

“The Effectiveness of Corpora in Promoting Vocabulary Acquisition”

Researcher:

Ahmad Abdu Ali Asey

Ala'rdha Education Office, Jazan Region, KSA.

ORCID ID: <https://orcid.org/0000-0002-9120-7837>



Abstract:

The use of corpora e.g., the Corpus of Contemporary American English (COCA) has little to none sufficient support for classroom efficacy in the literature. This paper addresses that ‘gap’ and emphasizes the importance of vocabulary teaching through ‘exposure’ to “authentic texts” found on different corpora. To attest for COCA’s efficacy, an experimental study is conducted on twenty-six Saudi elementary students divided into two groups, control and treatment. Progress is measured using pre-post-tests. Test scores are analyzed quantitatively using SPSS. Students’ perceptions regarding COCA are collected through one-to-one interviews. Results of the study are promising and substantiate the viable potentials of COCA as an educational tool in language teaching.

Keywords: COCA; Vocabulary; Authentic texts; Language teaching.

1. INTRODUCTION

1.1. Background to the Study

” The more one considers the matter, the more reasonable it seems to suppose that lexis is where we need to start from, the syntax needs to be put to the service of words and not the other way round.” (Widdowson in Lewis, 1993: 115)

Language teaching across all educational levels involves wide arrays of preplanning, considerations based on the language, and how to be used in its pertinent culture. The importance of teaching the vocabulary of a language goes hand in hand with the way grammar is taught. However, some linguists have emphasized the role of vocabulary and prioritized it over grammar. This is based on the premise that without grammar very little can be conveyed, and without vocabulary, nothing can be conveyed (Wilkins, 1972:111). Nevertheless, the effective teaching of one can substantiate the other. The situation in which the teaching of ESL/EFL may prelude to failure is when language learners lack sufficient vocabulary of the target language (TL). Therefore, vocabulary is a central and indispensable element in language teaching and learning. Its importance resonates with the use of language in different settings. It is also of paramount importance that vocabulary gets a fair proportion of the curriculum, and presented in a way that ensures relativity to the frequency of words in TL. Nation (1990: 40) states:

“One useful strategy in learning vocabulary is to master a large number of high-frequency words – those words that appear most often in written and spoken texts and are therefore the most useful ones to learn.”

Therefore, local evaluators for EFL/ESL programs are required in order to determine the functionality and practicality of the adopted curriculum from external sources. These evaluators should have a satisfactory acquaintance with the field of Corpus Linguistics (CL), specifically, the use of corpora e.g., the Corpus of Contemporary American English (COCA) which has little to none sufficient support for its classroom efficacy in the literature. This paper attempts to fill that gap and provide evidence of its implications. This will enrich and corroborate the proper usage of words in different settings, their variations, and present a valuable and contemporary awareness for language learners and teachers on how words function in isolation and in context. The reason behind such proposal –although far-fetched, is to put forth the idea of a unified English curriculum that is specifically designed based on a relevant corpus relative to the aspired goals to be achieved. In addition, most, if not all, language learners will undergo similar, if not the same, language proficiency tests e.g., TESOL English Language Proficiency Test (TELPT), Test of English as a Foreign Language (TOEFL), and International English Language Testing System (IELTS), which further supports the need of such an idea. Colossal as it may seem, a universal English curriculum based on CL is recommended for the purpose of quality, fair competition, and to accomplish a sustainable English proficiency, especially in Saudi Arabia. Thus, this paper aims to answer questions as to how CL may be the best-suited alternative to raise students’ language proficiency, and what are the benefits of vocabulary exposure within the scope of CL. It also aims at investigating the role corpora play and their efficacy in promoting vocabulary acquisition and retention through exposure to authentic texts ‘jotted down’ by the native speakers of the language. This may raise awareness to the importance of naturally collected texts in the teaching of vocabulary.

1.2. Vocabulary

Vocabulary acquisition in English language teaching was not given sufficient emphasis in the literature. In order for learners to navigate a language, a verbal repertoire is critical. Preferably, all language aspects must be given adequate attention and must be related to one another in a way that serves to develop competency and performance. The effectiveness of teaching vocabulary would have to entail constant exposure to authentic language found on different corpora. Corpora contain instances of authentic language use or texts that are stored electronically. Data extracted from a corpus can have significant indications as to how to go about teaching vocabulary.

The exposure of authentic texts may help boost retention and recollection of language items. Therefore, the aim would be on presenting relevant materials that language learners will come to encounter while watching movies, playing video games, or doing any activity that involves the use of English as a medium of instruction. CL provides such solutions; and learners can continue to learn on their own after being introduced to the concept of concordances and how they can be utilized on a computer. Such opportunities have the ability to equip students with the tools they need to become autonomous learners. The emphasis of CL in teaching vocabulary can further be substantiated by the benefits it brings to the learning environment. Such benefits are a) focus on linguistic performance, rather than competence; b) focus on linguistic description, rather than linguistic universals; c) focus on quantitative, as well as qualitative models of language; and d) focus on a more empiricist, rather than rationalist view of scientific inquiry (Leech, 1992).

In my years of experience in the field of EFL teaching, I have come to realize that learners with low-level of vocabulary tend to perform worse than their counterparts. Their motivation is also limited, and they often end up resenting the subject as a whole. Therefore, using CL may help recapture and corroborate the ‘entertainment’ factor in the learning environment through the use of corpora on computers. Such a method focuses on how learners can use the knowledge that is distributed outside the classroom, particularly on the Internet (Siemens, 2005); a process Johns (1994) has called “data-driven learning.”

In order to achieve the aforementioned objectives, the following research questions are raised:

1. What is the effect of teaching vocabulary using COCA on Saudi elementary students at Gharbi Al'Otaif, Al'Otaif, and Khabta schools?
2. What are the informants' attitudes towards COCA and its potentials as an education tool for vocabulary teaching?

