are your school or university interested in having accounts on online corpora or purchasing concordance software like Wordsmith or Antconc?	4	2	1
Item No. 10	10. Do you think it is beneficial to integrate corpus-based activities in EFL classrooms?	7	0	0
Item No. 11	11. Do you think that Corpus-Linguistics has to be integrated in teacher training programs in your country?	7	0	0
Item No. 12	12. Do you think that lack of technological equipment in your school/ university prevents you from using online corpora in your classes?	4	2	1
Item No. 13	13. Do you believe that students' low technological skills may hinder integrating corpus-based activities in the classroom?	3	3	0
Overall Rating		**7**	**0**	**0**

SUGGETIONS AND RECOMMENDATIONS:

-
-
-
-
-
-
-
-
-
-

Juror's/Expert's Signature **Date:**

(D): THE TOTAL RATINGS OF THE JURY FOR LEARNERS' QUESTIONNAIRE

Questionnaire's Items No.	The Items Statement	Valid	Not sure	Invalid
Item No. 5	**5.** I feel that the use of corpora (COCA Corpus) was helpful for me in learning vocabulary in the course ENG.140?	6	1	0
Item No. 6	**6.** I feel that using corpus-based activities /COCA based activities and printouts in learning vocabulary is more helpful and beneficial than using conventional methods like dictionaries and coursebooks reading texts?	5	2	0
Item No. 7	**7.** I feel that using Corpora and being exposed to massive authentic input contributes positively in learning many aspects and linguistic elements of a word?	7	0	0
Item No. 8	**8.** I feel that being involved in corpora individual activities and group tasks increases my vocabulary retention?	4	2	1
Item No. 9	**9.** I generally feel comfortable to use concordances lines to learn new words and to be involved in the experimental group for a 5-weeks period?	5	1	1
Item No. 10	**10.** After going through this study, I realized the importance of context and co-occurrences provided by the corpora queries to effectively learn new words, rather than using dictionaries.	6	0	1
Item No. 11	**11.** I think that I will use Corpora or COCA Corpus at my own in the future to learn new words or look up words?	7	0	0
Item No. 12	**12.** I feel that the queries concordance lines or context examples are overwhelming and may cause frustration or distraction especially.	5	2	0
Item No. 13	**13.** I feel that by the course of time I got familiarized to easily deal with some challenges of using Corpus-based Approach, like handling overwhelming searches data and do searches for most linguistics features easily.	5	0	1
Item No. 14	**14.** I feel that my language proficiency level impacts my ability to effectively use corpus-based printouts and activities.	7	0	0
Item No. 15	**15.** I have some difficulty in using the COCA corpus due to cut-off sentences in some concordance outputs.	6	0	1
Item No. 17	**16.** I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance outputs.	6	0	1
Overall Rating		**7**	**0**	**0**

SUGGESTIONS AND RECOMMENDATIONS:

-
-
-

Juror's/Expert's Signature **Date**

(E): THE TOTAL RATINGS OF INSTRUCTIONAL MATERIALS, PLANS, AND TESTS

Study Tools and Criteria of Assessment	Minimum	Maximum	Mean	St. Deviation
1. Teaching Plans: Criteria	**-0.5**	**1**	**0.6428**	**0.5149**
E. Fit the Time Framework	-0.5	1	0.5714	0.612
F. Appropriate to the Levels	0	1	0.5	0.651
G. Equal Content for both Groups	0	1	0.5714	0.606
H. Measurable Objectives	0.5	1	0.9285	0.188
2. Assessment Plans: Criteria	**-0.5**	**1**	**0.7321**	**0.2672**
E. Convenient in Respect to the Time Table	-0.5	1	0.5	0.4082
F. Appropriateness of Assessment Tools	1	0.5	0.7142	0.2672
G. Representativeness for The Proposed Objectives	1	1	1	0
H. Equality between the Two Groups	0	1	0.7142	0.3933
3. Activities and Tasks: Criteria	**-0.5**	**1**	**0.75**	**0.2868**
E. Achieving the planned Objectives	0.5	1	0.8571	0.2439
F. Fit the Adopted Approach for Each Group	0.5	1	0.7857	0.2672
G. The Variety of Types and Forms	0.5	1	0.9285	0.1889
H. Promoting Students' Participation and Passion	-0.5	1	0.4285	0.447
4. Pre-tests and Post-tests: Criteria	**-0.5**	**1**	**0.7714**	**0.3296**
F. Diversity of Used Questions	1	1	0.7857	0.3933
G. Representativeness of the Target Vocabulary	-0.5	1	0.4285	0.6726
H. Suitable Length and Time for the Test	-0.5	1	0.7142	0.3933
I. Meeting Different Achievement Levels	0.5	1	0.9285	0.1889
J. Convenient Distribution of Marks.	1	1	1	0

N.B.: Strongly Agree = 1 Neutral = 0 Disagree = -0.5 Agree = 0.5 Strongly Disagree =

APPENDIX J: COCA CORPUS HANDOUT/ GUIDE

The COCA corpus (new version released March 2020)

The corpora from English-Corpora.org are the world's **most widely-used corpora**. The Corpus of Contemporary American English (COCA) is by far the most widely-used of these corpora. In early 2020, we dramatically expanded the scope and size and features of COCA to make it even more useful for researchers, teachers, and learners.

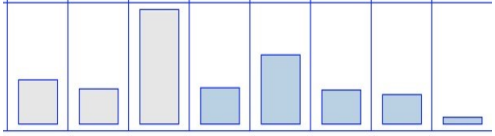
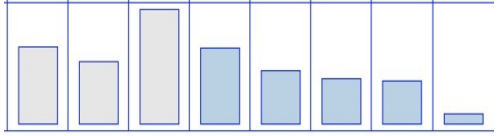
The corpus contains more than **one billion words** of data, including 20 million words each year from **1990-2019** (with the same genre balance year by year). This makes COCA the only corpus of English that is 1) large 2) recent and 3) has a wide range of genres. The following table shows the genres in the corpus.

Genre	# texts	# words	Explanation
Spoken	44,803	127,396,932	Transcripts of unscripted conversation from more than 150 different TV and radio programs (examples: All Things Considered (NPR), Newshour (PBS), Good Morning America (ABC), Oprah)
Fiction	25,992	119,505,305	Short stories and plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books 1990-present, and fan fiction.
Magazines	86,292	127,352,030	Nearly 100 different magazines, with a good mix between specific domains like news, health, home and gardening, women, financial, religion, sports, etc.
Newspapers	90,243	122,958,016	Newspapers from across the US, including: USA Today, New York Times, Atlanta Journal Constitution, San Francisco Chronicle, etc. Good mix between different sections of the newspaper, such as local news, opinion, sports, financial, etc.
Academic	26,137	120,988,361	More than 200 different peer-reviewed journals. These cover the full range of academic disciplines, with a good balance among education, social sciences, history, humanities, law, medicine, philosophy/religion, science/technology, and business
Web (Genl)	88,989	129,899,427	Classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages (by Serge Sharoff). Taken from the US portion of the GloWbE corpus.
Web (Blog)	98,748	125,496,216	Texts that were classified by Google as being blogs. Further classified into the web genres of academic, argument, fiction, info, instruction, legal, news, personal, promotion, review web pages. Taken from the US portion of the GloWbE corpus.
TV/Movies	23,975	129,293,467	Subtitles from OpenSubtitles.org, and later the TV and Movies corpora. Studies have shown that the language from these shows and movies is even more colloquial / core than the data in actual "spoken corpora".
	485,179	**1,002,889,754**	

GENRES Because COCA has so much data from each of these eight genres, it provides useful information about the frequency of words, phrases, and grammatical constructions across the genres – whether they are very informal (e.g. TV and movie subtitles or in spoken transcripts), more formal (e.g. academic articles), or somewhere in between (e.g. magazines and newspapers). For example, the following two charts show the frequency of the word *lucky* in the different genres, as well as the "*get* passive" (e.g. *he got promoted*).

lucky

get passive

SECTION	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	6370	4877	16697	5094	9249	4825	3916	741	34353	26869	50926	33009	21756	19638	17673	3867
WORDS (M)	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	49.53	39.25	130.37	40.39	78.17	38.27	32.17	6.19	267.10	216.24	397.63	261.69	183.87	155.74	145.17	32.28
SEE ALL SUB-SECTIONS AT ONCE																

The great general balance (and large size) of COCA also allows you to see the frequency of related phrases across genres. For example, the following table shows "*soft* NOUN".

		CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC
1		SOFT TISSUE	1120	62	67	74	39	36	100	35	707
2		SOFT DRINKS	1109	123	123	42	90	83	304	296	48
3		SOFT MONEY	790	21	34	12	446	8	78	153	38
4		SOFT SPOT	867	133	110	166	63	159	135	86	15
5		SOFT DRINK	721	43	60	48	77	68	191	199	35
6		SOFT VOICE	546	12	39	10	11	351	53	52	18
7		SOFT POWER	421	49	53	1	64	1	51	32	170
8		SOFT TISSUES	328	13	23	5	6	10	41	6	224
9		SOFT LANDING	274	42	19	18	38	17	56	63	21
10		SOFT LIGHT	247	22	25	9	1	121	38	21	10
11		SOFT PEAKS	216	2	4		1	2	144	62	1
12		SOFT SKIN	207	9	24	28	6	114	19	4	3
13		SOFT TOUCH	213	22	26	28	15	43	51	25	3
14		SOFT SKILLS	219	41	41		7		13	13	104

You can also compare two genres (or sets of genres). For example, the following table shows collocates (nearby words) of *chair* in ACADEMIC (left) and FICTION (right).

SEC 1 (ACADEMIC): 119,790,456 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	DIVISION	56	1	0.5	0.0	55.3
2	ENGINEERING	53	1	0.4	0.0	52.4
3	ABA	55	0	0.5	0.0	45.9
4	PROGRAM	46	1	0.4	0.0	45.4
5	EDUCATION	42	0	0.4	0.0	35.1
6	RESEARCH	34	1	0.3	0.0	33.6
7	FACULTY	38	0	0.3	0.0	31.7
8	ASME	36	0	0.3	0.0	30.1
9	COMMISSION	30	1	0.3	0.0	29.6
10	ROLE	33	0	0.3	0.0	27.5
11	SUBCOMMITTEE	24	1	0.2	0.0	23.7
12	DESIGN	28	0	0.2	0.0	23.4
13	MEMBER	45	2	0.4	0.0	22.2
14	RISE	44	2	0.4	0.0	21.7
15	DEAN	26	0	0.2	0.0	21.7

SEC 2 (FICTION): 118,322,084 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	KITCHEN	293	3	2.5	0.0	98.9
2	WING	111	0	0.9	0.0	93.8
3	WAY	87	1	0.7	0.0	88.1
4	FEET	220	3	1.9	0.0	74.2
5	FOOT	72	1	0.6	0.0	72.9
6	MOTHER	134	2	1.1	0.0	67.8
7	COMMAND	65	1	0.5	0.0	65.8
8	LEATHER	409	7	3.5	0.1	59.2
9	CANVAS	58	1	0.5	0.0	58.7
10	BEACH	67	0	0.6	0.0	56.6
11	FINGERS	53	1	0.4	0.0	53.7
12	LAWN	318	6	2.7	0.1	53.7
13	VISITOR	51	1	0.4	0.0	51.6
14	CAPTAIN	61	0	0.5	0.0	51.6
15	DAY	50	1	0.4	0.0	50.6

You can even compare synonyms of a given word in different genres. For example, this table shows the synonyms of *strong* in TV/MOVIES (left) and ACADEMIC (right).

