The Effect of Corpus-Based Language Teaching on Iranian EFL Learners' Vocabulary Learning and Retention

Ladan Ashkan[1] & Seyyed Hassan Seyyedrezaei[2]

[1] Department of English Language Teaching, Gorgan Branch, Islamic Azad University, Gorgan, Iran

[2] Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran

Correspondence: Seyyed Hassan Seyyedrezaei, Department of English Language Teaching, Aliabad Katoul Branch, Islamic Azad University, Aliabad Katoul, Iran. E-mail: srezaei.sh@gmail.com

Abstract

The use of corpora in second/foreign language (ESL/EFL) classes has established to be a valuable tool in teaching grammar, vocabulary and natural language use. The corpus-based approach to language teaching and linguistics has gained its prominence since the mid-1980s. However, there has been little research on investigating the corpus-based tasks openly in the classroom. The current research attempts to examine the effect of corpus-based teaching on EFL learners' vocabulary learning and retention of Iranian EFL learners. Forty pre university Iranian female students at Saei high school in Gorgan, aged 18 participated in this study. The number of participants in each group was 20. After administering the pretest, students in the experimental group were taught using corpus-based approach while students in the control group were taught using traditional methods. After instruction, a posttest was administrated to both groups. After two weeks of the first posttest, the second posttest was administrated to both groups to see the effect of corpus-based teaching on vocabulary retention (immediate retention). The design of the study was quasi-experimental, as there was no random selection. T-tests were employed to analyze the collected data from the vocabulary tests including pretest and posttests. The results of the study indicated a significant difference between the experimental and control group in favor of corpus-based vocabulary teaching. The result also showed that corpus-based teaching has a significant effect on EFL students' vocabulary retention and the effect did not fade away over time. This study has some pedagogical implications which can bring fruitful results for language teachers and learners and material developers.

Keywords: corpora, EFL learner, corpus-based teaching, vocabulary teaching, retention

1. Introduction

It is believed that general language ability highly relies on the Competency in knowing and using words in true contexts (Carter & McCarthy, 1988). However, nearly all learners whether English as a second or foreign language (ESL / EFL) learners face some difficulties with vocabulary and it becomes one of the leading difficulties of language instruction (Cobb, 2003). On the other hand, it cannot be ignored that vocabulary is an indispensable element of a language and of critical importance to the EFL learners (Zhang & Liu, 2014). In fact, vocabulary teaching is a significant part of language teaching. Zhang and Liu created an organized introduction methods and techniques of corpus based foreign language teaching. They assert that corpus-based teaching compared with traditional English teaching is recognized as a helpful tool which brings about authentic language input. Although corpus-based teaching has recently found its place in the classroom, a small number of attempts have been made to employ corpus-based tasks openly in the classroom by both EFL teachers and learners because of the difficulty students have in understanding this tool (Thomas, 2002).

According to Qiao (1995), corpus-based tool may be practical for language teachers, researchers and learners both in ESL and EFL contexts. It would develop vocabulary learning and knowledge by being exposed continually to some target words in authentic collections of data like reading texts. In other words, corpus-based instrument provides second language (L2) learners with real-life examples and far more exposures to some unknown vocabulary items. Therefore, language learners are in search for answers and have to assign more time in learning vocabulary through contexts.

In this study, corpus-based language learning was utilized. It is another method for teaching composition and is extremely compelling. Since it gives the learners language use in connection, it is valuable for the obtaining of syntax and vocabulary since they help learners to hold lexicon linguistic utilization designs better. This study has an essentialness for the learners on the grounds that they get to be inspired, and they accept more obligations regarding the vocabulary items, turn out to be more free authors, and more certain about vocabulary learning. The corpus-based methodology advances disclosure learning. Following in this study corpora have been utilized, this study gives numerous chances to the learners and teachers to advance language direction process.

2. Review of the Related Literature

2.1 Second Language Vocabulary Acquisition

The main reason behind learning a second language in general and vocabulary in specific is to achieve the ultimate aim, which is to know and understand information similar to that of native speakers of a language (Gass & Selinker, 2001). As a result, that would lead to the need to know the vocabulary size of native speakers. Nation & Waring (1997, as cited in Schmitt, 2000, p. 8), in their literature review of vocabulary size studies, inferred that a native speaker's vocabulary size is around 20,000 word families, and it is anticipated that a native speaker will add around 1,000 word families every year to his or her vocabulary size. A person will continue to learn new vocabulary throughout his or her lifetime.