The study attempts to test the following hypothesis:

- Saudi elementary level vocabulary repertoire at Gharbi Al'Otaif, Al'Otaif, and Khabta schools will improve as result of the exposure to aspects of CL, particularly COCA, e.g., collocations and frequency of words used in the target language`.

2. Review of the Literature

2.1. Corpus-Based Teaching

A corpus can safely be defined as the collection of various instances of texts as per used by native speakers. These texts are affectionately known as “authentic texts.” An authentic text is one created to fulfill some social purpose in the language community in which it was produced (Little et al. 1988: 27). In other words, the use of authentic texts, embracing both the written and spoken word, is helping to bridge the gap between classroom knowledge and a student's capacity to participate in real-world events (Wilkins 1976: 79). Alongside this recognition of the need to practise skills and strategies for the real world, there has been a growing awareness of the importance of active factors in learning, and the use of authentic texts is now considered to be one way of maintaining or increasing students' motivation for learning (Guariento et al., 2001: 347). They give the learner the feeling that he or she is learning the ‘real’ language; that they are in touch with a living entity, and the target language as it is used by the community that speaks it (ibid). One of the benefits of including corpus-based texts in English curricula is the frequent contact with authentic texts that have a high statistical occurrence in media outlets, e.g., movies, which, in turn, may enforce the retention of these texts and can readily be internalized.

Moreover, there have been numerous theories in the field of Second Language Acquisition (SLA) with their own unique orientations. For example, Krashen's Monitor Model (Krashen, 1981), DeKeyser's skill-learning theory (DeKeyser, 1998), and Instructed Language Learning (Ellis, 1984). These theories put more emphasis on the proper methodology of instruction in L2 acquisition rather than attempting to tackle the root of the issue and help create a curriculum that draws the attention of language learners who are motivated enough to acquire an L2, regardless of their language attitudes – instrumental or integrative. As an English teacher, I almost invariably find myself confronted with the questions as to how and when such information (in current English curriculum in Saudi Arabia) can be of help to us (students). By contemplating the reasons behind such questions, I have come to realize that language learners might not be the primary culprit, rather the apparent disassociation of what is being presented to them from the authenticity of TL.

2.2. Areas of Language Teaching

2.2.1. Phraseology

Phraseology is the study of phrases. It is a central element of CL. Sinclair (1991) concluded that the meaning of a word is found through several words in a sequence, through phrases (p170). Phraseology includes the study of collocations, and language occurring in preferred sequences. Collocations are the regular co-occurrence patterns of a target word with other words (Biber et al. 1998:6). Kennedy (2014) studied 'between' and 'through' and found that 'between' is usually used after nouns like differences, distinction, agreement, and meeting, whereas 'through' is more frequently found after verbs such as go, pass, run, and fall (p.107). Therefore, collocation is of much higher importance in terms of language use, acquisition and ultimate success in language learning. The more the learner is capable of producing the correct collocations, the less hesitation pauses he makes in long sequences of words and, consequently; the more competent in the language he becomes.

2.2.2. Preferred Sequences

Phraseology also includes the study of preferred sequences of words. Hunston (2002, pp.9–11) explains that learners often confuse between the adjectives 'interesting' and 'interested', and explanations of their different meanings do not usually help students use the words correctly; looking at the phrases "someone is interested in something", "an interesting thing", "what is interesting", and "it is interesting to see", can give students the ability to use the individual words correctly by providing an established pattern of use for each word. Only through corpus study can we find the details of phraseology—collocations, lexical bundles, and language occurring in preferred sequences.

2.2.3. Syllabus Design

The final area of language teaching that Corpus Linguistics addresses is syllabus design. Phraseology, lexicogrammar, register, ESP, nuances of language—all of these areas can be used to more accurately and efficiently design syllabi by helping to see what learners really need to know about language: frequency and collocation for vocabulary, grammar patterns for different registers, and specific knowledge for specific purposes. For example, the present perfect tense appears in almost every grammar textbook. It is usually defined as "recent past." However, a corpus study revealed that more than 80 percent of the time, present perfect is used to signify "indefinite past" (Mindt, 2000, p. 224). Statistics of this kind help textbook writers, course designers, and teachers set priorities for the classroom. If you, as a teacher, are armed with this type of knowledge, you can supplement course materials with information that is relevant for language learners. Furthermore, this will help with the augmentation and diversification of class corpus-based activities. Students, on the other hand, will benefit significantly from their encounter with the authentic texts of TL and that may even create an incentive for practicing the language amongst themselves, thereby alleviating or eradicating the factor of boredom that arises out of grammar-oriented curricula, which give students an ample amount of mental bandwidth to process.

2.3. Concept and Potentials of CL

Language learners should be acquainted with the nature of concordancing programs (CPs). As a pedagogical tool, CP has typically been associated with 'data-driven learning,' which entails learners taking the role of a researcher to work out the rules and regularities of specific lexical items using concordance data, often with explicit language learning goals (Yoon, 2016). CPs are tools that allow the learner to tap into vast collections of texts, called corpora, to help learners discover how language is actually used. There are also Web-based CPs available: see <http://www.lexutor.ca/conc/eng/> for example.

If you put in a word or phrase, for instance, these programs will search for examples of that item in context and return a listing. Although initially designed for research purposes, language teachers have adopted it as a tool for language learners.

COCA is the most widely freely available corpus of English, it has a built-in CP, and probably the most commonly used. As is the case with most corpora, there are different tabs (List, Chart, Collocates, etc.). A concise explanation will be provided of each to better understand the implications in L2 learning.

- List:** This tab is mostly used when asking a question that is likely to result in multiple answers. For example, what words most frequently occur before or after “cause”? This is achieved by using an asterisk (*) followed by a space before the selected word. The asterisk stands for either a word or punctuation mark (* cause).