The ability to focus in on specific genres means that you can find "just the right word" for a

SEC 1 (TV/MOVIES): 128,074,534 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	SPICY	581	53	4.5	0.4	10.3
2	BEEFY	64	7	0.5	0.1	8.6
3	TOUGH	11190	1579	87.4	13.2	6.6
4	STRAPPING	143	21	1.1	0.2	6.4
5	HOT	23639	3877	184.6	32.4	5.7
6	BRILLIANT	4418	1156	34.5	9.7	3.6
7	GREAT	107836	34522	842.0	288.2	2.9
8	BITING	658	230	5.1	1.9	2.7
9	BURLY	52	27	0.4	0.2	1.8
10	BRIGHT	4561	2542	35.6	21.2	1.7
11	SOUND	18257	11719	142.5	97.8	1.5
12	WELL-BUILT	20	14	0.2	0.1	1.3

SEC 2 (ACADEMIC): 119,790,456 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	ROBUST	2829	108	23.6	0.8	28.0
2	EFFECTIVE	28807	1596	240.5	12.5	19.3
3	DURABLE	683	65	5.7	0.5	11.2
4	CONCENTRATED	2750	298	23.0	2.3	9.9
5	STAUNCH	252	32	2.1	0.2	8.4
6	ZEALOUS	217	31	1.8	0.2	7.5
7	ARDENT	375	54	3.1	0.4	7.4
8	COMPELLING	2845	452	23.7	3.5	6.7
9	FERVENT	210	34	1.8	0.3	6.6
10	CLEAR-CUT	405	83	3.4	0.6	5.2
11	PERSUASIVE	1360	308	11.4	2.4	4.7
12	DEEP-SEATED	260	61	2.2	0.5	4.6

WORD 1 (W1): DEEP (7.33)

	WORD	W1	W2	W1/W2	SCORE
1	BREATH	7687	0	15,374.0	2,096.7
2	SPACE	1292	0	2,584.0	352.4
3	BREATHS	1285	0	2,570.0	350.5
4	POCKETS	764	0	1,528.0	208.4
5	WATER	1224	1	1,224.0	166.9
6	THROAT	513	0	1,026.0	139.9

WORD 2 (W2): PROFOUND (0.14)

	WORD	W2	W1	W2/W1	SCORE
1	HEARING	54	0	108.0	791.9
2	DISABILITIES	25	0	50.0	366.6
3	AFFECT	23	0	46.0	337.3
4	EFFECT	695	16	43.4	318.5
5	EFFECTS	214	7	30.6	224.2
6	DISABILITY	14	0	28.0	205.3

particular concept in a particular genre. For example, suppose that you want to know what word related to *potent* is the most frequent with a form of *argument* in academic English. As the following image shows, you simply search for *=potent ARGUMENT* and limit the search to academic, and you would see the following results, and then (as with any search results) can see the matching phrases in context,

In addition to comparing across genres, you can also compare words, to "tease out" differences between related words. For example, the following chart shows words that occur with *deep* and *profound*, showing that (for example) *deep breath* is common but *profound breath* never occurs, or that *profound effect* is common but *deep effec*t is not. You can also compare other related words, such as adjectives used with the words *men* and *women*, or verbs occurring near *Obama* or *Trump*.

HISTORICAL COCA is the only large corpus of English that has extensive data from the entire period of the last 30 years – 20 million words per year from 1990-2019 (with the same genre balance year by year). This means that in addition to seeing variation by genre, you can also map out recent changes in English in ways that are not possible with any other corpus – such as with the frequency of *awesome* from 1990-2019,.

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	24063	9902	6800	8494	2740	957	1819	1352	350	940	1127	1688	3075	4523	4359
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	24.23	76.99	54.73	66.32	21.72	8.09	14.43	11.11	2.92	6.76	7.63	11.52	21.21	31.14	30.12
SEE ALL SUB-SECTIONS AT ONCE															

And of course you can look at much more than just simple words or phrases. COCA is the only corpus that allows you to map out changes in syntactic constructions over the past 30 years, as with the "like construction" (*and I'm like, no way*) or the "end up V-ing" construction (*you'll end up paying way too much*) – both of which have increased in each five year period since the early 1990s.

SECTION	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	140	393	639	1145	1780	2581	1826	2340	2489	2849	2949	3293
WORDS (M)	139.1	147.8	146.6	144.9	145.3	144.7	139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	1.01	2.66	4.36	7.90	12.25	17.83	13.13	15.83	16.98	19.66	20.30	22.75
SEE ALL SUB-SECTIONS AT ONCE												

Just as we compared collocates (nearby words) in different genres above (for *chair* in ACAD and FIC), we can also compare collocates in different periods, to look at semantic change. For example, the following table shows collocates of *green* in the 1990s (left) and the 2010s (right).

We can also see what is being said about a topic over time. For example, the following shows the collocates of crisis in each five year period (and genre) from 1990-2019, which shows what we were worrying about in these different periods.

SEC 1 (1995-1999, 1990-1994): 286,833,557 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	GREEN RANGER	187	0	0.7	0.0	65.2
2	GREEN PEPPER	215	35	0.7	0.1	6.2
3	GREEN CROSS	52	9	0.2	0.0	5.8
4	GREEN VEGETABLES	73	32	0.3	0.1	2.3
5	GREEN PEPPERS	94	47	0.3	0.2	2.0
6	GREEN MAN	58	31	0.2	0.1	1.9
7	GREEN ACRES	52	31	0.2	0.1	1.7
8	GREEN BELL	154	92	0.5	0.3	1.7
9	GREEN PLANTS	61	37	0.2	0.1	1.7
10	GREEN GLASS	61	37	0.2	0.1	1.7

SEC 2 (2010-2014, 2015-2019): 290,003,115 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	GREEN GAZETTE	96	0	0.3	0.0	33.1
2	GREEN JOBS	87	0	0.3	0.0	30.0
3	GREEN PRACTICE	60	2	0.2	0.0	29.7
4	GREEN ENERGY	170	7	0.6	0.0	24.0
5	GREEN ARROW	192	8	0.7	0.0	23.7
6	GREEN BUILDING	130	18	0.4	0.1	7.1
7	GREEN SCREEN	85	12	0.3	0.0	7.0
8	GREEN ZONE	118	21	0.4	0.1	5.6
9	GREEN LANTERN	96	21	0.3	0.1	4.5
10	GREEN SPACES	97	23	0.3	0.1	4.2

CONTEXT	ALL	BLOG	WEB-GENL	TV/MOVIES	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009	2010-2014	2015-2019
DEBT	1513	370	419	4	264	2	76	203	175	115	63	25	29	420	72
CRISIS	1088	142	146	64	268	24	140	128	176	194	128	76	116	162	124
HEALTH	917	65	92	36	209	7	214	177	117	149	84	84	133	115	195
ENERGY	771	87	96	32	129	10	135	196	86	94	43	245	123	48	35
GULF	656		5	3	267	1	122	194	64	629	11	3	2	4	2
OIL	437	58	38	8	76	8	89	69	91	132	46	64	53	28	18
REFUGEE	381	10	16	6	122	6	57	65	99	47	48	19	15	27	199
SOLUTION	381	49	49	4	97	7	42	67	66	116	31	33	36	40	27
MIDLIFE	348	32	33	115	62	24	40	21	21	25	31	45	26	104	52
CREDIT	385	86	82	2	46	1	50	94	24	3	9	2	138	50	15
CLIMATE	396	114	80	2	65	4	86	14	31	2	1	12	48	33	106

As with all of the other corpora from English-Corpora.org, there is a very **wide range of searches**,

		CONTEXT ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 1,292 | UNIQUE 900 +
1	☐	WEARING A TOP HAT	22	
2	☐	WEARING A SHORT SKIRT	18	
3	☐	WEARING A HARD HAT	14	
4	☐	WEARING A WHITE BLOUSE	14	
5	☐	WEAR A SHORT SKIRT	13	
6	☐	WEARING A WHITE HAT	13	
7	☐	WORE A SHORT SKIRT	13	
8	☐	WEARING A RED TIE	12	

1	2019	FIC	PennLitJournal	A	B	C	know Jessica, the salesgirl with the husky voice who as usual wore cuffed hot pants. I discovered much to my surprise that she lived in southeast, to
2	2019	NEWS	Chicago Tribune	A	B	C	the 2400 block of Partridge Lane told police Feb. 19 that someone wearing a dark hat and jacket was looking into a neighbor's vehicle. # A resident i
3	2019	TV	The Blacklist	A	B	C	to the defendant? It was the only suit in the area wearing a fedora-style hat. I see. And did you arrest the defendant at that point? No
4	2018	NEWS	OregonLive.com	A	B	C	Security disability benefits, punched the guard after he was accused of wearing a stolen hat on May 12, 2017. # Nathaniel M. Williams, 25Columbia
5	2017	FIC	Bk:One-hundred-knuckledFist	A	B	C	because he looks toward the sound, too. # A woman wearing a bright-yellow blouse and black pants hums past his stand, armed with a small vacuu
6	2017	MAG	Fox News	A	B	C	evacuated Tuesday after an explosion was reported and police shot a suspect wearing an explosive belt who shouted " Allahu Akbar, " in what offici
7	2017	NEWS	USA TODAY	A	B	C	in Jerusalem's Old City on May 22, 2017. She wore a white skirt suit by Michael Kors and her customary sky-high heels in a brightly colored striped p
8	2016	FIC	Bk:LoveMaps	A	B	C	my apple! " Maya yelled from the porch. # She wore a white blouse tied at the waist and a blue-and-black plaid skirt. She skipped toward them,
9	2016	FIC	Dialogue: Mormon	A	B	C	. Approaching her, he saw it was Jennie O'Brien. She wore a drab skirt, and her hair was wound into an untidy bun upon her head. #

including: words, phrases, substrings, lemmas, part of speech, synonyms, and customized wordlists. For example, the search **WEAR * ADJ @CLOTHES** takes just about one second to search through the billion words to find strings like the following (and it doesn't require learning