Moreover, Nation (2006) asserted that L2 learners need to know around 98% of the written or spoken words in discourse in order to understand it very well. In order to reach this percentage in written texts, learners need to know around 8,000 to 9,000 word families. Put another way, learners need to know around 5,000 to 7,000 in order to understand a spoken discourse. However, Nation and Waring (1997, as cited in Schmitt, 2000) claimed that learners can cope with small vocabulary size of 2,000 to 3,000, but if they want to function in English without any unknown vocabulary, the vocabulary sizes which were stated above are necessary.

2.2 Learning Vocabulary in SLA Context

It is no doubt that learning vocabulary is an essential part for language mastery (Schmitt, 2008). Adding to a rich vocabulary is an unavoidable need for both L1 and L2 learners however because of incremental nature of word learning, it is an on-going test. In this manner, so far there has not been a strategy that best improves vocabulary learning (Schmitt, 2008; Gu, 2003).There are different ideas about the best way to learn vocabulary. Nation (2001) believed that form, collocation and word classes should be taught/learned incidentally but aspects of meaning, register and other constraints are better learned through direct explicit instruction. Schmitt (2008), however, puts more emphasis on intentional learning.

2.3 History of Corpus Linguistics

Corpus linguistics is a procedure inside of the field of linguistics that has been growing quickly since 1964 when the initially mechanized corpus, The Brown Corpus1, was finished. Corpus etymologists are for the most part intrigued by graphic or useful elucidations of language (Meyer, 2002), and study etymological marvels through the observational examination of substantial mechanized databases of language called corpora (corpus, sing.). A corpus is "a large and principled collection of natural texts" (Biber, Conrad, & Reppen, 1998, p. 4), which is compiled so that it is representative of the language in general, a language, or other subset of the language. Conrad (2005) and Tribble and Jones (1990) stated that corpora may contain language in light of composed writings, deciphered discourse, or both. These writings are put away electronically, and afterward dissected utilizing PC programming projects called concordance generators, concordancers, or, nonexclusively, concordancing programming.

Gathering a lot of content keeping in mind the end goal to dissect linguistics wonders was not another idea when corpus linguistics landed as a technique. As Meyer (2009) focused out in a late article, early word references depended on a substantial collection of distributed works and a huge number of reference slips of actually happening language.

2.4 Corpora in Linguistic Research and Language Teaching

Since their beginning, computerized corpora have been mainly used for research or "for finding out about language and texts" (Leech, 1997, p. 2). Today, nearly every sub discipline within linguistics uses corpora, to a greater or lesser extent, to inform their studies.

In spite of the fact that corpora have been utilized by etymologists for exploration purposes for more than forty years, analysts who are additionally language teachers are starting to have more prominent enthusiasm for abusing corpora for the instructing of second and outside languages. As indicated by Leech (1997), corpora can

have an immediate or aberrant impact on the language classroom. In a roundabout way, corpora are affecting the language classroom in light of the fact that they are being utilized by materials engineers to make enhanced reference materials (e.g., word references, punctuations, and thesauri) and course books. Moreover, corpora are being abusing by language teachers to advise syllabus and course plan (Flowerdew, 1993), and to make tests (Coniam, 1994; Shillaw, 1994). In addition, corpora have been utilized to make both general scholastic (Coxhead, 2000, 2002), and discipline particular (Wang, Liang, & Ge, 2008) wordlists. Wordlists like Coxhead's (2000) Academic World List (AWL) contain the most as often as possible happening headwords of a talk; on account of the AWL, the words are those which happen most habitually when all is said in done scholarly talk, paying little mind to teach. Coxhead's rundown depends on three standards: showing the most pertinent, valuable, and continuous lexical things to students first. The rundown has added to the prioritization of vocabulary for the EAP educational modules. In any case, while helpful for organizing vocabulary direction, wordlists should be taught utilizing a principled way to deal with showing vocabulary joined by fitting classroom methods keeping in mind the end goal to guarantee that students procure and can accurately and inventively utilize these words in their own particular discourse and composing. Corpus-based techniques and exercises can offer assistance. This conveys us to the examination of how corpora are directly affecting the language classroom.