Figure 1: Search Results of 'cause'

Figure (1) shows the frequency at which “cause” mostly occurs. The most frequent is the abbreviation of because (‘cause). Followed by the use of ‘cause’ preceded by the definite article the, etc. By clicking on each result, instances of relative language use will appear which will engage the learner in an analytical atmosphere trying to learn the correct usage by looking at words that come before and after the selected one, thus, inferring a general rule that governs its use.

- Chart:** When switching to the chart tab using the previous word, the results generated will be presented according to the frequency of a feature across different text types that comprise the corpus, e.g., spoken, fiction, magazine, newspaper, and academic.

Figure (2) clearly indicates that ‘cause’ is most commonly used in magazines and spoken English, and least used in fiction. Learners can utilize such information, academically, to improve their writing skills. If learners are interested in a set of words or grammatical construction, then the LIST option shows the frequency of each matching form and the total frequency in each section.

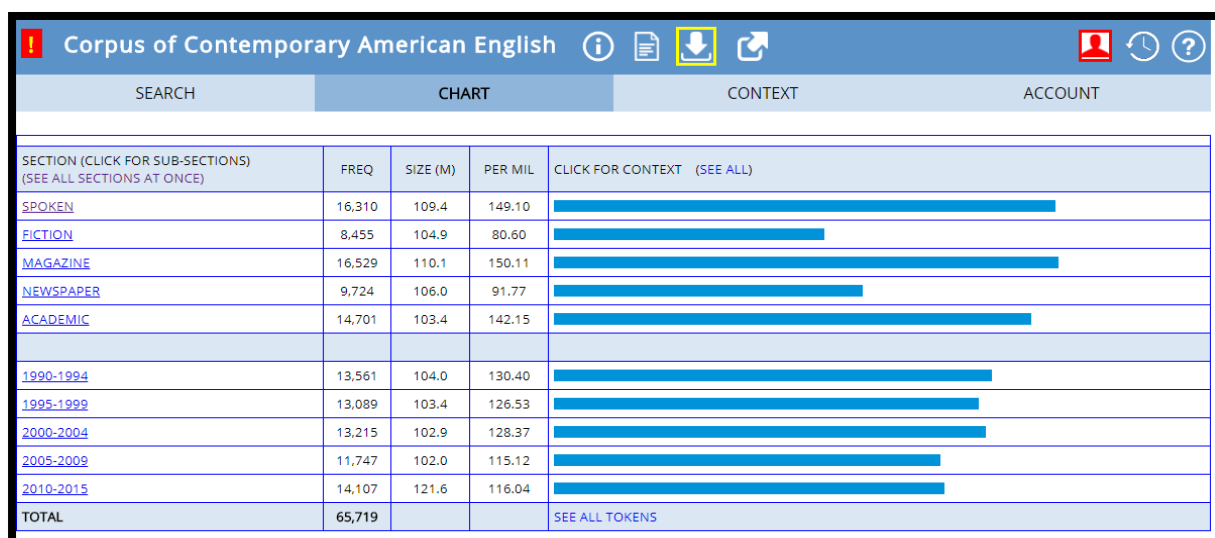


Figure 2: Results from tab 'Chart'

3- Collocates: This tab can be extremely rewarding as it provides the learner with words that tend to have a high statistical rate of co-occurrence. Learners will see what words occur near other words, which provides great insight into meaning and usage. For example, nouns after thick or look into, verbs before money, or any word near crack, believe, loud, or quickly. By utilizing the function of collocates effectively, learners will use collocational data to better understand English structure. For instance, let us use the same word ‘cause’ to look for nouns that mostly collocate with it. This is possible by typing the following in the search bar:

- In word/phrase bar we write [cause] between brackets.
- In collocates bar, we provide the part of speech, which is: `_nn*`

There is also a POS button next to collocates search bar, which stands for parts of speech. Learners can select the preferred POS to collocate with the word chosen.

The results (Figure 3) show that the noun that most often collocates with ‘cause’ is death. Other noun collocates are problems, damage, pain, cancer, disease, etc. Clicking on each result provides a list of instances of language use. By analyzing the results, we can safely deduce that the ‘cause’ has negative semantic prosody. In this way, the learner is engaged in the thinking processes underlying the creation of knowledge (Schank & Cleary, 1995).

Moreover, linguists have pointed out CPs potentials in the literature. Bloch (2009: 59) states:

“Concordancing technology can provide researchers, teachers, and students with a rich tapestry of examples of specific linguistic elements embedded in a variety of rhetorical contexts. Concordancing can also help the user to construct meanings and usage patterns based on sentences or pieces of discourse collected from published or transcribed texts.”

Bloch also asserts that CPs do not provide “correct” answers to queries about grammar and lexical questions but instead help create a learning environment where the user has to induce one appropriate answer from possibly many appropriate answers (ibid). This situation will yield and promote the learners’ autonomy.



Figure 3: Results of noun collocates of 'cause'

Furthermore, Yoon (2016: 223-224) conducted a study and concluded that using CPs may provide advanced L2 writers in academic settings with means and motivation to engage in robust meaning negotiations during their L2 written language production and therefore ultimately help them become more confident and autonomous as writers.

3. Design of the Study

3.1. Problem of the Study

Saudi school students are exposed to a large number of vocabulary the majority of which may not be of use if students were to be tested on any standardized language tests, e.g., Vocabulary Level Test (VLT). This issue may reflect poorly on other English Language skills, e.g., reading, writing, listening and speaking. In addition, learning vocabulary without taking into account the diverse germane aspects may prove to be a futile endeavor. Aspects like polysemy, collocation, and other lexical relations should be consulted for practical use of the language. Therefore, contact with CL is crucial for learners to be acquainted with such aspects so that autonomy is nurtured and encouraged in order to circumvent issues that may be raised of being restricted to the textbook.