		USE LARGE NGRAMS [?]	FREQ	TOTAL 8,108,778 | UNIQUE 17,446 +
1	☐	HEALTH CARE	64877	
2	☐	WORLD WAR	32709	
3	☐	LIVING ROOM	23171	
4	☐	LAW ENFORCEMENT	23057	
5	☐	WEB SITE	22382	
6	☐	CLIMATE CHANGE	21967	
7	☐	AIR FORCE	19768	
8	☐	REPUBLICAN PARTY	16928	
9	☐	STATE UNIVERSITY	16926	
10	☐	HEALTH INSURANCE	16378	
11	☐	FAMILY MEMBERS	16031	

		USE LARGE NGRAMS [?]	FREQ	TOTAL 178,237 | UNIQUE 6,395 +
1	☐	TAKE GOOD CARE	1226	
2	☐	MAKES PERFECT SENSE	1120	
3	☐	PAY CLOSE ATTENTION	706	
4	☐	FIND COMMON GROUND	683	
5	☐	BEGIN AUDIO CLIP	651	
6	☐	SAY GOOD NIGHT	641	
7	☐	TAKE FULL ADVANTAGE	558	
8	☐	GRATED PARMESAN CHEESE	553	
9	☐	COME FULL CIRCLE	536	
10	☐	COME RIGHT BACK	491	
11	☐	WORK FULL TIME	445	
12	☐	TRY NEW THINGS	439	

unnecessarily convoluted search syntax).

Because of COCA's advanced architecture, even searches for very general searches like **NOUN + NOUN** or **VERB ADJ NOUN** take just 1-2 seconds to search through the billion word corpus:

A unique feature of COCA, which makes it very useful for language learners and teachers, is the ability to **browse** through a list of the **top 60,000 words** (lemmas) in the corpus, and then to see an extremely wide range of information on each of these words.

For example, the following are just a few examples of high frequency words (about word #5000 in the 60,000 word list), medium frequency (~25,000), and low frequency (~45,000) words. For each word in the list, users can hear the word **pronounced**, see **videos** with that word in the text, find related **images** from Google Images, an see a **translation** for their preferred target language.

	RANK	FREQ	Word	PoS	Audio	Video	Image	ZH-CN
1	5200	11377	blogger	NOUN				
2	5201	11374	utterly	ADV				
3	5202	11372	trouble	VERB				
4	5203	11368	texture	NOUN				
5	5204	11365	head	ADJ				

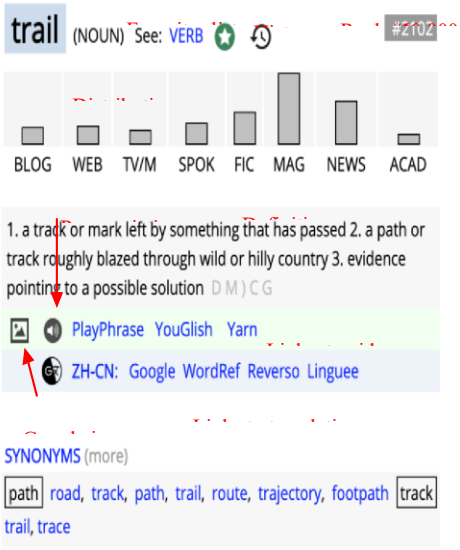
	RANK	FREQ	Word	PoS	Audio	Video	Image	ZH-CN
1	25201	577	unappealing	ADJ				
2	25202	577	tryst	NOUN				
3	25203	577	urologist	NOUN				
4	25204	577	scrabble	NOUN				
5	25205	577	demarcate	VERB				

	RANK	FREQ	Word	PoS	Audio	Video	Image	ZH-CN
1	45107	112	incubus	NOUN				
2	45108	112	unquantifiable	ADJ				
3	45110	112	husk	VERB				
4	45111	112	hunky-dory	ADJ				
5	45112	112	jacked	ADJ				

It is even possible to search the 60,000 word list by **pronunciation** (very helpful, because of difficult English spelling). For example, the following is a partial list of two syllable words (accented on the second syllable) that rhyme with *stay*:

Pronunciation	Rhymes with stay Type EXACT
Syllables / stress	○●○○○○○○○ ✖

11	3958	17233	array	NOUN				
12	4122	16221	convey	VERB				
13	4589	13766	survey	VERB				
14	4751	13055	cafe	NOUN				
15	4779	12934	betray	VERB				
16	5536	10325	obey	VERB				
17	5933	9149	ballet	NOUN				



TOPICS (more)

hike, hike, mountain, steep, park, terrain, mile, forest, creek, lake, wilderness, route, peak, river, trail, visitor, valley, hiker, snow, climb

COLLOCATES (more)

NOUN	campaign, mile, paper, blazer, park, mountain, blood, road
VERB	follow, hike, lead, blaze, ride, climb, cross, wind
ADJ	appalachian, steep, narrow, historic, scenic, rocky, indian, rough
ADV	left, downhill, steeply, uphill, aptly, underfoot, deftly, briskly

CLUSTERS (more)

trail •	trail in • trail to • trail blazers • trail for • trail with • trails in • trail from • trail on
• trail	campaign trail • paper trail • appalachian trail • hiking trails • long trail • blood trail • on trails • money trail
trail • •	trail of tears • trail of blood • trail of destruction • trail that led • trail leads to • trail of smoke • trail of bread • trail of broken
• • trail	on the trail • along the trail • up the trail • off the trail • down the trail • leaving a trail • left a trail • to the trail
trail • • •	trail of bread crumbs • trail that led to • trail to the top • trail through the woods • trails in the sky • trail of destruction across • trail that leads to • trail of broken hearts
• • • trail	on the campaign trail • out on the trail • from the campaign trail • end of the trail • hot on the trail • side of the trail • on the appalachian trail • miles of hiking trails

TEXTS / VIRTUAL CORPORA (more)

WEB:nps.gov • WEB:ponderosaca.com • NEWS:AssocPress • MAG:Backpacker • ACAD:InstrPsych • MAG:Bicycling • NEWS:USA TODAY • MAG:Backpacker • FIC:FantasySciFi • MAG:Backpacker • MAG:NatlParks • NEWS:USAToday • MAG:Backpacker • MAG:Backpacker • WEB:...documentaryfilms.com • MAG:NatlParks • MAG:Bicycling • MAG:Backpacker • MAG:Bicycling • MAG:Backpacker • NEWS:Atlanta • MAG:Backpacker • MAG:Backpacker • MAG:NatlParks • NEWS:Atlanta •

CONCORDANCE LINES (more)

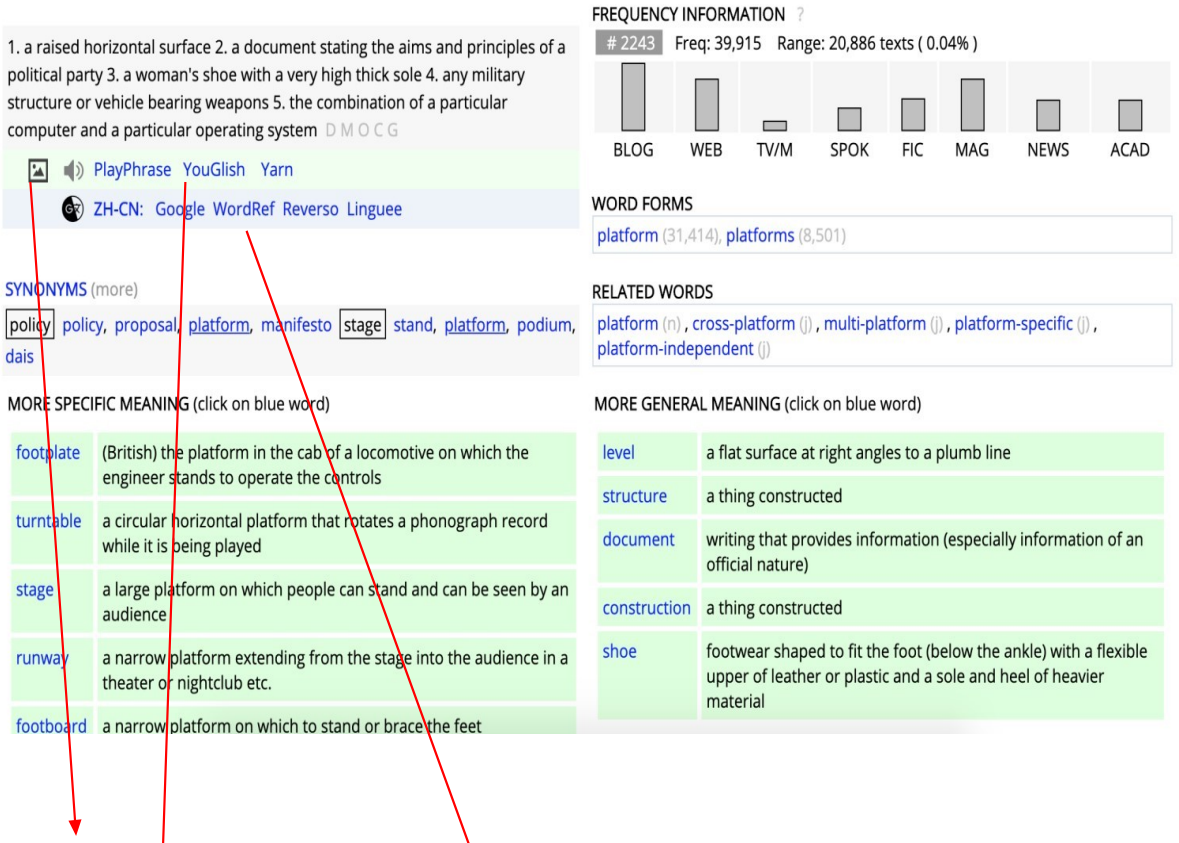
1	WEB: 2012: amazon.com	Along the way , they confront Army Rangers along the Appalachian	Trail	and conduct a late-night stealth mission to an abandoned NASA
2	BLOG: 2012: ...heother.blogspot....	me , a direct definition of ultra trail running . The	trails	are amazing . DCNR does an excellent job keeping the trails and
3	MAG: 1991: NatlParks	to protect a continuous corridor are ongoing . The North Country	Trail	Association , a 500-member organization established in 1980 ,
4	MAG: 1991: NatlParks	. # For more information , contact the Overmountain Victory	Trail	Association , c/o P.O. Box 632 , Manassas Park , VA 22111
5	MAG: 2003: Bicycling	and anti-bike website operator John Parulis , describing the	trails	at China Camp , east of San Rafael " Look at this
6	WEB: 2012: yelp.com	try ... # Perks of coming here is hiking the various	trails	at the top , and taking the ski lift back down .
7	BLOG: 2012: seattlemet.com	that the city did not " re-interview " businesses along the	trail	before approving its completion ; and the fact that an
8	TV: 2016: Tomorrow, When the W...	OK ? Rain gear 's there too . Should be some	trail	bikes out the back if they have n't been taken . Awesome
9	NEWS: 1998: Houston	because only a few of the wealthiest teams - the Portland	Trail	Blazers (owned by billionaire Paul Allen) , the New York
10	MAG: 2010: PopMech	, I also fail to notice that the side of the	trail	butts a 50-foot cliff off the creek bed . With my blissful
11	MAG: 2009: OutdoorLife	to find one . He showed me some images from a	trail	camera that was placed on the stand farthest from camp , the
12	TV: 2014: Avengers Assemble	, arizona . How did ... Technopathic static . Covering one	trail	can sometimes make another . I told you I 'd back your
13	MAG: 2007: AmSpect	first four miles or so were quite steep , as the	trail	climbed steadily up the drainage through thick timber above the
14	MAG: 1997: Backpacker	and its remote eastern tributaries , the vast and contorted Burr	Trail	country , the huge and austere Kaiparowits Plateau , and the slot
15	MAG: 2009: Backpacker	, so we divert onto a slippery embankment that rejoins the	trail	downriver . Later , we recognize what 's left of an elk
16	MAG: 1997: Backpacker	and describes himself as a warm sleeper . Often on the	trail	for 10 or more days at a time , Mike values a
17	MAG: 2008: Bicycling	paths . Barnack can see riders heading up the Tollgate	Trail	from his kitchen window . But the Tims also ride . Barnack
18	NEWS: 2018: OregonLive.com	TriMet and PGE are working a section of the Springwater Corridor	Trail	from Southeast Tacoma to 45th Place and the widely used trail is
19	NEWS: 1992: WashPost	tour , the bus drops hikers off at the Skyline Drive	trail	head and picks them up at the bottom . A historian and
20	MAG: 1997: Backpacker	a free backcountry trip planner and general map , or call	Trails	Illustrated , (800) 962-1643 for its map to the park

