

## THE ROLE OF DATA-DRIVEN LEARNING APPROACH IN TEACHING ENGLISH VOCABULARY TO INTERMEDIATE LEARNERS

**Otakhon Rustam ugli Radjabkulov**

Uzbekistan State University of World languages

**Gulnoza Giyosiddinovna Radjabova**

Senior teacher, Uzbekistan State University of World languages

### ABSTRACT

Vocabulary development is an essential aspect of learning a second or foreign language. Therefore, the focus of this research will be on how to increase vocabulary acquisition through the use of data-driven learning approach. The goal of this study was to examine the impact of data-driven learning in acquiring English language vocabulary. As a result, 39 intermediate students participated in this research. Participants were first divided into two groups: Experimental and Control. The participants were taught 80 vocabularies during the course of the treatment's eight sessions. The words were taught through Data-Driven learning to the experimental group, whereas the words were taught conventionally to the control group. Two groups were evaluated (Post-test) after training, and the results were analyzed using JASP statistical software. The mean, variance, and standard deviation of the data were calculated. The results of the tests revealed that the experimental group had made significant development. The use of data-driven learning boosted word learning and retention in the experimental group, but not in the control group, and there was a significant difference between the two groups. Finally, it was determined that data-driven learning was a successful approach for all students in the experimental group.

**Keywords:** teaching vocabulary, DDL, Corpus linguistics, COCA, intermediate learners

### INTRODUCTION

The introduction of information technologies into our lives has greatly changed our way of life. At the same time, it has also had an impact on pedagogy and language teaching. In particular, in recent decades, the use of corpus linguistics in language teaching has been growing significantly. Corpus linguistics is the study of language based on corpora, the authentic huge database of language. Corpus data can be used to study and research different areas of language and approaches of corpus data research also

have some different characteristics. One of them is Data-Driven Learning( DDL) was originally coined by Tim Johns which is the use of corpus linguistics and its tools, techniques in language teaching pedagogy. When teaching English language, English language vocabulary is an essential aspect to teach, because it is a language rich in synonyms and collocations, and each word can be used differently in different contexts. As a result, all language instructors should think about how to teach vocabulary. Harmer (1990) stated that “If the language construction is regarded as the bone of the language, then vocabulary offers it vital organs, and flesh and blood.” In support of Harmer's opinion, it can be said that vocabulary is the main pillar of language. According to D. A.Wilkins, “without grammar, many things cannot be expressed by language, while without vocabulary, nothing could be expressed”. As a result, language teachers employ a variety of instructional resources to help students understand new terms. EFL and ESL teachers mainly use textbooks and traditional methods in teaching vocabulary in English. Although it is convenient and very effective for EFL teachers to teach students using textbooks in teaching vocabulary, they are not rich in real examples where these words are used to reveal the full meaning of certain words and examples are rarely taken from authentic sources. When it comes to use of corpus linguistics and corpus data in teaching vocabulary, Corpora or corpuses (singular corpus) are large electronic collections of spoken or written naturally occurring language (Ma & Mei, 2021; McEnery & Wilson, 2001; O'Keeffe et al., 2007; Reppen, 2010; Sinclair, 1991) Corpus linguistics is new discipline of linguistics. Many linguists suggest that Corpus linguistics is methodology which is considered as a way of analyzing of real language data for investigating a certain language for different purposes (Weisser 2016, Proctor 2012, Gries 2011, Swales 2016). On the other hand, According to, Togninini-Bonelli (2001), Corpus linguistics is a “pre-application” methodology which define its a collection of rules and knowledge before they are applied. At the same time she claims that Corpus linguistics has a theoretical status and a role which contribute other applications and some areas of linguistics. In the same vein, Mahlberg defines corpus linguistics as “an approach to the description of English with its own theoretical framework” Linguists have defined corpus linguistics in different ways, some consider corpus linguistics to be a method, others define it as a methodology, and some scholars consider corpus linguistics to be a separate theory or a branch of language. Corpus linguistics is effective not only in the development of linguistic fields, including lexicography, grammar, stylistics, translation studies and other fields, but also in teaching foreign languages. Corpus-based approach in teaching English is considered efficient



because Corpus entails natural occurring language data which gives opportunity to teachers and learners to analyze and learn lexemes in different contexts in detail.