3.2. Significance of the Study

This study is fundamental for everyone working in the field of EFL/ESL. It provides textbooks writers with an idea of the commonly used words in day-to-day life, which may effective to convince them to escape infusing teaching materials with words that have a lower statistical occurrence rate. The learners themselves will acquire such words at a later stage subconsciously through exposure, or learn them through conscious efforts autonomously.

This study is also fundamental for language teachers for it provides them with a new outlook as to how to effectively teach a language, and enable them to set priorities in the classroom, and supplement course materials with information that is relevant for students. Teachers may also be enticed to indoctrinate the use of corpora into their classrooms and be willing to revolutionize the way they teach languages. The magnitude of similar studies can highlight aspects of language teaching that may otherwise go unnoticed to teachers who are not current with recent development in the field of language teaching.

3.3. Limitation of the Study

The study is conducted at Gharbi Al'Otaif, Al'Otaif, and Khabta elementary schools at Al A'ardha District in the most southern border of Saudi Arabia. Although this study may be a good representative of Al A'ardha District, findings cannot be extrapolated to other schools across the Kingdom. Further studies are crucial in different parts of the kingdom to reveal a more precise unanimity to overcome the limitation.

3.4. Methodology

3.4.1 Instruments

The instruments of this research include corpus-inspired materials to be taught to students similar to the illustrated example in (Figure 4). Pre-post-tests are used to gauge students' rate of progression, interviews are conducted with the informants to gather their opinions of the study, its methodology and design, and their attitudes toward the importance of vocabulary in language learning.

3.4.2 Informants

Since English in KSA is taught starting from the fourth grade, it would only be cogent to restrict the choice of informants to sixth graders. They are at a transitional stage and being informed of the implications of corpora may prove to be fruitful in years to come. In addition, their knowledge is inclusive of past years, which acts as a guiding post to better represent their respected educational stage. Moreover, due to the complexity of the research tool (COCA) in relation to the sensitive age of informants, sampling procedures were strict. Students who do not have access to COCA at their homes were excluded from the initial sampling process (40 students, 20 per group). The final number of informants plummeted to (26) students divided into two groups, control and treatment. As a result, external validity is jeopardized. However, efforts were made to overcome the Hawthorne effect by communicating clearly the purpose of the study to both groups as well as being discreet in observations.

3.4.3. Design

The study adopts the experimental research design emphasizing random selection. The sample consists of two groups. The first one is the control group, who will be receiving instruction, the unmarked form of tuition in the standard learning environment. The other group, who is the treatment group, will be exposed to special conditions using the research tool. The purpose of the control group, axiomatically, is to provide a baseline for comparison. Progress is measured using pre-tests before the intervention, and post-tests after the treatment has completed. Such tests are corpus-derived in both pre-and-post successions, developed by the author and inspired by COCA. Results of the pre-post-tests are analyzed quantitatively using SPSS. Moreover, interviews will be conducted to obtain information about informants regarding their attitudes towards COCA, and how they feel about continuing to consult corpora in the future.

3.4.4. Materials

The materials that are going to be presented to students will take into account Hutchinson and Waters (1987) divisions of needs into two categories. The first category is called target needs (TNs). TNs refer to what learners need to do in the target situation. TNs are further divided into three groupings.

- Necessities represent the aspired, and final objectives learners should come to understand in order to function in a target situation efficiently. Necessities are determined by the instructor regarding vocabulary items that students may encounter at the next educational stage.
- Lacks refer to what the learners already know and do not know. This indicates a gap between the current performance and target performance (necessities). Lacks fit into present knowledge that must be gauged in order to bridge the gap and enable learners to reach the target destination during the learning process. This would have noteworthy implications on the selection of favorable materials.
- Wants refer to what learners wish to gain or learn. They merely mirror the personal aims and expectations of learners for desires do not occur independently of a person. This is achieved through negotiating the syllabus with informants. Individual wants must be addressed, and school facilities should accommodate them.

The second category is the learning needs. The inclusion of corpus-based teaching serves this purpose and fulfills these needs. This is because learners are going to be interacting with the corpus; thus, creating an environment in which self-discovery and self-direction are embraced.

3.4.5. An Illustration

This section includes a preview of how COCA is utilized to give students an insight into how to go about using it. “How many – How much” is one of the lessons that is going to be presented to students using COCA (Figure 4). The purpose of which is to find noun collocates that go with either “How many or how much.” Students are encouraged to interact with the results by forming sentences using “How many” with nouns that proceed. Examples in context are looked at for deeper understanding as to how to form sentences within spoken and written genres. Students are then exposed to the ‘extended context’ on a few selected examples - usually the first five. An exercise is provided to them using the ‘extended context’ by copying the sentence containing the designated word(s), deleting them and students try to fill in the blanks. Now, when it comes to differentiating the use of both, the ‘compare’ function is consulted which can provide learners with a different perspective on how to infer words’ usage. The ‘compare’ function has two search boxes (figure 5) that are used to compare the collocates of two words, and to see how they differ in meaning and practice.

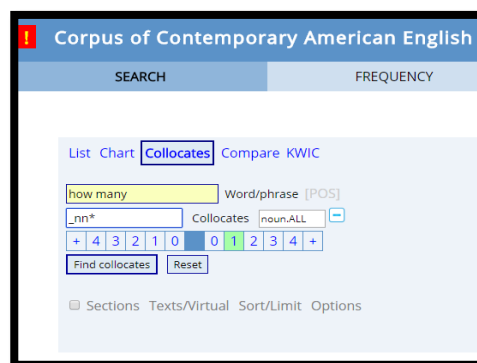


Figure4 : Noun collocates search

(Figure 6) shows noun collocates that occur with both “How many” and “How much.” The ‘compare’ function immerses students in an analytical atmosphere in which they have to analyze nouns associated with each and attempt to sort out any irregularities to reach a general rule regarding words’ usage. When first presented to them, students concluded that “How many” precedes nouns with –s plural ending. This was a step in the right direction. Sorting out the results by frequency instead of ratio, the word ‘people’ appeared as the most frequently used with “How many.” Since it is plural with no –s ending, students are yet again put to the test to dig even deeper and challenge their first deduction. Since ‘people’ is plural and does not take an –s, further analysis is required.