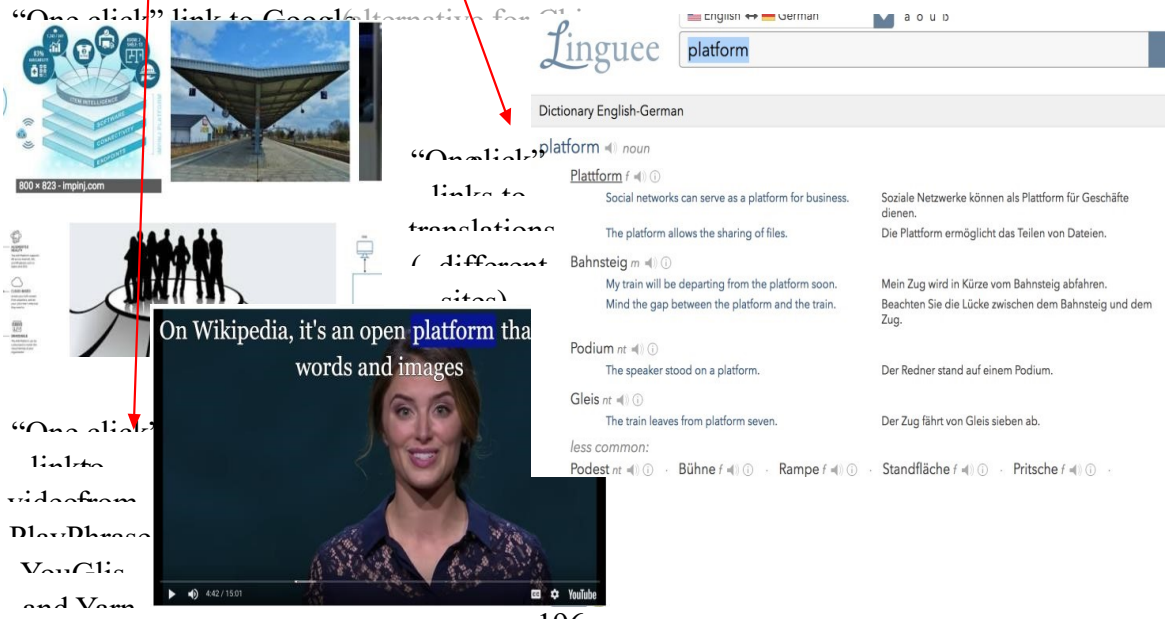
Each of the top 60,000 words in the corpus has a home page such as the following, with links to

other pages with more information. Users can save words to a “favorites” list for later review, and go back through a history of all of their “word” pages

Each of the top 60,000 words also has more detailed pages, including a “dictionary” page, related topics, collocates, clusters, websites, and concordance lines. Samples of each of these pages are given below.**DICTIONARY** page

Includes a **definition**, links to Google **Images**, **pronunciation**, **videos**, and **translations** (to desired





language) at (up to) four different sites. Also includes **synonyms** and words with more specific

TOPICS HORMONE NOUN

	Same text	Word	PoS
1	253	estrogen	n
2	245	supplement	n
3	227	stimulate	v
4	224	therapy	n
5	209	cancer	n
6	203	disease	n
7	202	prescribe	v
8	199	symptom	n
9	192	diet	n
10	187	breast	n
11	174	stress	n
12	171	nutrient	n
13	168	obesity	n
14	161	protein	n
15	161	researcher	n
16	160	induce	v
17	160	injection	n
18	160	tissue	n
19	158	medicine	n

TOPICS COSMOLOGICAL ADJ

	Same text	Word	PoS
1	108	universe	n
2	82	bang	n
3	73	physics	n
4	71	cosmic	j
5	70	cosmology	n
6	69	galaxy	n
7	68	quantum	n
8	62	physicist	n
9	59	particle	n
10	57	astronomer	n
11	55	gravity	n
12	54	expansion	n
13	53	theory	n
14	50	relativity	n
15	48	atom	n
16	45	constant	j
17	45	correspond	v
18	44	gravitational	j
19	44	mathematical	j

TOPICS VIBRATE VERB

	Same text	Word	PoS
1	40	vibration	n
2	30	frequency	n
3	25	physics	n
4	23	atom	n
5	22	particle	n
6	21	quantum	n
7	19	faint	j
8	19	string	n
9	17	molecule	n
10	17	sound	n
11	17	wave	n
12	16	electron	n
13	16	gravity	n
14	16	magnetic	j
15	16	mathematical	j
16	16	motion	n
17	16	physicist	n
18	16	vacuum	n
19	15	grin	v

meanings (**hyponyms**) and more general meanings (**hypernyms**) (both from WordNet). It also includes **frequency information** (including rank order, number of tokens, and two "range" measures of how well the word is spread throughout the ~500,000 texts, as well as its distribution across the eight main genres in the corpus. Finally, it also includes the frequency of the different **forms** of the lemma (e.g. verb forms), and **related words**.

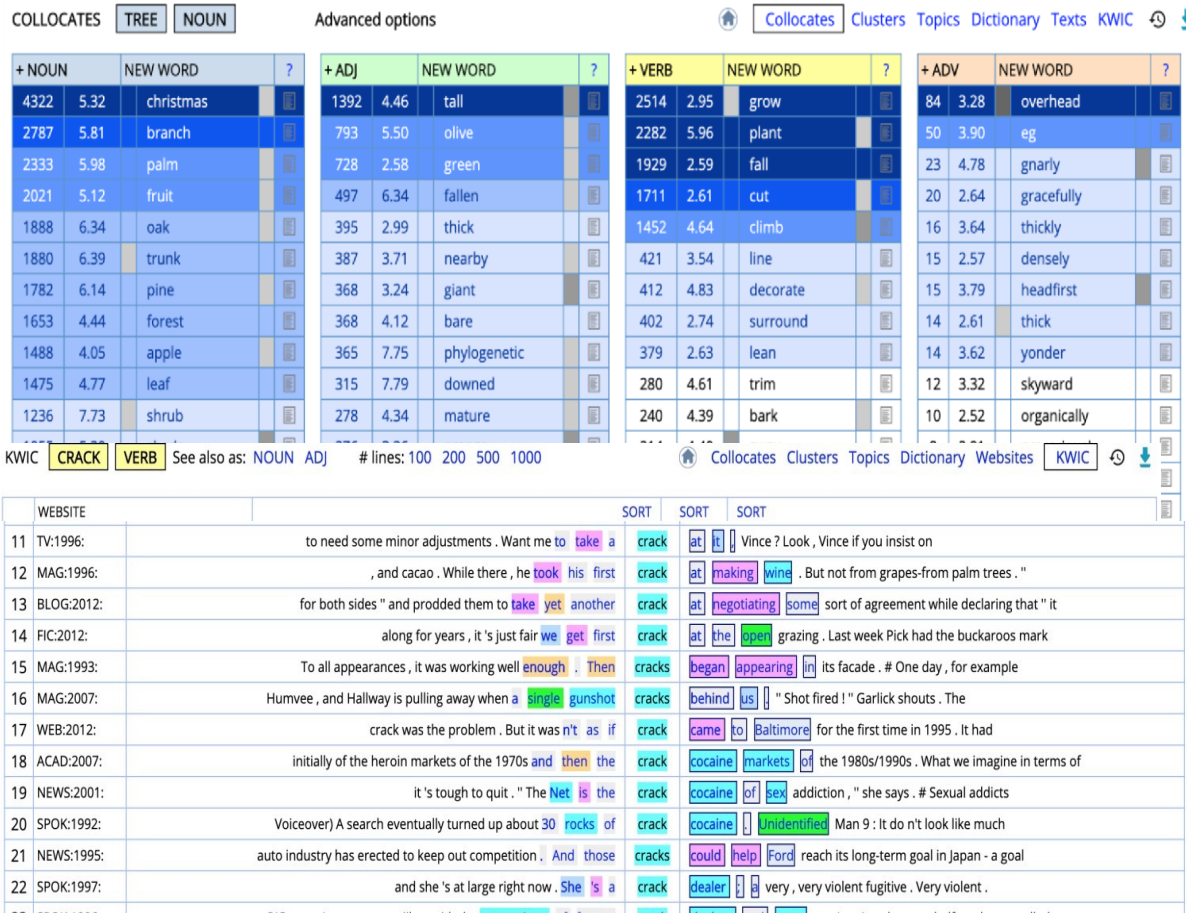
TOPICS page

Shows words that tend to co-occur in the 22 million webpages in the corpus. In many cases these provide better insight into meaning and usage than collocates (the standard tool for finding textually related words), and yet we're not aware of any other corpus that has these. Users can click on any of these words to follow a "chain" of related words.