According to Leen and Lee (2015), Due to corpora's data-driven nature, vocabulary learning and corpus analysis are intimately tied to one another.

Learning vocabulary is a crucial aspect of learning a foreign language. Some vocabulary is notoriously difficult to recall and use it appropriately. As a result of this, academics looked at using data –driven learning to improve student learning. Data driven learning approach is very practical in teaching foreign languages, including teaching English, and has several advantages. One of them is that it brings a real authentic language environment into the classroom, and the authentic data allows learners to explore for them. Unlikely, contexts in textbooks and teaching materials using traditional language teaching, students face real language contexts. Learners explore corpus data and learn from unfiltered English what vocabulary is used in what context and what grammar is used in what context. Gabrielatos (2005) describes it ‘condensed exposure’ which contribute to vocabulary expansion or heightened awareness of language patterns.

Individuals have the capacity to remember or recall things such as vocabulary, events, memories, and so on after a short or long period, according to Hornby's definition (2004). We have the ability to analyze and grasp all information through visual analyzing. In data-driven learning approach, learners research corpus data and learn vocabulary from authentic contexts.

In the DDL literature, learners are sometimes referred to as researchers (Bernardini 2001: 22), travelers (Johns 1997: 101), or detectives under the tagline "Every pupil a Sherlock Holmes" (Johns 1997: 101). Learners are encouraged to use a variety of activities to observe corpus data, create hypotheses, and formulate rules in order to gain understanding of language (inductive approach), or to verify the accuracy of rules from their grammar or textbook (deductive approach). As a result, they become more engaged, active, and ultimately autonomous in their learning. The student is "empowered" (Mair 2002), which boosts his or her confidence and sense of worth. Moreover, Corpus-based Data-Driven Learning (DDL), according to Zhu (2011), is a sophisticated computer-assisted teaching method that requires EFL students to examine real English samples provided by corpora and summarize the usage of the target vocabulary that EFL learners need to master. This approach could increase students' initiative and excitement for learning the language, as well as their capacity for critical thought and independent study (Zhu, 2011). Data driven learning ensures that students can analyze the

context themselves, memorize words and use these learned words correctly depending on the situation of use.

The research was based on the results of the following questions related to the basis of theoretical background and rationale of the investigation.

1. How effective is the usage of Data-driven learning in teaching vocabulary for intermediate students?

2. Is there any difference in vocabulary learning progress between teaching with and without corpus based data?

## METHODOLOGY

Qualitative and quantitative research methods are relevant to obtain results in this study. The data were obtained by running an independent t-test on two different groups as well as using the paired t-test for taking the pre and post overall results. Since the research aimed at investigating the effectiveness of data-driven learning, the research respondents were divided into two groups with one receiving the treatment. Treatment lasted eight sessions, each lasting 45 minutes. During the eight sessions of treatment, the participants were taught 80 vocabularies by the teacher. It signifies that learners studied 10 vocabularies in each session. Because "retention in short-term memory is not successful if the number of pieces of information surpasses seven," according to Grains (1986).

Before starting the research process, the corpus, corpus tools and their use were explained to the experiment group population, and a list of vocabulary that they need to learn in each lesson was presented. They learned words by analyzing corpus data through The Corpus of Contemporary American English (COCA <https://www.english-corpora.org/coca/>). This process was carried out under the guidance of a teacher in a room equipped with the Internet and computers.

On the other hand, new vocabulary was taught to a control group verbally by the teacher. Verbal explanation, for example, implies that she vocally conveyed the sole meanings, synonyms, and antonyms, and learners learned new vocabulary without having to look at their patterns and contexts. Finally, the researcher spent 45 minutes in each session teaching the students the new vocabulary. The pupils were assessed after 4 weeks and 8 sessions of instruction and vocabulary teaching; this exam was known as a post-test. There were a total of 7 tasks including 50 questions. And every single learner responded to the questions. The data was then supplied by the outcomes. (Pre-test consisted of 7 tasks based on 60 questions)

## Research subjects and venue

The pupils at the specialized secondary school number 269 in Sergeli district and at Secondary school number 72 in Yunusabad district in Tashkent were chosen as the population. Twenty students who were selected from secondary school 269 became members of the experimental group which consisted of 11 male students and 9 female students. Nineteen students, who were from Secondary school number 72, became the members of the control group which included 10 male and 9 female students. All the students were selected from 9 classes to 11 classes respectively.