2.5 Language Corpus in Vocabulary Teaching

A useful tool in teaching vocabulary is analysis of corpus information. It produces some valuable information for both students and teachers about how language is used in real-life situations. A corpus is a collection of authentic texts (written or spoken transcripts) that are stored in an electronic form (Partridge, 2006, p. 103). Its size can range from a few sentences to millions of words. Linguistic information is typically presented in the form of concordances (Tribble & Jones, 1997). A concordance is a list of all the occurrences of a particular word or phrase in a corpus, presented within the context (usually a few words to the left and right of this word). Concordances are obtained using the software called a concordancer.

Tim Johns was one of the first teachers who used a concordancer, and he was the author of the data-driven learning (Johns, 1991). DDL is an approach to language learning based on the assumption that the use of authentic language together with a concordancer will enable the learners to observe the language as it is used in real-life situations. What is more, in DDL the learning process is based on the learner's discovery of rules and patterns of language use

2.6 Vocabulary Retention

It is referred to that and the educated material is held in the memory, the learners can get advantage from it when the season of reviewing it comes. This is the thing that we call retention and recovery. Souleyman (2009) noticed that retention is a component of memory that can be characterized as including more perplexing capacities as retaining or learning, retention, review, and acknowledgment. He includes that there are processes precede retention which is seeing, admission, and capacity in the transient memory and later in the long term memory.

Vocabulary retention is an essential factor in learning English as a foreign language. Mohammed (2009, p. 16) defines vocabulary retention as "the ability to keep the acquired vocabulary and retrieve it after a period of time to use it in different language contexts". Zhang (2002) stated that one of the biggest challenges for EFL learners is how they can effectively remember, retain, and retrieve the newly learned English vocabulary. There are two kinds of vocabulary retention: immediate retention and delayed retention. Souleyman (2009) defined immediate retention as the level of retention of the newly comprehended piece of information as measured by a test after the experimental treatment immediately. It can also be referred to as medium term retention. On the other hand, he defines delayed retention as the level of retention of the target piece of information newly acquired through the experimental treatment, as measured by a test on that new information. In this particular case, the delayed test was given to the learner's month or more lately. Delayed retention can thus be referred to as long-term retention of the items.

3. Methodology

3.1 Participants

The participants of the study were 40 female pre university students at age 18. They were divided into two groups of 20. They were students at Saei high school in Gorgan, Iran. The study lasted for 20 sessions (2 sessions a week). Each class had the same teacher. The teacher and the researcher were the same.

3.2 Instruments

The instruments in this study were one pretest and two posttests. A 40 multiple-choice item test of vocabulary was developed based on the English book at high school (pre university level) by the researcher as the assessment tool. The vocabulary items were largely chosen from the ones which were taught during the course.

The criterion for selecting the words was their frequency. It should be mentioned that the posttests were the same as pretest.

At first, the constructed test was subjected to a pilot study, i.e., trying out the newly written test before final administration; items analysis (to check item facility, item difficulty and choice distribution) was run on this test. Through item analysis, poor items were modified. Validity is generally considered the most important issue in psychological and educational testing because it concerns the meaning placed on test results and it refers to the degree in which the test or other measuring device is truly measuring what was intended to measure (Cooper, 1998).

To construct the pretest, the researcher developed a 40 vocabulary items based on the text book (pre university book) objectives. The form of the items was Multiple Choice (MC) and piloted with 30 students. Of course, the validity and reliability indices were also estimated. The KR-21 Reliability index for pretest and posttests of vocabulary were 0.86. The validity of the test was reassured by consulting three high school language teachers regarding face and content validity. Their comments were used to revise the test.

4. Results

As it was mentioned before, the study was done at school and since the researcher had only access to two classes which were pre-determined by school authorities, therefore one of them was selected as the control and the other one as the experimental group of the study. To check the reliability of the instrument used in this study, the test was piloted on a population of students similar to the original one.

Table 1. Descriptive statistics of the piloting of the test of vocabulary

Mean	Variance	Std. Deviation	N of Items
21.35	4.108	2.027	40

According to the above table, the mean and standard deviation turned out to be 21.35 and 2.027 respectively. Afterwards, to check the normality assumption of the distributed scores in each group (experimental/control), a one-sample Kolmogorov-Smirnov test was run (Table 2).