COLLOCATES Page

Words that occur nearby. Click on any word to see the collocates for the new word, or click on the "text" icon to see the word in context. The gray boxes show the most normal placement of the node word (e.g. *Christmas tree*, but *tree trunk*). You can also sort by Mutual Information score and set frequency thresholds (Advanced Options).

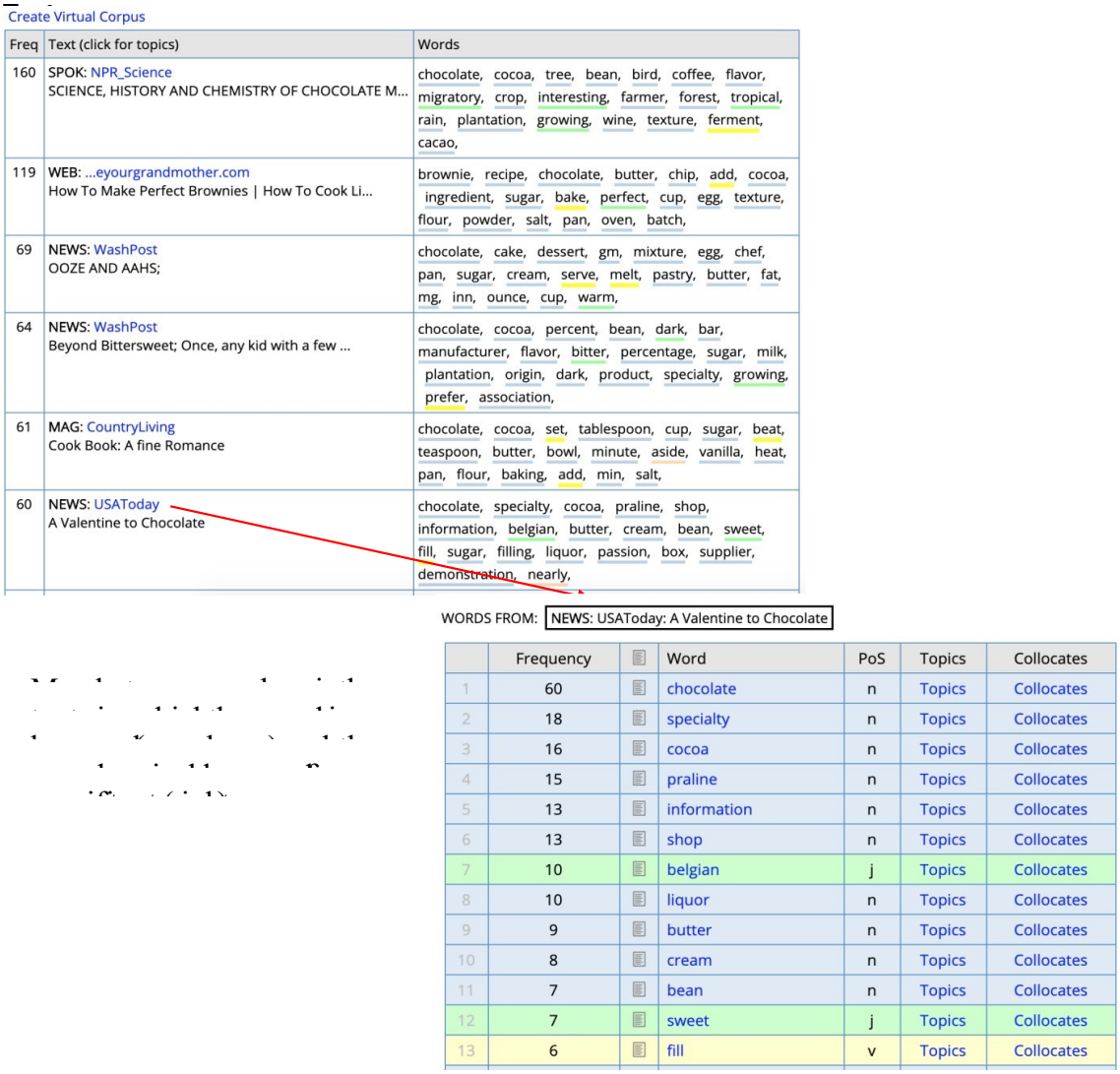
CLUSTERS page



The most common 2, 3, and 4 word strings. You can decide how “wide” or “tight” you want the clusters to be (e.g. whether they include highly frequent words like articles and prepositions).

KWIC / concordance page

See 100 – 1000 random lines, with the surrounding words highlighted for part of speech. You can re-sort the words by words to the left or right, all with the goal of seeing the patterns in which the word
occurs.



Create Virtual Corpus

Freq	Text (click for topics)	Words
160	SPOK: NPR_Science SCIENCE, HISTORY AND CHEMISTRY OF CHOCOLATE M...	chocolate, cocoa, tree, bean, bird, coffee, flavor, migratory, crop, interesting, farmer, forest, tropical, rain, plantation, growing, wine, texture, ferment, cacao,
119	WEB: ...eyourgrandmother.com How To Make Perfect Brownies | How To Cook Li...	brownie, recipe, chocolate, butter, chip, add, cocoa, ingredient, sugar, bake, perfect, cup, egg, texture, flour, powder, salt, pan, oven, batch,
69	NEWS: WashPost OOZE AND AAHS;	chocolate, cake, dessert, gm, mixture, egg, chef, pan, sugar, cream, serve, melt, pastry, butter, fat, mg, inn, ounce, cup, warm,
64	NEWS: WashPost Beyond Bittersweet; Once, any kid with a few ...	chocolate, cocoa, percent, bean, dark, bar, manufacturer, flavor, bitter, percentage, sugar, milk, plantation, origin, dark, product, specialty, growing, prefer, association,
61	MAG: CountryLiving Cook Book: A fine Romance	chocolate, cocoa, set, tablespoon, cup, sugar, beat, teaspoon, butter, bowl, minute, aside, vanilla, heat, pan, flour, baking, add, min, salt,
60	NEWS: USAToday A Valentine to Chocolate	chocolate, specialty, cocoa, praline, shop, information, belgian, butter, cream, bean, sweet, fill, sugar, filling, liquor, passion, box, supplier, demonstration, nearly,

WORDS FROM: NEWS: USAToday: A Valentine to Chocolate

	Frequency		Word	PoS	Topics	Collocates
1	60		chocolate	n	Topics	Collocates
2	18		specialty	n	Topics	Collocates
3	16		cocoa	n	Topics	Collocates
4	15		praline	n	Topics	Collocates
5	13		information	n	Topics	Collocates
6	13		shop	n	Topics	Collocates
7	10		belgian	j	Topics	Collocates
8	10		liquor	n	Topics	Collocates
9	9		butter	n	Topics	Collocates
10	8		cream	n	Topics	Collocates
11	7		bean	n	Topics	Collocates
12	7		sweet	j	Topics	Collocates
13	6		fill	v	Topics	Collocates

Summary

COCA has a number of features that sets it apart from any other corpus. These include its **size** (1.0 billion words), how up-to-date it is (texts through Dec 2019), the wide range of **genres** (TV/Movie subtitles, spoken, blogs, web, fiction, magazine, newspaper, academic), and its **searches** (range of query types, and the ease and speed of its searches), including the ability to limit to and compare across genres and time periods.

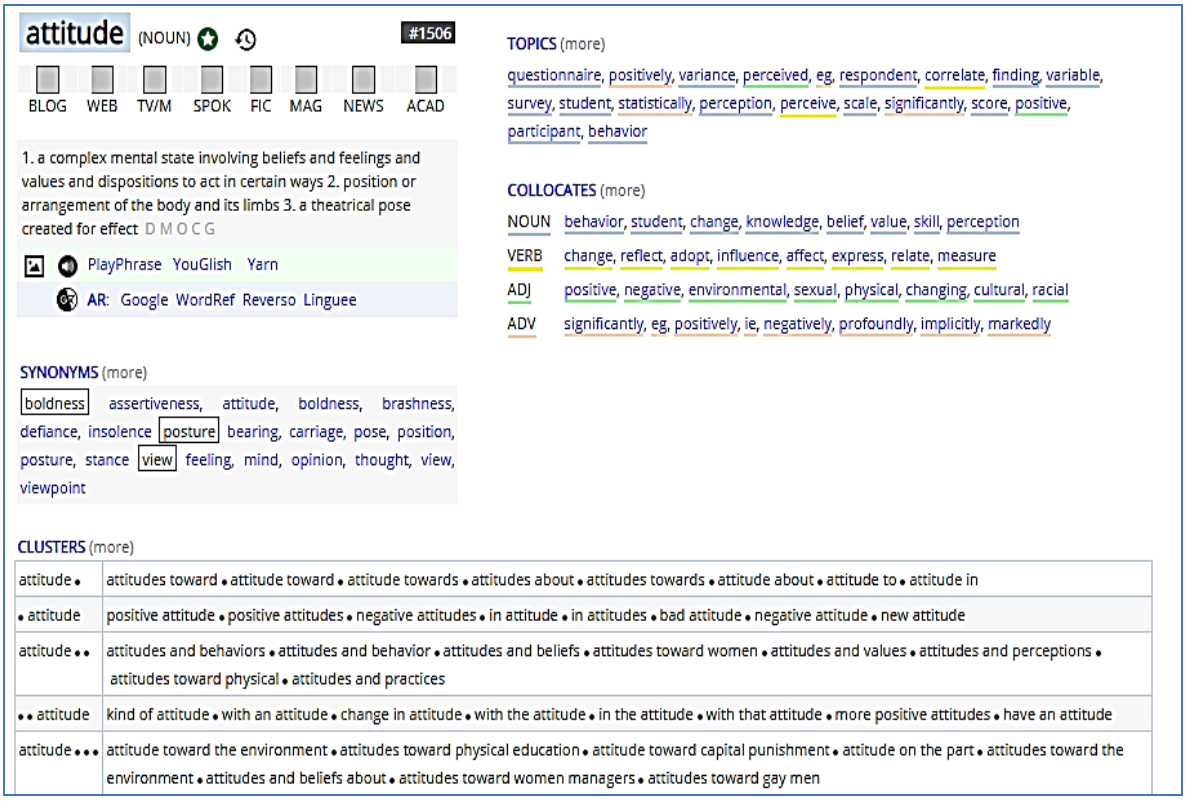
In addition, it is different from most of the other corpora from English-Corpora.org in the attention that it gives to the **top 60,000 words** in the corpus, and the wide range of information for each word, including frequency information, definitions, synonyms, WordNet entries, related topics, concordances (new display in COCA), clusters, websites that have the word as a "keyword", and KWIC/concordance lines.

