



***Learners Attitudes Towards Utilizing
Corpusbased Approach in the Saudi EFL Context***

Ahmed Fathy Farag-Allah Youssef

**PhD Research Scholar, English Department,
Sudan University of Science and Technology,**

SudanEnglish Language Lecturer

, Department of English Language,

Faculty of Humanities and Administrative Sciences,

Burayadah Colleges

Email: eng.department.bpc@gmail.com

Tel: +966546511761





ABSTRACT:

Corpus Linguistics has been noted to be a trendy approach in language research in the last few decades as it is closely related to technological advancements. In the current study, the researcher has investigated the learners' attitudes towards using corpus-based approach in their EFL classroom and figuring out the potential challenges in using corpora. The researcher has employed quantitative methodology to investigate the study questions. A questionnaire has been employed on the experimental group's learners (N=25) to collect their views on three themes, benefits of corpus-based approach, their overall perceptions on the use of corpora, and potential challenges in corpora use in vocabulary learning. The obtained data has been descriptively analyzed using SPSS program to answer test the study hypothesis. Findings showed that learners had positive attitudes towards using corpus-based approach and they highly rated its benefits, while they illustrated facing some challenges in using corpora which could be met over the time and by much practice. Some recommendations could be formulated from the apparent results of this study: First, bridging the gap between corpus research results and the actual teaching practices in the Saudi EFL context could be achieved by the gradual adoption of corpora use. Third, supporting learners with sufficient training to possess technical skills of online corpora use and concordancing skills of linguistic analysis and interpreting.

Key Words: *Corpus-based approach, COCA Corpus, vocabulary learning, Learners Autonomy, Learners Attitudes, Concordancing skills*



Introduction

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 is 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). Corpus has been defined by Crystal (1997) as “*a large collection*



of linguistic data, either written texts or transcription of recorded speech, which can be used as a starting point of linguistic description or as a means of verifying hypotheses about a language". This collection of massive authentic data is electronically stored in readable and searchable forms which enable users doing a variety of linguistic analysis and description. The term 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 materials, concordancing, concordancers, etc. (Barbieri & Eckhardt, 2007).

In a classroom, Corpora can be utilized directly or indirectly in teaching English; i.e. indirectly in designing teaching materials and in syllabus design, while directly through the direct use of corpora inside or outside the classroom (Gavioli & Aston, 2001, Römer, 2011). In employing corpus-based approach learners are expected to be independent researcher and autonomous language learners to explore language patterns and cooccurrences, rather than passive knowledge recipients (Conrad, 2005).

This study is a part of a broader experimental study that investigated the effectiveness of corpus-based approach in learning vocabulary in the Saudi EFL tertiary level. In this paper the researcher will focus only on learners' perceptions on corpus-based approach after going through a six-week treatment. The first spark that initiated the researcher's interest in this topic is attributed to his successful and rich experience in teaching legal translation to languages department four years back using parallel corpora to teach legal texts and terminology.

Problem Statement

The researcher has noticed that a great bulk of recent studies on ELT have agreed upon the importance of learners' involvement, authenticity of language input, and contextualized vocabulary



learning. Corpus linguistics has already claimed to be providing authentic data input, enhance learners' autonomy, and promote learner-centered classroom. 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. 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 handled the idea of forming specialized corpora or checking or using corpora for linguistic analysis.

To the researcher's knowledge, no studies have surveyed EFL Saudi learners' perceptions on this approach to enhance corpus application on an evidenced-based suggestion. To fill in this gap, the researcher has designed this study to thoroughly investigate learners' attitudes towards corpus as a promising tool in learning vocabulary in the Saudi EFL context, hoping this study could improve and enhance the experience of corpus use. Moreover, investigating the potential challenges and difficulties perceived by learners who have gone through real classroom treatment, could honestly pinpoint the weaknesses of this approach and then consolidating corpus-based approach models and guidelines of application. Therefore, this investigation could bridge the gap in the Saudi context through an action classroom research and quantitative methodology and contribute to the corpus literature.

Questions of the Study

The following research questions have been formulated:

1. What are learners' attitudes towards corpora benefits in learning vocabulary?
2. What are learners' overall perceptions on the use of corpus-based approach in vocabulary learning classes?

-
-
3. What are their attitudes towards the potential challenges and difficulties in using corpus-based approach?

Objectives of the Study

The researcher has made the following objectives:

- 1- Investigating experimental group's learners' attitudes towards the benefits of corpus-based approach adoption in learning vocabulary?
- 2- Determining Experimental group learners' overall perceptions on the use of corpus-based approach in their vocabulary learning classes.
- 3- Investigating experimental group's learners' perceptions on the potential challenges or difficulties of corpus-based approach application.