### Research instruments

Internet access computers to register and use COCA corpus were chosen for vocabulary development and teaching to experimental group. For the control group, vocabulary was taught in the traditional way through the English Vocabulary in use text book and in a classroom equipped with a traditional blackboard.

### Limitation of the study

The current study was only limited only to students of a school in Tashkent. In order to come to a strong agreement on effectiveness of DDL(Data-driven learning) on teaching vocabulary and progress in acquisition of new words in a foreign language, the research should be replicated with the same tests and techniques. Furthermore, although teaching vocabulary through data driven learning was useful in this study, it is unlikely to be carried out in the absence of modern computers and the Internet.

## RESULTS AND DISCUSSION

The effectiveness of DDL was investigated according to the results of pre and post-tests for both control and experimental groups. The researcher ran Paired T-test to obtain results.

Table 1

### Descriptive

	N	Mean	SD	SE
Pre -test	20	68.900	7.440	1.664
Post-test	20	72.200	7.135	1.595

The tables above show that the difference in vocabulary progress of students in the control group is not that significant as the mean (the middle or average value of the entire set) ranges from 68,900 in the pre-test to 72,200 in the post-test.



**Paired Samples T-Test**

Measure 1	Measure 2	T	df	P	Cohen's d
Post-test	- Pre -test	4.931	19	< .001	1.103

The table above illustrates that p-value stands at .001. The smaller p-value favors the alternative hypothesis of the research which is the effectiveness of the Data driven learning and corpus data in students' vocabulary progress.

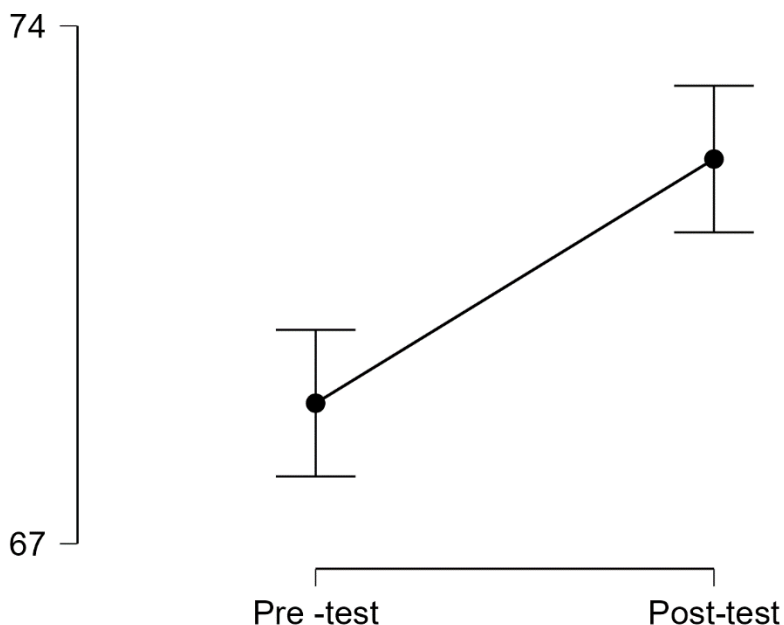
**Table 3. Descriptive Plots.****Pre -test - Post-test of Experimental Group**

Table 3 shows that students' average scores shifted from 68 to 72 in the post-test. The effectiveness of DDL is high as the table above shows that Cohen's d stands at 1.103 which is considered absolutely huge.

Note that

d=0.20 indicates a small effect

d=0.50 indicates a medium effect

d=0.80 indicates a large effect

**Table 4 . Descriptive Statistics**

	Groups	N	Mean	SD	SE
Pre-test result	CG	19	67.368	7.833	1.797
	EG	20	68.900	7.440	1.664

The table illustrates that the mean for both groups before treatment is not that significant. The dispersion (the spread out of data) about the mean value stands at 7.833 for the control group and at 7.440 for the experimental group. It means that there is a small variation since the data points are close to the mean.