All of these features make COCA the ideal corpus for researchers, teachers, and learners.

APPENDIX K: PRINTOUTS AND EXERCISES FOR CORPUS-BASED EXPERIMENTAL GROUP

A Sample for Using *Full Description of a Word* Query Type through COCA Corpus for N. (ATTITUDES)

A Shot Screen Taken from COCA Corpus from the Search Data for the Word "ATTITUDE"



attitude (NOUN) #1506

BLOG WEB TV/M SPOK FIC MAG NEWS ACAD

1. a complex mental state involving beliefs and feelings and values and dispositions to act in certain ways 2. position or arrangement of the body and its limbs 3. a theatrical pose created for effect D M O C G

PlayPhrase YouGlish Yarn

AR: Google WordRef Reverso Linguee

TOPICS (more)

questionnaire, positively, variance, perceived, eg, respondent, correlate, finding, variable, survey, student, statistically, perception, perceive, scale, significantly, score, positive, participant, behavior

COLLOCATES (more)

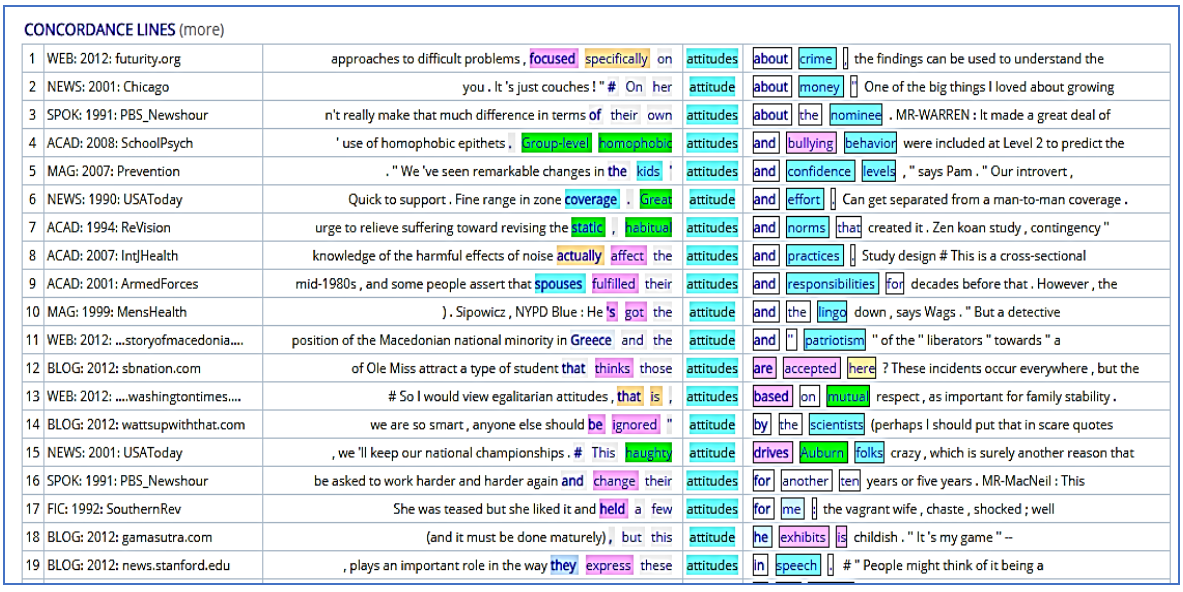
NOUN	behavior, student, change, knowledge, belief, value, skill, perception
VERB	change, reflect, adopt, influence, affect, express, relate, measure
ADJ	positive, negative, environmental, sexual, physical, changing, cultural, racial
ADV	significantly, eg, positively, ie, negatively, profoundly, implicitly, markedly

SYNONYMS (more)

boldness assertiveness, attitude, boldness, brashness, defiance, insolence posture bearing, carriage, pose, position, posture, stance view feeling, mind, opinion, thought, view, viewpoint

CLUSTERS (more)

attitude •	attitudes toward • attitude toward • attitude towards • attitudes about • attitudes towards • attitude about • attitude to • attitude in
• attitude	positive attitude • positive attitudes • negative attitudes • in attitude • in attitudes • bad attitude • negative attitude • new attitude
attitude ••	attitudes and behaviors • attitudes and behavior • attitudes and beliefs • attitudes toward women • attitudes and values • attitudes and perceptions • attitudes toward physical • attitudes and practices
•• attitude	kind of attitude • with an attitude • change in attitude • with the attitude • in the attitude • with that attitude • more positive attitudes • have an attitude
attitude •••	attitude toward the environment • attitudes toward physical education • attitude toward capital punishment • attitude on the part • attitudes toward the environment • attitudes and beliefs about • attitudes toward women managers • attitudes toward gay men



CONCORDANCE LINES (more)

1	WEB: 2012: futurity.org	approaches to difficult problems , focused specifically on	attitudes	about crime ;	the findings can be used to understand the
2	NEWS: 2001: Chicago	you . It 's just couches ! " # On her	attitude	about money "	One of the big things I loved about growing
3	SPOK: 1991: PBS_Newshour	n't really make that much difference in terms of their own	attitudes	about the nominee	. MR-WARREN : It made a great deal of
4	ACAD: 2008: SchoolPsych	' use of homophobic epithets . Group-level homophobic	attitudes	and bullying behavior	were included at Level 2 to predict the
5	MAG: 2007: Prevention	. " We 've seen remarkable changes in the kids '	attitudes	and confidence levels	, " says Pam . " Our introvert ,
6	NEWS: 1990: USAToday	Quick to support . Fine range in zone coverage . Great	attitude	and effort ;	Can get separated from a man-to-man coverage .
7	ACAD: 1994: ReVision	urge to relieve suffering toward revising the static , habitual	attitudes	and norms that	created it . Zen koan study , contingency "
8	ACAD: 2007: IntJHealth	knowledge of the harmful effects of noise actually affect the	attitudes	and practices ;	Study design # This is a cross-sectional
9	ACAD: 2001: ArmedForces	mid-1980s , and some people assert that spouses fulfilled their	attitudes	and responsibilities for	decades before that . However , the
10	MAG: 1999: MensHealth	) . Sipowicz , NYPD Blue : He 's got the	attitude	and the lingo	down , says Wags . " But a detective
11	WEB: 2012: ...storyofmacedonia....	position of the Macedonian national minority in Greece and the	attitude	and " patriotism	" of the " liberators " towards " a
12	BLOG: 2012: sbnation.com	of Ole Miss attract a type of student that thinks those	attitudes	are accepted here	? These incidents occur everywhere , but the
13	WEB: 2012: ....washingtontimes....	# So I would view egalitarian attitudes , that is ,	attitudes	based on mutual	respect , as important for family stability .
14	BLOG: 2012: wattsupwiththat.com	we are so smart , anyone else should be ignored "	attitude	by the scientists	(perhaps I should put that in scare quotes
15	NEWS: 2001: USAToday	, we 'll keep our national championships . # This haughty	attitude	drives Auburn folks	crazy , which is surely another reason that
16	SPOK: 1991: PBS_Newshour	be asked to work harder and harder again and change their	attitudes	for another ten	years or five years . MR-MacNeil : This
17	FIC: 1992: SouthernRev	She was teased but she liked it and held a few	attitudes	for me ;	the vagrant wife , chaste , shocked ; well
18	BLOG: 2012: gamasutra.com	(and it must be done maturely) , but this	attitude	he exhibits is	childish . " It 's my game " --
19	BLOG: 2012: news.stanford.edu	, plays an important role in the way they express these	attitudes	in speech ,	# " People might think of it being a

Note: This Figure is taken from COCA Corpus using Word Full Info. Query option, through the Corpus website www.english-corpora.org/COCA/

A sample for using KWIC Type of query in COCA Corpus for the adjective "VALUABLE":

SEARCH | WORD | CONTEXT | OVERVIEW

L - - - - 1 2 3 R * RE-SORT ?

K FOR MORE CONTEXT [?] SAVE LIST CHOOSE LIST ---------- CREATE NEW LIST [?] SHOW DUPLICATES

2016	SPOK	NPR_FreshAir	A	B	C	correspondent . You write that cocaine starts to become really	valuable	a little further down the supply chain , not where the coca
2014	SPOK	CBS: Face The Nation	A	B	C	small degree , but there is information that we think is	valuable	about organizational activities that we are fairly confident
2002	FIC	Commentary	A	B	C	so that even to your bitterest enemies you 're much more	valuable	alive than dead . " This , too , though its meaning
2000	ACAD	Education	A	B	C	comments , is that this cooperative oral exam format provides a	valuable	alternative method of assessing student learning . Students
1999	MAG	NaturalHist	A	B	C	Photograph A faucet is locked , above , to protect a	valuable	amenity piped water Opposite page : Karen Josefat carries water
2012	BLOG	...onicles.blogspot.com	A	B	C	water supply of many eastern cities , and a unique and	valuable	American culture that has endured for generations . # Please
2012	WEB	conservativebyte.com	A	B	C	analysis Col. West is not reelected we have lost a very	valuable	American in congress . The worst part is that anti-American
2012	WEB	cjr.org	A	B	C	'm back to " the way something is written is still	valuable	and affects the amount of information which the public actually
2012	BLOG	...ntiagingbydesign.com	A	B	C	, you can repeat something like : " I am a	valuable	and hard working individual who is worthy of a fulfilling new
2009	ACAD	CommCollegeR	A	B	C	for both . So I feel that what I did was	valuable	and helpful to the students . " The verb advocate seemed to
2012	WEB	ncbi.nlm.nih.gov	A	B	C	the extent of individual benefit derived . Young people have	valuable	and interesting contributions to make to the debate about
2008	ACAD	Education	A	B	C	of career decisions and will result in a more focused ,	valuable	and productive outlook with respect to the high school course of
1999	ACAD	InfoSystems	A	B	C	of these benefits will reveal the conditions under which COA is	valuable	and where it is likely to be used . # Direct Costs
1992	TV	Seinfeld	A	B	C	n't he just pay a mover ? He 's got some	valuable	antiques . He 's worried they 'll break something . Soon he
2012	WEB	antipope.org	A	B	C	is that if you flood the market , they are n't	valuable	any more . This really caused huge problems for Spain once they