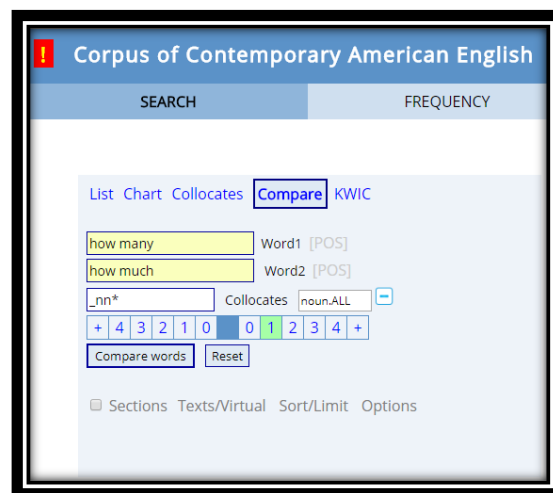


Figure5 : Utilizing 'compare' function

After several attempts, a suggestion that the ‘countability factor’ may be in play was proffered by one of the students. When asked, he replied that ‘people’ could be counted; thus, can be used with numbers, whereas nouns on the right column in (figure 6) cannot. The suggestion was immediately put to the test using as many noun collocates as possible so that the information is inculcated. Other uses for “How much,” however, was not provided, at least for the time being, in order to not distract students and maintain their focus on a one particular context at a time.

2/21/2018 jQuery UI Dialog - Default functionality

SEE CONTEXT: CLICK ON NUMBERS (WORD 1 OR 2)
 SORTED BY RATIO: CHANGE TO **FREQUENCY** [HELP...]

WORD 1 (W1): HOW MANY (0.68)					WORD 2 (W2): HOW MUCH (1.46)						
	WORD	W1	W2	W1/W2	SCORE		WORD	W2	W1	W2/W1	SCORE
1	HOURS	387	0	774.0	1,132.7	1	MONEY	2447	1	2,447.0	1,672.1
2	YEARS	738	1	738.0	1,080.0	2	FUN	281	0	562.0	384.0
3	TIMES	3473	6	578.8	847.1	3	DAMAGE	238	0	476.0	325.3
4	DAYS	324	1	324.0	474.1	4	WORK	194	0	388.0	265.1
5	MILES	133	0	266.0	389.3	5	TROUBLE	188	0	376.0	256.9
6	TROOPS	92	0	184.0	269.3	6	WATER	242	1	242.0	165.4
7	MINUTES	87	0	174.0	254.6	7	WEIGHT	220	1	220.0	150.3
8	VOTES	86	0	172.0	251.7	8	PRESSURE	110	0	220.0	150.3
9	CASES	83	0	166.0	242.9	9	EFFORT	104	0	208.0	142.1
10	MONTHS	82	0	164.0	240.0	10	CONTROL	100	0	200.0	136.7
11	JOBS	159	1	159.0	232.7	11	PAIN	98	0	196.0	133.9
12	NIGHTS	76	0	152.0	222.4	12	INFLUENCE	80	0	160.0	109.3
13	SHOTS	75	0	150.0	219.5	13	DIFFERENCE	79	0	158.0	108.0
14	CALORIES	145	1	145.0	212.2	14	SUPPORT	74	0	148.0	101.1
15	GAMES	69	0	138.0	201.9	15	ENERGY	145	1	145.0	99.1
16	POINTS	64	0	128.0	187.3	16	INFORMATION	142	1	142.0	97.0
17	WAYS	64	0	128.0	187.3	17	CASH	67	0	134.0	91.6
18	PLAYERS	63	0	126.0	184.4	18	PROGRESS	63	0	126.0	86.1
19	STEPS	61	0	122.0	178.5	19	HELP	61	0	122.0	83.4
20	PIECES	60	0	120.0	175.6	20	FUEL	60	0	120.0	82.0
21	LIVES	119	1	119.0	174.1	21	TIME	1540	13	118.5	80.9
22	CARS	53	0	106.0	155.1	22	FAT	56	0	112.0	76.5

Figure6 : 'Compare' function results

4. Results and Discussion

4.1. Results

During the study period, students in the control group were engaged in an immersive learning environment fostered by COCA. The study carried on for a month with 4 sessions per week. Students underwent two orientation sessions in the first week to get accustomed with COCA. Whether the orientation sessions were expedient or not remains to be seen. The results of the study are collected in accordance with informants' performance on both pre-and-post tests. Table (1) shows the number of participants and scores in the pretest of both control and treatment (before intervention) groups. Findings presented below are statistical extracts out of SPSS.

			Pretest	Posttest	Gain
Group	Control	1	8.00	10.00	2.00
		2	18.00	13.00	-5.00-
		3	6.00	7.00	1.00
		4	5.00	8.00	3.00
		5	10.00	12.00	2.00
		6	9.00	8.00	-1.00-
		7	13.00	9.00	-4.00-
		8	11.00	10.00	-1.00-
		9	5.00	4.00	-1.00-
		10	6.00	13.00	7.00
		11	8.00	7.00	-1.00-
		12	8.00	7.00	-1.00-
		13	2.00	2.00	.00
	Total	N	13	13	13
Treatment	1		3.00	2.00	-1.00-
	2		4.00	12.00	8.00
	3		11.00	14.00	3.00
	4		7.00	9.00	2.00
	5		5.00	11.00	6.00
	6		6.00	13.00	7.00
	7		15.00	15.00	.00
	8		2.00	12.00	10.00
	9		7.00	15.00	8.00
	10		16.00	18.00	2.00
	11		11.00	9.00	-2.00-
	12		10.00	14.00	4.00

	13		13.00	15.00	2.00
	Total	N	13	13	13
Total	N		26	26	26

Table 1: Pretest and Posttest Scores (control & treatment).