Literature Review and Previous Studies

Learners should have opportunities to develop a feel for language and enhance their discovery learning through given direct access to corpora use as stated by Huang (2011). One of the recent corpus studies is Boontam and Phoocharoensil (2018) which was conducted on 30 Thai learners to investigate the effect size of employing corpus-based approach. The researcher applied a treatment for six weeks (one 50-minute class every week). Results of pre and posttests illustrated some improvements in vocabulary learning as a result of using corpus-based activities and exercises. Moreover, results of questionnaires and interviews revealed learners' positive attitudes towards using corpus approach. Another study, Al-Mahbashi et al (2016) which has been applied on 77 female university students in Sana'a University, Faculty of Education to investigate the effectiveness of DDL approach on writing skills. They have employed COCA Corpus (Corpus of Contemporary American English) (COCA) in creating the corpus-based hands-on and tasks for the experimental group.



The results of pre and posttests analysis for both groups, experimental and control groups, revealed significant differences between the mean scores of their post test scores in favor of the experimental group which applied corpus-based approach. Similarly, the differences in delayed post test scores were significant in favor of the experimental group. the researcher concluded that corpus-based group outperformed its counterpart in both learning gains and retention rates. They concluded stating *'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)'*.

On the other hand, some studies investigated the use of corpora in learning other languages, rather than English, like Çelik and Elkatmiş (2013). They investigated the effect of corpus-based approach on learning Turkish punctuation marks compared to traditional lecture-based approach. They conducted an experimental study on 171 university students at Kirikkale University, in Turkey. They used an achievement test on punctuation marks. Results of post-tests have revealed significance differences between both groups in favor of the experimental group. Furthermore, based on the questionnaire responses and interviews the majority of experimental group's learners in revealed positive perceptions on studying punctuation marks through obtained concordances lines

Methodology

As I previously mentioned that this study is a part of a broader one which 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 and learners



attitudes towards this approach. This type of research, quasi-experimental or experimental research designs 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 the broad experimental 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.

For this study, the researcher has employed quantitative methodology to answer the study's three questions. As learners have been give a questionnaire of 16 items at the end of the experiment (a six-week treatment of corpus-based approach to learn three chapters of reading topics and 116 items of vocabulary).

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. Lastly, the quantitative data, taken from tests results and questionnaires responses, has been statistically analyzed using Statistical Program for Social Sciences Package (V.25).

Mainly, 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.

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.*

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

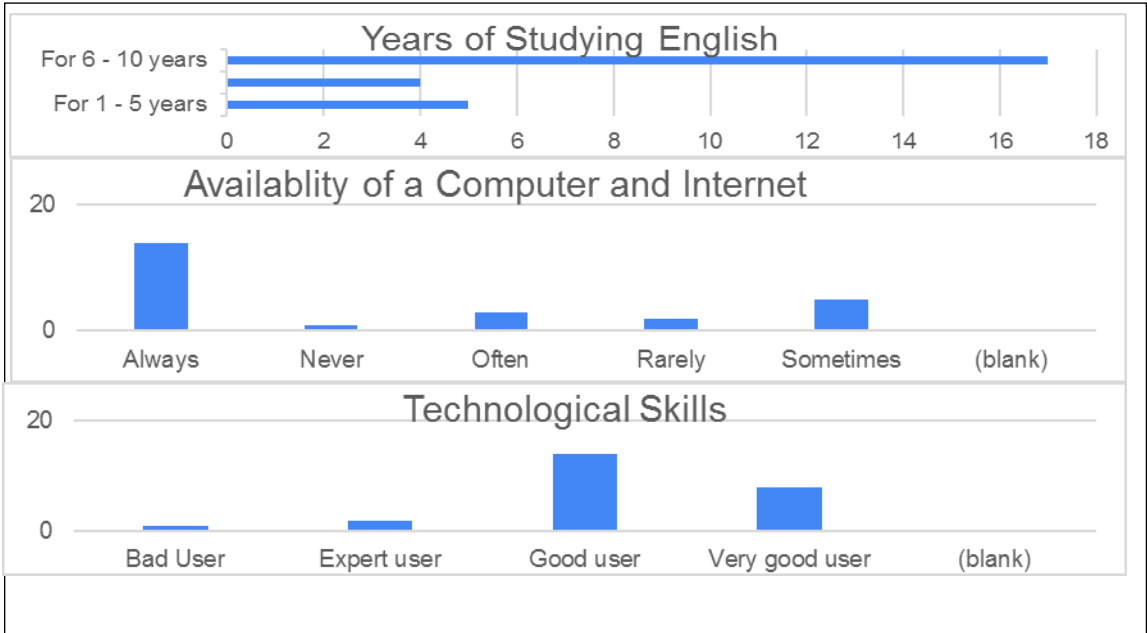


the control and the experimental condition following a chance procedure (Gravetter & Forzano, 2018; Hatch & Lazaraton, 1991; Kirk, 2009; Loewen & Plonsky, 2016; Nunan, 1992). One class has become 29 students and has been assigned as the control group (traditional approach for teaching reading and vocabulary), 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. 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. The attitudes questionnaires have been used only with experimental group's learners (25 students).