**Table 5. Independent Samples T-Test for the post test of both groups**

	T	df	P	Cohen's d
Post test	1.723	37	0.093	0.552

The table illustrates that Cohen's d stands at 0.552 for both groups.

d=0.20 indicates a small effect

d=0.50 indicates a medium effect

d=0.80 indicates a large effect

The effectiveness of DDL on the vocabulary achievement of students is fairly high since the value stands at more than 0.552.

As can be seen, the experimental group made significant improvements since they were trained through Data-driven learning and acquired new words through the use of concordance, real contexts, and other corpus aids, indicating that this method of vocabulary acquisition was successful. Because they were trained verbally, the control group scored lower than the experimental group, indicating that the teacher had just taught new vocabulary by an explanation of meaning, synonyms, and antonyms without any authentic contexts. As a result, their advancement toward the experimental group was slower. According to Table 4, there was a quasi-significant difference in teaching vocabulary through corpus based data rather than without corpus materials.

## CONCLUSION

The primary goal of this research was to determine the impact of Corpus based data on vocabulary acquisition. For many instructors and researchers, the teaching of foreign language vocabulary is a hot topic. According to Chen(2004), “integration of corpora into vocabulary classrooms not only provides learners with faster searching tools and better quality of contexts that traditional dictionaries are not likely to achieve but also enhances their learning motivation” (Chen, 2004, p.5). The use of corpus data for pedagogical purposes does not have a long history, because computerized corpora themselves appeared in the middle of the twentieth century. The use of corpus linguistics in pedagogy is related to the way in which corpus data is used. Corpus based information can be used to create syllabus, dictionary, textbooks and other teaching materials, which some linguists call indirect approach. On the other hand, the use of corpus data directly in teaching an aspect of the language is considered Data Driven Learning. In the data-driven learning approach, the learner explores the language more independently through corpus data. This approach can be used in teaching grammar or teaching vocabulary and teaching other aspects of language. It is considered as a beneficial approach. The results of the above research show that data driven learning has several advantages in teaching and learning vocabulary. Through concordance list, a language learner can see a certain word again and again and analyze its use in different contexts, which ensures that the learner remembers that word and can use it correctly in different real situations. At the beginning, it should be recognized that using data-driven learning for teaching purposes can be difficult. Because Data Driven Learning is directly related to the corpus and its tools, language learners may have difficulties in using the corpus effectively. Before conducting this research, I investigated several studies on this topic. I found that one of the shortcomings of some of these studies is that when teachers use data-driven learning in language teaching, they allow learners complete learner autonomy, and the lack of guidance on the proper use of the corpus tools reduced their effectiveness. In order to inform the students of the experiment group about computer literacy and the use of the corpus, I explained to them how to use the corpus a few days before starting the research.

Some English words have denotational and connotational meanings. Such words can be taught in the traditional way, through textbook materials, but the learner faces difficulties in using the learned words in real life in different contexts in different meanings. However, when learning vocabulary through DDL, it allows students to analyze the denotative and connotative meanings of the words together. According to Cobb, Presenting





vocabulary in a set of contexts derived from corpora and/or in a concordance format offers contexts that are rich, varied, and abundant but also brief, incomplete, and devoid of a continuous storyline. Despite these drawbacks, these contexts may present some opportunities for contextual word learning that are absent from other more traditional text types.(Cobb,1999). The pupils were trained and examined in order to verify this subject. This subject was proven by their grades. In general, the students who were instructed to learn through data-driven learning made more progress than the students who were instructed verbally after 8 sessions and 4 weeks, according to the findings. As a consequence, data –driven learning is preferable to verbal instruction in vocabulary teaching since this study was beneficial to both the researcher and the students.

This study demonstrates the use of DDL in teaching vocabulary, and the findings support the use of DDL in both teaching and learning. If the corpus is chosen and used correctly, it can help pupils develop a more in-depth and helpful in this approach. In the course of this research, COCA (Corpus of Contemporary American English) corpus was used because this corpus is easy to use and partially free if you are fully registered. The research demonstrates that using data-driven learning in teaching and learning has favorable outcomes, since corpus based data can aid students' learning in a variety of ways. As we can see, learning vocabulary benefitted the experimental group more than the control group. It has been determined that learning using corpus resources is more successful than learning without them. After four weeks, the benefits of employing corpus resources were clear. There are a number of advantages to this research. When students acquire vocabularies with only translations, it is possible that they may forget such terms in the future. However, when they learn words with authentic contexts and usage, the pupils' learning improves. Also, ability of remembering or recalling those words will improve.