Table (1) provides case summaries of the scores in both pre- and post-tests of the two groups as well as the differences in scores (the test was out of 20). Table (2) displays the mean interval between groups in pretest and it is inconsequential.

Group	Mean	N	Std. Deviation
Control	8.3846	13	4.07305
Treatment	8.4615	13	4.55733
Total	8.4231	26	4.23484

Table 2: Mean Ratio in Pretest

The Treatment group performed slightly better than the control group in the pretest by a ratio of 0.0769. The difference ratio is not that momentous and may produce valid results later on, especially, if the improvement rate of the treatment group after post-test proves to be sizeable in relation to the results of the pretest. Another significant aspect was tested to ensure the reliability of pretest results. A univariate test for between-subjects effects in pretest (Table 3) clearly indicates that the Sig. value is $0.964 > 0.5$, so there is no statistically significant difference between groups in pretest.

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	.038 ^a	1	.038	.002	.964
Intercept	1844.654	1	1844.654	98.753	.000
Group	.038	1	.038	.002	.964
Error	448.308	24	18.679		
Total	2293.000	26			
Corrected Total	448.346	25			

Table (3): Univariate Test of Between-Subjects Effects in Pretest

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	92.346a	1	92.346	6.980	.014
Intercept	2783.115	1	2783.115	210.352	.000
Group	92.346	1	92.346	6.980	.014
Error	317.538	24	13.231		
Total	3193.000	26			
Corrected Total	409.885	25			

Table (4): Univariate Test of Between-Subjects Effects in Posttest.

A univariate test similar to the one conducted on pretest was also used on posttest results to gain perspective and ensure homogeneity of means. As evident from Table (4) there is a considerable improvement in posttest for the treatment group. This is demonstrated by the p value in the univariate test ($.014 < 0.05$) indicating a significant result that negates the null hypothesis. The difference in means between the two groups in posttest is revealed on Table (5) report, thus, substantiating the effectiveness of the intervention between tests. Table (6) also presents similar perception corroborating the same stance. There is a statistically significant difference between groups in posttest as determined by ANCOVA ($F = 11.070, p = .003$). Discussion of the findings will be carried out on the next section.

Group		Pretest	Posttest
Control	Mean	8.3846	8.4615
	N	13	13
	Std. Deviation	4.07305	3.25616
Treatment	Mean	8.4615	12.2308
	N	13	13
	Std. Deviation	4.55733	3.98233
Total	Mean	8.4231	10.3462
	N	26	26
	Std. Deviation	4.23484	4.04912

Table (5): Mean Difference between pre-post-tests

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	222.233 ^a	2	111.117	13.619	.000	.542
Intercept	171.732	1	171.732	21.049	.000	.478
Pretest	129.887	1	129.887	15.920	.001	.409
Group	90.321	1	90.321	11.070	.003	.325
Error	187.651	23	8.159			
Total	3193.000	26				
Corrected Total	409.885	25				

Table (6): ANCOVA Posttest & Pretest by Group

As has already been mentioned, and despite random sampling, the treatment group performed slightly better than control in pretest. However, the mean difference was not significant. Figure (7) sheds light on this issue by providing an illustrated representation of means differences in both pre- and post-tests.

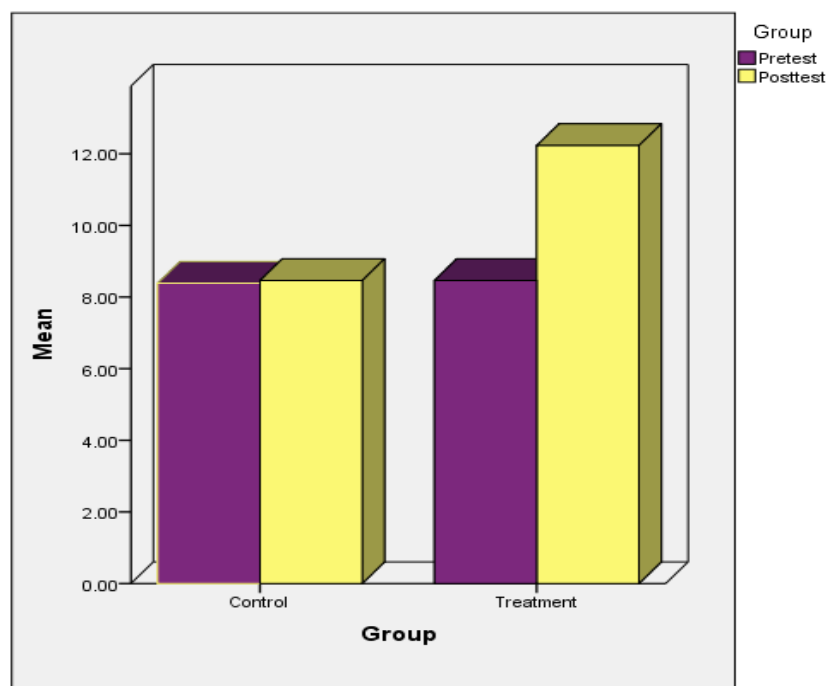


Figure 7: Means Differences of both Groups (Pre & Post).