Experimental Group's Participants Demographic Data

Earlier, the researcher collected the personal data of participants and quantitatively analyzed it. The main findings for this part can be summarized in the following bar charts in figure (1)

Figure 1: Participants Demographic Data



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.

Data Collection Tools

To collect the data, the researcher has employed a questionnaire of 16 items that have been used at the ninth week of the treatment.

Questionnaire

In order to investigate Saudi EFL learners’ attitudes towards and views on integrating corpus-based approach in their vocabulary classrooms, a questionnaire has been conducted immediately



after the delayed posttests, at the commencement of the ninth week of the treatment (See Appendix A for Learners' Questionnaire). 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) study.

To make sure that the questionnaire is valid and reliable for collecting the data and measuring the assigned objective, 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. First, the researcher received some valuable suggestions and feedback from the Jury and accordingly he did some changes to the items of the questionnaire. 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 1 Questionnaire Validity According to the Jury Evaluation), there was a high percentage of satisfaction among the jury on the validity of the questionnaire.

Table 1 Questionnaire Validity According to the Jury Evaluation

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

Data Analysis

The questionnaire has been 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 the survey has been quantitatively analyzed by using SPSS software, through descriptive statistics analysis features to answer the study's three questions.

Results and Discussions

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. 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 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.

Students' Perceptions on the Use of Corpus-based Approach

The following table is presenting an overall overview for the three main themes of the questionnaire, overall perceptions, corpora benefits, and potential difficulties. According to table (Table 2), 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% were still neutral and only 3.4% disagreed.

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 2: Summary of Descriptive Statistics for the questionnaire 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 in vocabulary learning; this theme included the items 6, 7, 8, and 10. The following table (*Table 3*) shows the mean scores for participants response for each item. It can be seen that the majority of 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 and believed in the effectiveness of COCA corpus activities and tasks in vocabulary learning. 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. This finding answered the first questions of the study about learners' perceptions on corpus-based approach benefits in vocabulary learning. Thus, the first hypothesis has been met and proved as learners revealed high positive attitudes towards benefits of using corpora.

Table 3: 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



Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
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*) 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. 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 overcome some challenges and problems over the course of time and they got familiarized to this new approach.

Table 4: 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



Items	N	Percentages of Ratings					Mean	Std. Deviation
	Statistic	1	2	3	4	5	Statistic	Statistic
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

Items	N	Percentages of Ratings					Mean	Std. Deviation
		Statistic	1	2	3	4	5	Statistic
Overall Evaluation for the Theme	24	2%	5.75%	27.88%	40.40%	24.05%	3.833	0.8912

To recap, this theme had four items that measured general satisfaction of participants about using corpus-based approach. This theme represented the second question of the study (*What are learners' overall perceptions on the use of corpus-based approach in vocabulary learning classes?*). 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.

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 5*) 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 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.

Table 5: 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

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. This theme represented the third question of the study (*What are their attitudes towards the potential challenges and difficulties in using corpus-based approach?*). 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.



Implications and Suggestions

At the end of the experiment, learners of the experimental group have revealed positive attitudes towards the use of corpus-based approach in their vocabulary learning 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 by much practice. In addition, the majority (approximately 73.1%) demonstrated their desire to use corpora at their own in the future, while only 7.7% disagreed. This finding is in alignment with the majority of corpora literature which proved the effectiveness of corpora utilization in vocabulary learning and retention and learners' positivity of attitudes. It is aligned with the results of these corpora studies (Chan and Liou, 2005; Kaur and Hegelheimer, 2005; Hadley, 2002 and 2017; Sirphicharn, 2003; Horst et al., 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.

Hadley (2017) 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 a 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.

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 can 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.



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 rulesetc. 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.

Suggestions for Teaching Practices

Given the results of this study, the researcher has formulated the following implication for an effective corpus use in the classroom:

1. Based upon the researcher noticing in this experiment, 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. 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.

3. 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 most applicable for various learning objectives.

4. 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.

Recommendations for Further Study

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 second recommended study 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 of teachers' lack of training in EFL context. Another suggested study could 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.