Table 2. One-sample Kolmogorov-Smirnov test (for experimental/control groups)

	Test score (experimental)	Test score(control)
N	20	20
Normal Parameters Mean	33.70	26.95
Std. Deviation	2.25	1.63
Kolmogorov-Smirnov Z	.172	.162
Asymp. Sig. (2-tailed)	.123	.177

As represented in Table 2, it is clear that the experimental group (Z= .172, p= .123) and control group (Z= .162, p= .177) were normally distributed.

4.1 Investigation of the First Hypothesis

The first hypothesis is that corpus-based vocabulary teaching has no significant effects on Iranian EFL intermediate learners' vocabulary learning. To investigate whether there is any statistically significant difference between the posttest scores of learners receiving corpus-based vocabulary teaching compared to those in control group, an independent sample t-test was calculated (Table 3).

Table 3. Descriptive statistics for each group's performance on posttest

	Group	N	Mean	Std. Deviation	Std. Error Mean
Posttest1	Experimental	20	33.70	2.250	.503
	Control	20	26.95	1.637	.366

Table 4. An independent sample t-test of the posttest scores of experimental and control group

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Posttest1	Equal variances assumed	5.033	.031	10.84	38	.000	6.750	.622	5.490	8.009
	Equal variances not assumed			10.84	37.71	.000	6.750	.622	5.486	8.013

According the results, there is a significant difference in the posttest scores of corpus-based vocabulary teaching group (M=33.70, SD=2.250) and control group (M=26.95, SD=1.637); t (38) =10.84, *p* = .000. These results suggest that teachers' use of Corpus-based vocabulary teaching has statistically significant effect on Iranian EFL intermediate learners' vocabulary learning.

4.2 Investigation of the Second Hypothesis

The second null hypothesis of the present study is that corpus-based vocabulary teaching has no significant effects on Iranian EFL intermediate learners' vocabulary retention. In order to investigate this null hypothesis from different aspects, an Independent Sample T-test was run.

To examine whether there is any statistically significant difference between the second posttest scores of experimental and control groups, an independent sample t-test was run.

Table 5. Descriptive statistics for each group's performance on posttest 2

	Group	N	Mean	Std. Deviation	Std. Error Mean
Posttest 2	Experimental	20	31.75	2.173	.486
	Control	20	24.25	24.25	.280

Table 6. Independent sample t-test of the second posttest scores of experimental and control group independent sample t-test of the second posttest scores of experimental and control group

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	T	df	Sig.(2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Posttest 2	Equal variances assumed	6.255	.017	13.374	38	.000	7.500	.561	6.365	8.635
	Equal variances not assumed			13.374	30.349	.000	7.500	.561	6.355	8.645

As it is shown in tables 5 and 6, there is a significant difference in the second posttest scores of experimental group (M=31.75) and control group (M=24.25); t (38) =13.374, p = .000. These results suggest that there is a significant difference between the vocabulary retention of Iranian EFL learners who receive corpus-based teaching compared to those receiving traditional instruction (control group). Therefore, it can be concluded that learners receiving corpus-based teaching performed better in the second posttest than those belonged to control group.

5. Discussion and Conclusion

After analyzing the data which were obtained from experimental and control groups the following results were revealed. First of all, a pretest that consisted of forty multiple choice items was administered in both experimental and control groups to see if there is any significant difference between their knowledge of vocabulary in the beginning of the research. Then the results which were obtained from posttests were compared. This time the data analysis showed a significant difference in the mean scores of the two groups as a result of running corpus-based teaching in experimental group. Therefore, the first null hypothesis can be rejected and it can be proven that corpus-based teaching has statistically significant effect on vocabulary learning. Finally, the scores from posttest2

(that administrated 2 weeks after posttest1) were analyzed to see the effect of corpus-based teaching on retention of vocabularies. After analyzing the data, the data analysis showed a significant difference in the mean scores as a result of the running corpus-based teaching in experimental group. Therefore, the second null hypothesis, which indicated the effect of corpus-based teaching on vocabulary retention, could be rejected as well.

As the aforementioned data analysis demonstrated, the first null hypothesis was rejected because there was statistically significant difference between the experimental and control group as a result of administering corpus-based teaching during the term in experimental group. The second null hypothesis was rejected as well; because after two weeks of the first posttest, the second posttest (which was the same as the posttest1) was administered to both groups for retention of the vocabulary again there was statistically significant difference between the experimental and control group. That was due to using corpus-based teaching in experimental group too. Therefore, it can be said that corpus-based teaching is not as an instrument of power, but as a democratic instrument of learning; learners for real world communication.