A sample for using COLLOCATES Type of query in COCA Corpus for the verb "SURVIVE":

COLLOCATES SURVIVE VERB Advanced options Collocates Clusters Topics Dictionary Texts KWIC

+ NOUN		NEW WORD	?
520	2.58	attack	
361	3.17	winter	
354	4.27	crash	
319	2.78	species	
288	2.62	attempt	
183	2.81	struggle	
167	6.81	america	
146	5.51	ordeal	
128	4.36	assassination	
125	2.89	odds	
123	3.43	us	
121	6.24	onslaught	
114	3.65	blast	
110	3.30	encounter	
109	3.24	bacteria	
108	3.96	scrutiny	

+ ADJ		NEW WORD	?
205	4.75	intact	
99	3.01	harsh	
75	6.26	unscathed	
68	2.98	brutal	
36	2.53	freezing	
32	4.33	hardy	
26	3.35	miraculous	
24	2.51	horrific	
16	3.14	turbulent	
16	4.41	harrowing	
14	4.09	unharmed	
13	3.33	tumultuous	
12	3.08	arid	
11	2.71	prehistoric	
11	2.81	frigid	
11	3.50	adaptable	

+ VERB		NEW WORD	?
576	3.11	manage	
571	6.09	thrive	
525	3.93	struggle	
170	5.91	prosper	
144	2.51	enable	
137	4.46	reproduce	
122	2.93	adapt	
83	4.32	flourish	
7	2.99	behead	
5	2.81	winter	
4	2.54	scavenge	
4	3.80	procreate	
3	2.79	harrow	
3	3.23	covenant	
3	3.82	prostitute	
3	5.61	subfreezing	

+ ADV		NEW WORD	?
384	3.16	somehow	
371	3.55	barely	
120	6.25	miraculously	
52	2.91	financially	
40	2.64	economically	
30	3.39	narrowly	
25	2.65	underground	
24	2.94	indefinitely	
6	3.25	handily	
5	2.89	situ	
4	2.83	precariously	
4	4.35	tenaciously	
3	2.98	heroically	
3	3.02	strong	
3	3.39	piecemeal	

A sample for using <u>CLUSTERS Type</u> of query in COCA Corpus for the verb "miss":

CLUSTERS MISS VERB See also: NOUN LIMIT: Loose Medium Tight Phrasal Collocates Clusters Topics Dictionary Texts KWIC

6557	miss you	11136	i miss	1231	missing the point	3002	do n't miss	131	miss a local story	1368	n't want to miss
3797	missed you	7984	i missed	1231	miss out on	2260	gon na miss	105	missing and exploited children	470	you ca n't miss
3690	miss it	6810	n't miss	1230	missing out on	2048	going to miss	87	missing out on something	457	i do n't miss
3150	missed it	4437	you miss	932	missed out on	1653	want to miss	67	missed the last two	336	i would n't miss
3058	missing from	4316	you missed	653	missed the point	990	did i miss	64	missing the whole point	306	you do n't miss
2542	miss him	3418	have missed	557	miss the point	964	did n't miss	64	missed the boat on	280	i did n't miss
2197	missing in	2933	he missed	501	misses the point	950	ca n't miss	59	missing the point here	258	not to be missed
1793	miss her	2258	will miss	480	missed you so	696	would n't miss	57	missing the big picture	184	n't wan na miss
1764	missing out	1657	we missed	472	missing in action	672	do you miss	53	miss all the fun	179	i must have missed
1568	miss out	1470	had missed	406	miss you so	658	i will miss	50	missed the first two	168	you wo n't miss
1455	missing something	1432	we miss	352	miss you too	581	wo n't miss	45	missed it the first	160	not want to miss
1332	miss me	1255	she missed	351	missing a beat	562	did you miss	45	miss you when you	147	i think you missed
1219	missed out	1160	went missing	325	missed the first	543	you are missing	44	miss the days when	146	do n't you miss
952	miss them	1109	just missed	310	miss a beat	458	may have missed	44	missed the cut in	146	you may have missed
941	missing for	1075	not miss	253	missed you too	454	will be missed	42	missing in my life	146	seems to be missing
868	miss fine	1062	never miss	243	miss a story	444	i really miss	40	missed the first time	144	would n't have missed
857	missed him	1056	would miss	236	missed the last	431	has been missing	40	miss me when i	140	you did n't miss

NOTE: *the clusters are arranged according to their frequency level in each column, whereas each column represents a specific number of words and a specific structure for the cluster.*

Name: _____ **Date:** _____

Appendix C: Activities and printouts for Vocabulary teaching using COCA Corpus

A. ***Read the concordance lines below and answer the following questions.***

Attitudes:

approaches to difficult problems , focused specifically on	attitudes	about crime , the findings can be used to understand the
you . It 's just couches ! " # On her	attitude	about money " One of the big things I loved about growing
n't really make that much difference in terms of their own	attitudes	about the nominee . MR-WARREN : It made a great deal of
' use of homophobic epithets . Group-level homophobic	attitudes	and bullying behavior were included at Level 2 to predict the
. " We 've seen remarkable changes in the kids '	attitudes	and confidence levels , " says Pam . " Our introvert ,
Quick to support . Fine range in zone coverage . Great	attitude	and effort . Can get separated from a man-to-man coverage .
urge to relieve suffering toward revising the static , habitual	attitudes	and norms that created it . Zen koan study , contingency "
knowledge of the harmful effects of noise actually affect the	attitudes	and practices . Study design # This is a cross-sectional
mid-1980s , and some people assert that spouses fulfilled their	attitudes	and responsibilities for decades before that . However , the

1. What does the word "attitude" mean?
2. What part of speech is it?
3. What are the prominent collocates in these sentences?
4. Tick the appropriate synonyms for the word attitudes? View, stance, opinion, feeling, richness, academic.
5. Write a sentence on your own using this word?

B. ***Read the following synonyms and clusters for the noun challenge then answer the following questions.*** **Challenges:** n.

SYNONYMS (more)

test contest, encounter, task, test, trial

CLUSTERS (more)

challenge •	challenge to • challenge for • challenges to • challenges in • challenge in • challenges for • challenges facing • challenge from
• challenge	biggest challenge • new challenges • real challenge • big challenge • new challenge • legal challenges • greatest challenge • major challenge
challenge • •	challenges and opportunities • challenges faced by • challenges we face • challenge for me • challenges posed by • challenges associated with • challenge for us • challenges that we
• • challenge	to the challenge • as a challenge • to the challenges • about the challenges • for the challenge • meet the challenges • with the challenges • with the challenge
challenge • • •	challenge is to make • challenge is to find • challenges that we face • challenges and opportunities for • challenge is to get • challenges of our time • challenges of the future • challenge to the law
• • • challenge	can be a challenge • up to the challenge • one of the challenges • it was a challenge • part of the challenge • will be a challenge • it is a challenge • rise to the challenge

1. What does the word "challenge" mean in Arabic?
2. Can you use one of two words clusters examples in a correct sentence about yourself?
3. What part of speech is this word, does it have another part of speech?
4. Can you write two examples using three words cluster before and after?

(Teacher instructions on Collocations)

Students will use a Collocates Query Search Type to explore collocate information for words as well as understand what a collocate means. Students will be divided into groups of three

Introduce the idea of collocates: Collocates are words that other words ―hang out with‖ or words that are commonly found together. For example in English you can say: have a meeting as in ―I have a meeting at 3 p.m. but you cannot say ―do a meeting‖ as in ―I do a meeting at 3 p.m. Those words don't go together.

Students are probably familiar with the idea of collocates without knowing it. Ask students to fill in the blank of the following phrases:

1. Merry _____ (Christmas)

2. Happy _____ (Birthday, New Year)

3. Ride your _____ (bike, motorcycle)

4. Shake his _____ (hand)

There is no rule why any of these words belong together. That's a part of leaning words.

A frequency dictionary gives us information about other words a word hangs out with. The dictionary will tell us what the part of speech is, and if the word comes before or after. Look at the following example:

(Source: *A Frequency Dictionary of Contemporary American English,* Davies & Gardner, 2010)

This entry tells us that the word *family* is a noun. Adjectives that occur with the word family are: (adjectives are usually before the word) *whole family, extended family, royal family, entire family* etc. Nouns that occur with the word (nouns are usually after the word) are *family member, family friend, family child, and family life.*

Verbs that occur with the word family are *live in a family, raise a family, feed the family.* (The small dot tells us if the word family comes before or after the collocate if it's not the normal pattern).

To look up words in a frequency dictionary, do not look in alphabetical order like a regular dictionary. You must look alphabetically in the index, and it will give you a number to look up. The numbers show the order of how often the words are used in English.

147 family *n*
adj whole, extended, royal, entire, poor, nuclear, immediate, wealthy, middle-class, low-income noun member, friend, child, life, .history, support, parent, .planning verb live, raise, own, visit, gather, feed., belong, extend, reunite, .flee

A. Read the following concordance lines and try to figure out the frequent collocates with the adverb "away from"

But we are , from my perspective , we are far	away from	a firm decision so far . # MODERATOR : Mr. Muller
place as I am. # " I 'm a few chapters	away from	a Ph.D. , Ivan , " she said . " But
irritation , she leaned forward to nudge the bulging case-file	away from	a teetering pizza carton , and a trailing piece of wrap
so much ? # I know you 're straining to keep	away from	A VIEW TO A KILL , but any bottom-barrel Bond list
in transition . It was always going to prove difficult moving	away from	an over-reliance on the Brazilian wing wizard and in many
and behaviors around freedom to make mistakes ? Does it take	away from	attention to detail and quality , or does it open innovation
RIC / The Black and Tans turn tail and flee /	Away from	Barry 's coll-yum . ' Do you know that song ?
) and turned around . # Dan was probably a step	away from	bed , but he never showed it . He came right
I never will be able to shelve it . Time taken	away from	being in the forum will include : story production , new

1. Can you think of a synonym for this adverb?
2. What is the most frequent grammatical structure that this adverb forms (Colligation)?
3. Is it followed by a noun or a verb or what else?
4. Write an authentic example about yourself using this adverb?

B. Read the following concordance lines and try to TRANSLATE /DEFINE the word "dormitory":

……..Georgia State University has four new ***dormitories*** arranged in a traditional quad south of Georgia Tech near the …….

In the old days, when girls lived in their own ***dormitories*** and boys came for dates, the girl had to come back…….

….. links to friends and advertise their websites with fliers in a ***dormitory*** bathroom at their colleges. Mr. Brown was at ……..