For additional comprehensive measures, and for corroboratory purposes, a 'Gain' variable was created in SPSS to calculate the difference in 'gain' among participants. The 'Gain' variable deducts pretest scores from posttest scores (posttest scores – pretest scores) to provide yet another impermeable data that reflects the effectiveness of COCA as a tool in language teaching. Figure (8) validates the means difference between groups in posttest, and establishes the differences in 'Gain' variable across groups in posttest scores indicated by the blue bars below.

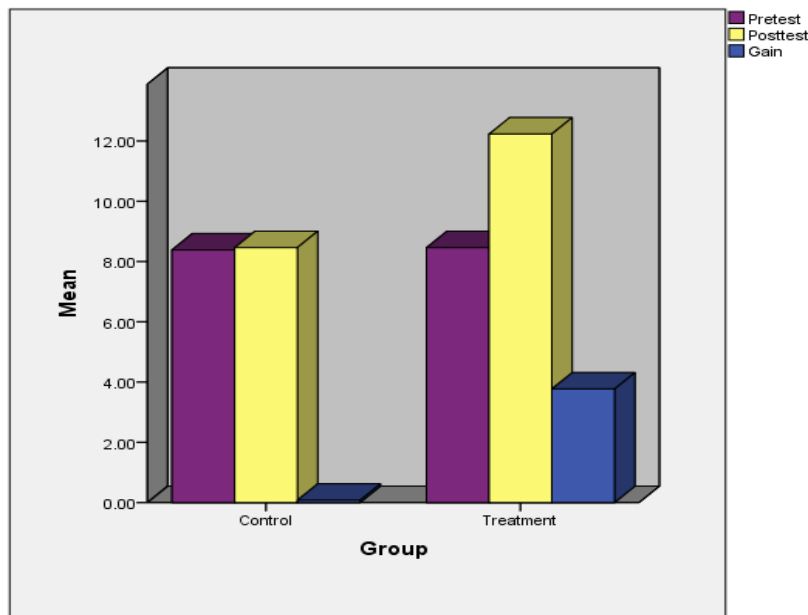


Figure 8: Gain Variable across Groups

4.2. Discussion of Results

For the research’s hypothesis to be validated, an examination of whether the means of both groups in posttest do differ is necessary. Moreover, tests of between-subjects effects are crucial to find if there are significant differences between subjects. In SPSS interpretation, sig. value greater than .05 mean acceptance of the null hypothesis. However, sig. value less than .05 mean rejection of the null hypothesis. Table (6) clearly indicates that the difference between the two groups is quite significant ($p < .05$).

Evidently, p value ($.014 < .05$) in Table (4) rejects the null hypothesis, which means that the research hypothesis is accepted. Therefore, the treatment or intervention has de facto had significant effect on students’ performance in posttest. Moreover, Levene’s test of equality (Table 7) consolidates this stance whereby $p = .465 > .05$ (not a significant result) which tests the null hypothesis that the error variance of the dependent variable is equal across groups.

F	df1	df2	Sig.
.552	1	24	.465

Table 7: Dependent Variable: Posttest

Figure (7) reveals the means difference (0.0769) in pretest and describes the acquired ratio (as does Table 5) of each group in both tests (8.3846 vs. 8.4615) for control in pre- and post- tests respectively, as opposed to 8.4615 vs. 12.2308 for treatment); hence the negligible means difference in the pretest. This is indicative of the fact that students in the treatment benefited from the exposure to authentic materials using COCA and their rate of improvement substantiates the optimistic potentials of COCA as an educational tool.

Figure (8) validates the means difference between groups in posttest, and establishes the differences in the overall ‘Gain’ variable across both groups. The difference rate manifested by the ‘Gain’ variable is prodigious. Treatment group not only surpassed control in test scores, they also improved individually. Exposure to authentic texts and navigating the corpus seemed a major role to the manifestation of the achieved results.

Therefore, the results of the quantitative analysis provide considerable evidence that using corpora does have its merits on Saudi elementary students at Gharbi Al'Otaif, Al'Otaif, and Khabta schools, thus answering the first of the research questions. Students valued COCA as a tool and as a diverting mechanism for promoting vocabulary acquisition. Students also felt comfortable interacting with COCA and were able to polish their inductive skills and foster autonomy. The results also attested to the effectiveness of COCA in promoting vocabulary acquisition and retention using representative texts spoken by the native speakers of the language. The result is increasing awareness to the prominence of the naturally collected texts in language teaching.

4.3. Students Perception

An interview (Appendix B) was conducted to bid a qualitative perspective to the quantitative results. Responses to the interview's open-ended questions indicated that students had positive perception towards using COCA in classroom at the expense of conventional teaching methods. This view supports the above quantitative findings as COCA was found to be helpful and entertaining in language teaching. For instance, Abdulaziz's said:

"COCA was useful for learning new words and how to use them, although the examples for each is too abundant, which can be confusing sometimes."

Despite the benefits of COCA in vocabulary acquisition, Abdulaziz felt that it is too rich and extensive that he sometimes feels confused. This reflects the usefulness of COCA as a tool and provides insights to some of the issues encountered by students. This is by no means counterproductive, since Abdulaziz is one of the students who improved by a gain variable of (3). Moreover, Sari's views supported Abdulaziz's by showing that COCA was useful in promoting linguistic performance.

"COCA was useful for it provides many examples, and can be used by almost anyone. And may be better suited for older students."

In this regard, Sari viewed COCA as a tool for promoting autonomy. Although he expressed doubts about other uses different from those administered in class. This is rather foreseeable due to his age being a six grader. Sari's performance increased by a gain variable of (4). Saud, on the other hand, had a different perspective and perceived COCA beyond the scope of his classmates.

"I find COCA to be helpful in reading and writing. As I read, I get to see how to sequence words."