Another benefit of utilizing data-driven learning in vocabulary instruction is that in data-driven learning, learners are more centered than teachers. However, the teacher does not give complete learner autonomy and helps them learn by analyzing words correctly. Although this approach to teaching vocabulary is a rather difficult process, The advantages mentioned above and the results of the research show that DDL can be widely used in teaching English vocabulary, taking into account the level of students and the conditions of the educational institution.

## REFERENCES

1. Bernardini, S. (2001) “‘Spoilt for choice’: a learner explores general language corpora,” in G.Aston (ed.) *Learning with Corpora*, Houston: Athelstan, pp. 220-49
2. Chen, Y. H. (2004). The use of corpora in the vocabulary classroom. *The Internet TESL Journal*, 10(9), 1-5
3. Cheng, W. (2007) “Concgramming: a corpus-driven approach to learning the phraseology of discipline-specific texts,” *CORELL: Computer Resources for Language Learning* 1: 22-35.
4. Cobb, T. (1999) Breadth and Depth of Lexical Acquisition with Hands-on Concordancing. *Computer Assisted Language Learning* 12 (4), 345-360. Available at [http://www.lextutor.ca/cv/breadth\\_depth\\_99.pdf](http://www.lextutor.ca/cv/breadth_depth_99.pdf) (16.08. 2016.)
5. Gabrielatos, C. (2005) “Corpora and language teaching: just a fling or wedding bells?” *TESLEJ* 8(4): 1-35. <http://www-writing.berkeley.edu/TESEJ/ej32/a1.html>
6. Gries, S. (2011). Methodological and interdisciplinary stance in corpus linguistics. “Perspectives on corpus linguistics”
7. Harmer J, 2001. *The Practice of English Language Teaching*, 3rd edn. London: Longman.
8. Hornby AS, 2004. *Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press.
9. Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of datadriven learning. *CALL Austria*, 10, 14-34
10. Johns, T. (1997) “Contexts: the background, development and trialling of a concordancebased CALL program,” in A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds) *Teaching and Language Corpora*, London & New York: Longman, pp. 100-15.
11. Ma, Q., & Mei, F. (2021). Review of corpus tools for vocabulary teaching and learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177-190
12. Mair, C. (2002) “Empowering non-native speakers: the hidden surplus value of corpora in continental English departments,” in B. Kettemann and G. Marko (eds) *Teaching and Learning by Doing Corpus Analysis*, Amsterdam & New York: Rodopi, pp. 119-30
13. M. Weisser. *Practical Corpus Linguistics: An Introduction to Corpus-based Language Analysis*. Malden, MA: Wiley-Blackwell, 2016. xviii + 287 pp
14. McEnery, A. M., & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh, UK: Edinburgh University Press.



15. O'Keeffe, A., McCarthy, M. J., & Carter, R. A. (2007). From corpus to classroom: language use and language teaching. Cambridge: Cambridge University Press.
16. Proctor, F. (2012). 'Functional and corpus approaches of English grammar'. USA: Nyx Academics LLC.
17. Reppen, R. (2010). Using corpora in the language classroom. New York: Cambridge language Press.
18. Swales, J. M. (2006). Corpus linguistics and English for academic purposes. "In Information technology in languages for specific purposes" 1(2 ) 19- 33. Springer, Boston, MA.
19. Sinclair, J. (1991). Corpus, concordance, collocation. Oxford: Oxford University Press.
20. Tognini-Bonelli, E. (2001). "Corpus linguistics at work" ( Vol.6). Amsterdam, Philadelphia. John Benjamin Publishing.
21. Wilkins, D. A. (1972). Linguistics in Language Teaching. Cambridge: MFT Press.
22. Zhu, Huimin. Data-driven learning: the new trend of English vocabulary teaching. CAFLE.(2011) No.1, p. 46-50.