…. In the spring, just before I moved from the ***dormitory*** into an apartment across town…..

1. What does this word mean?
2. Can you find a synonym for this word?
3. What part of speech is it?
4. Write down two sentences about your college's dormitory?

C. Read the following concordance lines and choose ONE word that comes with all of them:
challenges – lifestyles – opportunities - dormitories

leading rebounder in the series . # " He is getting	and taking advantage of them , " Hill said of Grant .
energy will lead the mayor electorally . Two possible	await back in his home state : to run for senator if
(Voiceover) You 've got him inebriated , he 's got the	because he 's passing by there about the right time .
get in line and have their events there . Another missed	because of the fear and vanity of some clueless CEO . Sigh
lot of blog entrepreneurs are able to find completely new job	by using their own site visitors . This can be a person
mob lawyer and that maybe strings were pulled to create certain	for Adama . Those are definitely interesting and complex
from six to 55 inches , these bass can provide fishing	for all kinds of anglers . # As a group , true
' demand for online courses . # Online education also provides	for community colleges to fulfill their workforce and economic

D. Read these concordance lines then try to fill in the gaps using one of these key words:

(get along – major – excuses – reasons – values – worth)

Those _____ are embodied in our Constitution and they were the basis of the Republican Party when…..

…….has a ' D ' under the date, it 's --------- **at** least $125 . However , if there is no mintmark ,

They 'll make up --------- **about** how the Mayan calendar was inaccurate , or they 'll just

show that white, middle class Americans held common ---------- **about** the appropriate expression of emotions.

they won't and they're selling off some of their ---------- oil **company** ,Aramco, so that they can raise money for

three weeks before an extremely tight presidential election. # Alabamians have good ---------- to be nervous.

So what do you do when you don't ---------- with your boss? Well sometimes the best advice is really to just move..

WRITE THE PART OF SPEECH FOR EACH KEY WORD:

- **get along** ( )
- **major** ( )
- **excuses** ( )
- **reasons** ( )
- **values** ( )
- **worth** ( )

WRITE THREE SENTENCES OF YOUR OWN USING THE PREVIOUS STUDIED WORDS:

1. ----------
2. ----------
3. ----------
4. ----------

E. Match each word to its appropriate concordance lines:

1. Attend college ( )
2. Get a long ( )
3. Miserable ( )
4. Memorable ( )
5. Strangers ( )
6. Unfamiliar places ( )
7. Worth the cost ( )
8. Academically ( )

a. to provide incentives to people living in America illegally to ---------- through the popular vote…..
b. care of a seven-year-old? Yeah. We ---------- great. And besides, I got to learn sometime. I mean…
c. becomes a survival skill when navigating in ---------- , a situation, for example, in which international students frequently find..
d. about the purpose of higher education and whether it's ---------- Even on this narrow aspect of the supply and demand of…….
e. and longer summer, means that more students will struggle ---------- and teachers will be forced to spend more time on catch-up and remedial 911.
f. few animals now a days. It literally makes me ---------- that hunters just kill them for a thrill…..
g. Scuba Diving # One of the best and ---------- experiences is seeing the underwater beauty of this island. Through this you would…..
h. I am perpetually humbled to receive it, like a ---------- invited to dinner and fed the best food in the house.

F. According to the previous concordance lines try to figure out the part of speech (P.O.S) for each key word and its Arabic meaning:

K.W	P.O.S	ARABIC MEANING
1. Attend college ( )		----------
2. Get a long ( )		----------
3. Miserable ( )		----------
4. Memorable ( )		----------
5. Strangers ( )		----------
6. Unfamiliar places ( )		----------
7. Worth the cost ( )		----------
8. Academically ( )		----------

G. Match each word to its appropriate concordance lines:

1. Atmospheric conditions (	a) At 5 a.m., the storm -- now known to ---------- as a post-tropical cyclone -- was roughly 90 miles west of Philadelphia and was
2. Extreme weather ( )	b) of " big three " of mobile platforms has the ---------- situation. BlackBerry sales have decreased almost 35%, and overall market share fell
3. Health disorders ( )	c) is that over these last four years, together, ---------- but surely, we have been pulling ourselves out of that hole that we
4. Floods ( )	d) forced to face flooding and drought in the summer and ---------- in the winter, and what grows in these conditions? not crops but…
5. blizzards ( )	e) by natural disasters such as the Japanese tsunami, Thai ---------- and -- most recently -- Hurricane Sandy. Lean inventories and supply chains are
6. slowly ( )	f) depression is now considered one of the main mental ---------- based on worldwide data and indicators (World Health Organization 2018), and
7. worse ( )	g) in the northern hemisphere, that can result in persistent ---------- such as droughts, heat waves and flooding "... # Other leading ice
8. meteorologists ( )	h) Another example is to study the influence of the ---------- such as the temperature and air pressure on the shower development. Under…..
9. stroke ( )	i) 82 per cent of premature deaths from heart disease and ---------- could be avoided if the main risk factors; tobacco, unhealthy diet and
10. asthma ( )	j) a preexisting condition -- let's say, diabetes or ----------. And here's something that really gets me -- that if you get
11. in contrast to ( )	k) ---------- persistent myths about government spending in Wisconsin, new figures released by the
12. at least ( )	l) a friend you can't trust. # Storm ---------- (with the odd exception) have always been honest about their intentions

H. According to the previous concordance lines try to figure out the part of speech (P.O.S) for each key word and its Arabic meaning:

K.W	*P.O.S*	*ARABIC MEANING*
1. Extreme weather	( )	----------
2. blizzards	( )	----------
3. slowly	( )	----------
4. worse	( )	----------

5. bio meteorologists ( ) ----------
6. stroke ( ) ----------
7. asthma ( ) ----------
8. in contrast to ( ) ----------

E. Check the Highest Frequency Genres for Each Word through Using (Word FREQUENCY)

SEARCH TYPE:

H. Frequency Genre **TARGET WORDS**	**Highest Frequency Genres**						
	Academic	**Spoken**	**Web**	**TV/Movies**	**Fiction**	**Magazine**	**News**
asthma							
meteorologists							
in contrast to							
blizzards							
Extreme weather							
slowly							
Health disorders							
Atmospheric conditions							

GROUP WORK TASKS – HOME ASSIGNMENTS - SAMPLE A

Complete the table below, following the stated instructions:

- Work in GROUPS, you and your PARTNERS together to fill in the table.
- Go to the website (www.english-corpora.org/COCA/
- Search for the words (low carb diet – nutritionists – diabetes – fast food - desserts)
- Use Key Word In Context (KWIC) Type of search.
- AS A LAST OPTION, you can use the available links for bilingual dictionaries on the website

Words/ elements	low carb diet	nutritionists	Diabetes	Fast food	desserts
Part of speech					
meaning					
Collocations before					
Collocations after					
synonyms					

2. Write ONE personal sentence for each word (low carb diet – nutritionists – diabetes – fast food - desserts).

1. ----------
2. ----------
3. ----------
4. ----------
5. ----------

GROUP WORK TASKS – HOME ASSIGNMENTS – SAMPLE B

Complete the table below, following the stated instructions:

- Work in pairs, you and your partner together to fill in the table.
- Go to the website (www.english-corpora.org/COCA/
- Search for the words **(preferences – soy products – diet – complex carb - probably).**
- Use Key Word In Context (KWIC) Type of search.
- AS A LAST OPTION, you can use the available links for bilingual dictionaries on the website

Words/ elements	preferences	soy products	diet	complex carb	probably
Part of speech					
meaning					
Collocations before					
Collocations after					
ONE Sample concordance line					

2. Write ONE personal sentence for each word (preferences – soy products – diet – complex carb - probably).

1. ----------
2. ----------
3. ----------
4. ----------
5. ----------

GROUP WORK TASKS – HOME ASSIGNMENTS – SAMPLE C

Complete the table below, following the stated instructions:

- Work in pairs, you and your partner together to fill in the table.
- Go to the website (www.english-corpora.org/COCA/
- Search for the words **(humidity – moods – pneumonia – affect - influence).**
- Use Key Word In Context (KWIC) Type of search.
- *AS A LAST OPTION*, you can use the available links for bilingual dictionaries on the website for translation, synonyms and words pronunciation.

Words/ elements	humidity	moods	pneumonia	affect	influence
Part of speech					
meaning					
Most Frequent Clusters					
Synonyms					
ONE Sample concordance line					

2. Write ONE personal sentence for each word (humidity – moods – pneumonia – affect - influence).

1. ----------
2. ----------
3. ----------
4. ----------
5. ----------

GROUP WORK TASKS – HOME ASSIGNMENTS – SAMPLE D

Complete the table below, following the stated instructions:

- Work in pairs, you and your partner together to fill in the table.
- Go to the website (www.english-corpora.org/COCA/
- Search for the words **(variety – safety – privacy – interests - lifestyles).**
- Use Key Word In Context (KWIC) Type of search.
- *AS A LAST OPTION*, you can use the available links for bilingual dictionaries on the website for translation, synonyms and words pronunciation.

Words/ elements	variety	safety	privacy	interests	lifestyles
Part of speech					
meaning					
Most Frequent Clusters					
Other related forms for the word (noun – verb – adjective- adverb - others)					
ONE Sample concordance line					

Then, Write ONE personal sentence for each word (variety – safety – privacy – interests - lifestyles).

1. ----------
2. ----------
3. ----------
4. ----------
5. ----------

GROUP WORK TASKS – HOME ASSIGNMENTS – SAMPLE (E)

Complete the table below, following the stated instructions:

- Work in Groups, you and your classmates together to fill in the table.
- Go to the website (www.english-corpora.org/COCA/
- Search for the words **(humidity – privacy – interests – lifestyles – pneumonia – diet – probably - diabetes).**
- Use Word Frequency (FREQUENCY) Type of search.

AS A LAST OPTION, you can use the available links for bilingual dictionaries on the website for translation, synonyms and words pronunciation.

H. Frequency Genre	**Highest Frequency Genres**						
TARGET WORDS	**Academic**	**Spoken**	**Web**	**TV/Movies**	**Fiction**	**Magazine**	**News**
humidity							
privacy							
interests							
lifestyles							
pneumonia							
diet							
probably							
diabetes							

Write a CONCORDANCE LINE FROM NEWS genre for the word (Privacy): ----------

Write a CONCORDANCE LINE FROM ACADEMIC genre for the word (pneumonia): ----------

Write a CONCORDANCE LINE FROM SPOKEN genre for the word (diet): ----------

Write a CONCORDANCE LINE FROM TV/MOVIES genre for the word (probably): ----------

Write a CONCORDANCE LINE FROM FICTION genre for the word (lifestyle): ----------

---