References

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511804489

Carter, R., & McCarthy, M. (1988). *Vocabulary and language teaching*. Longman: New York.

Cobb, T. (2003). Do corpus-based electronic dictionaries replace concordancers? In B. Morrison, G. Green,& G. Motteram (Eds.), *Directions in call: Experience, experiments, evaluation* (pp. 179-206). Hong Kong: Polytechnic University.

Coniam, D. (1994). Using corpus word frequency data in the automatic generation of English language cloze tests. In L. Flowerdew & A. K. K. Tong (Eds.), *Entering text* (pp. 29-44). Hong Kong: Language Centre. http://dx.doi.org/10.17507/jltr.0602.10

Conrad, S. (2005). Corpus linguistics and L2 teaching. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 393-409). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213-238. http://dx.doi.org/10.2307/3587951

Coxhead, A. (2002). The academic word list: A corpus-based word list for academic purposes. In B. Ketterman& G. Marks (Eds.), *Teaching and language corpora* (pp. 252-267). Cambridge: Cambridge University Press. http://dx.doi.org/10.1111/1467-8489.00016

Flowerdew, J. (1993). Concordancing as a tool in course design. *System, 21*(2), 231-244. http://dx.doi.org/10.1016/0346-251X(93)90044-H

Gass, S., & Selinker, N. (2001). *Second language acquisition: An introductory course*. New Jersey: Lawrence Township.

Gu, P. Y. (2003). Vocabulary learning in a second language: Person, task, context and strategies. *The Electronic Journal of English as a Second Language*, *7*(2), 32-38.

Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data driven learning. In T. Johns & P. King (Eds.), *Classroom Concordancing* (pp. 27-45).University of Birmingham: Centre for English Language Studies.

Leech, G. (1997). Teaching and language corpora: A convergence. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and language corpora* (pp. 1-23). London: Longman.

Meyer, C. F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511606311

Meyer, C. F. (2009). In the profession: The "empirical tradition" in linguistics. *Journal of English Linguistics, 37*(2), 208-213. http://dx.doi.org/10.1177/0075424209334008

Mohammed, E. F. (2009). *The effectiveness of TPRS in vocabulary acquisition and retention of EFL prep.stage students and their attitude towards English language* (Unpublished M.A. thesis). Egypt: Mansoura University.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524759

Nation, I. S. P. (2006). Language education: Vocabulary. *Encyclopedia of Language and Linguistics, 6*(1),

494-499. http://dx.doi.org/10.1016/B0-08-044854-2/00678-7

Partridge, B. (2006). *Discourse analysis*. New York: Continuum.

Qiao, H. (1995). Navigating in corpus space with the mini-concordancer: An autonomous language learning perspective. *On-CALL, 9*(3), 23-37.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. http://dx.doi.org/10.1177/1362168808089921

Souleyman, H. M. (2009). *Implicit and explicit vocabulary acquisition with a computer-assisted hypertext reading task: Comprehension and retention* (Unpublished doctoral dissertation). University of Arizona, USA.

Thomas, J. (2002). *Concordancing for all*. Paper presented at the Fifth Teaching and Language Corpora Conference, Bertinoro, Italy.

Tribble, C., & Jones, G. (1990). *Concordances in the classroom: A resource book for teachers*. Essex, U.K.: Longman.

Wang, J., Liang, S. L., & Ge, G. C. (2008). Establishment of a medical academic word list. *English for Specific Purposes, 27*(4), 442-458. http://dx.doi.org/10.1016/j.esp.2008.05.003

Zhang, L. J. (2002). Metamorphological awareness and EFL students' memory, retention, and retrieval of English adjectival lexicons. *Perceptual and Motor Skills, 95*(3), 934-944. http://dx.doi.org/10.2466/pms.2002.95.3.934

Zhang, Y., & Liu, L. (2014). A corpus-aided approach in EFL instruction: A case study of Chinese EFL learners' use of the infinitive. *English Language Teaching, 7*(7), 152-158. http://dx.doi.org/10.5539/elt.v7n7p152

Copyrights