REFERENCES

- Al-Mahbashi, A. et all. (2015). The Effect of Data Driven Learning on Receptive Vocabulary Knowledge of Yemeni University Learners, The Southeast Asian Journal of English Language Studies – Vol 21(3): 13 – 24, Retrieved from: <http://ejournals.ukm.my/31/article/view/9511/0>
- Aston, G. (2002). The learner as corpus designer. Teaching and learning by doing corpus analysis, Brill Rodopi: 7-25.
- Barbieri, S& Eckhardt. F (2007). Applying corpus-based findings to form-focused instruction: The case of reported speech*. Language Teaching Research 11, 3 (2007); pp. 319–346
- Boontam, P and Phoocharoensil, S. (2018). Effectiveness of English preposition learning through Data-Driven Learning (DDL). 3L: The Southeast Asian Journal of English Language Studies – Vol 24(3), 125 – 141.



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

Chan, T. P. & Liou, H. C. (2005). "Effects of web-based concordancing instruction on EFL students' learning of verb–noun collocations." *Computer Assisted Language Learning* 18(3): 231-251.

Chen, J., et al. (2004). Discovering parallel text from the World Wide Web. *Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation- Volume 32*, Citeseer.

Cobb, T. (1999). Breadth and Depth of Lexical Acquisition with Hands-on Concordancing. *CALL Journal* 12 (4), pp. 345-360

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

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 900+ million words, 1990-present. Available online at <http://www.americancorpus.org>.

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."

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

Gravetter, F.J & Forzano, L. B. (2018). *Research Methods for the Behavioral Sciences*. Cengage Learning, Inc. 978-1337-613316

Hadley, G. (2002). An introduction to data-driven learning. *RELC journal*, 33, 99-124.

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

Hatch, E. M. and A. Lazaraton (1991). *The Research Manual: Design and Statistics for Applied Linguistics*, Newbury House Publishers.



Horst, M., et al. (2005). "Expanding academic vocabulary with an interactive on-line database." *Language Learning & Technology* 9(2): 90-110.

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

Hunt, A. and D. Beglar (2005). "A framework for developing EFL reading vocabulary." *Reading in a foreign language* 17(1): 23-59.

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

Kirk, D. (2009). *Physical education futures*, Routledge.

Kirn, E., et al. (2003). *Interactions 1: reading*, McGraw-Hill 2002.

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

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

Loewen, S. (2018). *Instructed second language acquisition. The Palgrave handbook of applied linguistics research methodology*, Springer: 663-680.

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

Nation I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press: Cambridge

Nunan, D. and N. David (1992). *Research methods in language learning*, Cambridge University Press.

Pérez-Paredes, P., et al. (2012). "Learners' search patterns during corpus-based focus-on-form activities: A study on hands-on concordancing." *International Journal of Corpus Linguistics* 17(4): 482-515.

Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225.
<http://dx.doi.org/10.1017/S0267190511000055>

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



Sripicharn, P. (2003). Evaluating data-driven learning: the use of classroom concordancing by Thai learners of English. *Thammasat Review* (pp 203-236). ISSN: 2630-0303. Retrieved from: <http://40909-Article%20Text-93904-1-10-20151013.pdf>

Varley, S. (2008). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22(2), 133-152.

Yoon & Hirvela, (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, Volume 13, Issue 4, December 2004, Pages 257-283. Retrieved from: <https://doi.org/10.1016/j.jslw.2004.06.002>

Zhang, W. (2008) "In Search of English as a Foreign Language (EFL) Teachers' Knowledge of Vocabulary Instruction." Dissertation, Georgia State University.

Zimmerman, C. B. (1997). Historical trends in second language vocabulary instruction. In J. Coady & Th. Huckin. *Second language vocabulary acquisition, a rationale for pedagogy* (pp. 5-19). Cambridge: Cambridge University Press.

Appendix A: Learners Questionnaire.

1. Introduction

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



2. Personal Language and Technology Backgrounds:

1. Your Name (Optional): ----- 2. How long have been learning English? -----
3. Who do you describe your technological skills? -----
4. Do you have access to a computer and internet connection at home or school? -----

3. MAIN PART: Read the following statements and choose the closest response to you, from Strongly Agree (5) to Strongly Disagree (1) according to your personal experiment in using COCA Corpus recently.

Questionnaire's Items	The Items Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
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?					
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?					



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?					
Item No. 8	8. I feel that being involved in corpora individual activities and group tasks increases my vocabulary retention?					
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?					
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.					
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?					
Item No. 12	12. I feel that the queries concordance lines or context examples are overwhelming and may cause frustration or distraction especially.					
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.					
Item No. 14	14. I feel that my language proficiency level impacts my ability to effectively use corpus-based printouts and activities.					
Item No. 15	15. I have some difficulty in using the COCA corpus due to cut-off					



	sentences in some concordance outputs.					
Item No. 17	16. I have some difficulty in using the corpus due to unfamiliar vocabulary on concordance outputs.					