Saud's perception agrees with previous studies that ascertained the effectiveness of corpora in writing (Yoon, 2016). The statement here is that COCA is a 'versatile domain' in itself and can be used to target different aspects of language. Saud performance remained intact after treatment. Furthermore, Abdulelah shared his views of COCA and endorsed it for years to come. He stated:

"I find COCA relatively easy to navigate, and I am willing to use it more and more. Also, it is current since examples are from 2017."

Abdulelah appears overconfident since one of the pitfalls of using corpora is navigation. However, his endorsement of COCA is a reflection of how relatable and entertaining COCA is. Abdulelah's "boastful" attitude may be justifiable since he improved by a gain variable of (8). In fact, his self-confidence is a reflection of the fact that he was one of those who benefitted the most from COCA.

Therefore, it is fair to assume that the potentials and effectiveness of COCA are rather subjective. Each student can benefit from COCA in a way different from the other; COCA is a 'versatile domain'. However, vocabulary acquisition sits at the core of these benefits, and acts as a hub to other aspects of language. The positive feedback is implicative of the stance upon which this paper was inspired, and elementary students appeared to have a good hold on the prospects that COCA afford.

5. Conclusion & Recommendation

The paper identified a current problem that led to the disassociation in the minds of Saudi EFL learners between what is taught in relation to what is used in reality. The role of authentic texts in classrooms was viewed to be crucial. Such texts found in corpora have the potential to reinstate the lost association between how language is structured and presented in English curricula in Saudi Arabia and real-world usage. COCA was employed to provide a solution and present a valuable contemporary awareness for language learners on how words function in isolation and in context. The findings of this study suggest that the use of COCA in classrooms as an educational tool bears many benefits. Saudi elementary students perceived COCA as a viable tool to language learning, particularly vocabulary acquisition. The practicality of COCA was manifested by the degree of improvement students exhibited in pre- and post-tests scores. This study has also established the effectiveness of vocabulary teaching using corpora. Thus, exposure to authentic texts aided in boosting retention and recollection of language items. Moreover, the difference of means between the two groups in the posttest is suggestive of the fact that students will benefit from corpus-based teaching. This also lends additional support to carry out further research in CL.

The current study has therefore acknowledged and addressed a gap in previous research regarding the role of COCA in language teaching. However, despite the efficiency of COCA, results cannot be generalized to other contexts. Due to this limitation, research is encouraged on higher educational settings as well as research in elementary stages in different parts of the Kingdom. It is to be noted that some students may view COCA differently from their peers. Therefore, using COCA as an educational tool may reveal multitude of implications regarding its pedagogical implementation. The key factor is for language learners to be acquainted with authentic texts. These texts are affectionately known as “authentic texts”.

Corollary, teachers, researchers, and course designers will have the ability to integrate authentic materials from specific rhetorical contexts in ways that reflect new theoretical and instructive developments in language teaching e.g., teaching of grammar (Ellis, 2002). These changes may be scary for some instructors, but it is fundamental for creating a wealth of opportunities for learners to navigate the language in any place and at their own time, both within and outside the classroom. Consequently, coping with such trends in language teaching will revolutionize the way we teach.

(This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.)

References:

- Biber, D. Et al. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511804489>
- Bloch, J. (2009). The design of an online concordancing program for teaching about reporting verbs. *Language Learning & Technology*, 13(1).
- DeKeyser, R. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. *Focus on form in classroom second language acquisition*, 42-63.
- Ellis, R. (1984). *Classroom Second Language Development: A Study of Classroom Interaction and Language Acquisition*. Pergamon.
- Ellis, R. (2002). The place of grammar instruction in the second/foreign language curriculum. *New perspectives on grammar teaching in second language classrooms*, 17-34.
- Guariento, W., & Morley, J. (2001). Text and task authenticity in the EFL classroom. *ELT journal*, 55(4), 347-353. <https://doi.org/10.1093/elt/55.4.347>
- Hunston, S. (2002). Introduction to a corpus in use. In *Corpora in Applied Linguistics* (Cambridge Applied Linguistics, pp. 1-24). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524773.002>
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511733031>
- Johns, T. (1994). 19 From printout to handout: Grammar and vocabulary teaching in the context of Data-driven Learning. *Perspectives on pedagogical grammar*, 293. <https://doi.org/10.1017/CBO9781139524605.014>
- Kennedy, G. (2014). Between and through: The company they keep and the functions they serve. In *English corpus linguistics* (pp. 107-122). Routledge.

- Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford University Press.
- Leech, G. (1992). Corpora and theories of linguistic performance. *Directions in corpus linguistics*, 105-122. <https://doi.org/10.1515/9783110867275.105>
- Lewis, M. (1993). *The lexical approach: the state of ELT and a way forward*. Hove, England: Language Teaching Publications.
- Little, D., Devitt, S., & Singleton, D. M. (1988). *Authentic texts in foreign language teaching: theory and practice*. Authentik.
- Mindt, D., & Cornelsen, F. (2000). *An empirical grammar of the English verb system: to the memory of Franz Cornelsen*. Cornelsen.
- Nation, I. S. P. (1990). *Teaching and Learning Vocabulary*. New York: Newbury House.
- Shank, R.C., & Cleary, C. (1995). *Engines for education*. Hillsdale, NJ: Lawrence Erlbaum.
- Siemens, G. (2005). *Connectivism: A learning theory for the digital age*. *International Journal of Instructional Technology & Distance Learning*, 2(1). Retrieved Oct 22, 2017, from http://www.itdl.org/Journal/Jan_05/article01.htm
- Sinclair, J.M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Wilkins, D.A. (1972). *Linguistics in language teaching*. London: Edward Arnold.
- Wilkins, D. (1976). *Notional Syllabuses*. Oxford: Oxford University Press.
- Yoon, C. (2016). Concordancers and dictionaries as problem-solving tools for ESL academic writing. *Language Learning & Technology*, 20(1), 210.