

THE EFFECT OF DATA DRIVEN LEARNING TO IMPROVE VOCABULARY IN THE SECOND YEAR OF UNIVERSITY STUDENTS

Asti Ramadhani Endah Lestari
asti.r.endah.lestari@gmail.com
Universitas Indraprasta PGRI Jakarta

ABSTRACT

The goal of the present study is to compare the effectiveness of using the Data-Driven Learning approach and the traditional method of teaching vocabulary like a consultation with a grammar book and a lecturer's presentation. Two intact classes (N = 37) and (N = 38) taking Vocabulary subjects comprised the experimental and control groups, respectively. The researcher administered Nation's standardized vocabulary size test as a pre-test to ensure that the participants were at the same level of vocabulary knowledge. During the semester, the experimental group was taught using COCA. The finding showed that the experimental group exhibited a significantly marked increase in vocabulary results after the experiment. In General, the study revealed that DDL has a beneficial side-effect in improving university students' vocabularies as it enables students to perform an active role in the learning process.

Keywords: Data-driven learning, vocabulary

INTRODUCTION

Despite the fact that vocabulary learning is important (Ali & Anwar, 2021; Yang et al., 2021), learning in Asia, especially Indonesia didn't pay much attention on learning vocabulary (Barabadi & Khajavi, 2017; Catalan, 2003; Ivone, 2005). The emergence and recognition of "bottom-up processing" skills focusing on learning lexicon in language learning instead of the "top-down processing" skills focusing on grammar learning has acknowledged the key role of vocabulary in developing a second language (Barabadi & Khajavi, 2017).

Vocabulary is an essential aspect of learning a foreign language. (Richards & Renandya, 2002) say that vocabulary skills are the foundation of the other four language skills; listening, speaking, reading, and writing. Students who have a good understanding of vocabulary will master English faster. Understanding vocabulary helps students understand readings or utterances conveyed by others. Without adequate vocabulary, a person will not be able to write or speak properly and effectively. Understanding foreign language vocabulary is not only simply knowing

the equivalent words in the mother tongue but also understanding a metalinguistic awareness of the vocabulary. The information about a vocabulary can be in the form of correct spelling, pronunciation, synonyms, antonyms, and examples of using the word in a sentence.

There are various media that can help vocabulary learning, one of them is corpus. The corpus is a collection of several texts as a source of research related to linguistics and literature. The corpus in the plural form is called the corpora. Sinclair in Wynne (2005) describes a corpus as a collection of texts in electronic form that are selectively selected to serve as data in linguistic research. McEnery & Hardie (2011) said that a corpus is a body of text that has been carefully selected so that it can be used as a source of research data about a language or a variation of language to the fullest. Cheng (2011) says that the corpus is a compilation of languages that are classified based on specific categories to be studied to achieve a correct understanding of the use of the language. The corpus displays data about the use of words in a language.

Svartvik in Facchinetti (2007) explains that corpus linguistics is the study of language use in observed reality by collecting data from language and writing, then analyzing using a computer (corpus software) and describing based on a particular point of view and purpose. In line with Facchinetti, Widyawati (2014) said that corpus is a representation of a language in its actual form. The data displayed in the corpus are selected from real-life communication. These data were obtained in the form of spoken and written texts. Spoken texts can be taken from movies, television shows, and even everyday conversations. While written sources were obtained from various sources such as newspapers, speeches, novels, textbooks, films, television shows, etc.

Kaur & Hegelheimer (2005) describe that corpus has three advantages; authenticity, learner autonomy, and data-driven learning. Corpus shows evidence of authentic language because the data contained in the corpus are taken from sources that are used in the context of actual language use. Corpus also supports learning autonomy systems. Corpus helps students examine the function or structure of language from the source directly. By using corpus, the teacher's role is changed into a learning facilitator and is no longer a source of learning. In addition, the corpus helps data-driven learning (data-driven learning). Students learn a language based on quantitative and empirical data obtained from the processing of corpus data. The corpus provides many benefits in various linguistic studies such as lexicography, semantics, discourse analysis, grammar, language variety, contrastive linguistics, etc. An example of corpus research related to discourse analysis in the context of learning English is that conducted by Tai (2016). Tzu Yu Tai uses the corpus as a basis for analyzing discourse markers in 2 English textbooks and an English test in a comprehensive assessment program for junior high school students in Taiwan. Xiao conducted another study focusing on the uses

of the corpus to analyze the acquisition of modal verbs in students in China. The results of the study show that students in China tend to overuse modal verbs. This is due to the influence of Chinese as the mother tongue.

Corpus also has great benefits for language learning and research. In the field of research, the corpus helps researchers to generalize research results because the corpus has a level of validity and reliability Gablasova, Brezina, & McEnery (2017) and Kennedy (2014). The corpus transforms linguistic data which is usually displayed in qualitative form into quantitative data. While in the field of learning the corpus provides linguistic evidence that can be used as examples of language variations for students. From the results of classroom experiments conducted by Jafarpour, Hashemian, & Alipour, (2013) regarding language learning that uses a corpus and those that do not use a corpus, it was found that language learning that uses a corpus provides more significant results in improving students' writing abilities than the classroom who do not use the corpus. The corpus helps students discover word collocations so they can use a more varied vocabulary in writing.

Kwary (2018) said that the corpus allows users to explore input in greater depth. The corpus helps students explore all information about the vocabulary being studied. The corpus also helps teachers and researchers create perspectives about a language in greater depth. This is reinforced by Cai & Zhang (2013) who say that the level of authenticity provided by a corpus can help students build real learning scenarios. and get access to the real language. The corpus helps students understand vocabulary through the context of actual word usage.

Based on the explanation of the theory above, it can be concluded that the corpus is a collection of texts that are used as a source of linguistic and literary research. Even though at the beginning of its appearance there were many oppositions, now the corpus has proven to be of great benefit to various kinds of linguistic studies. Corpus has three advantages, namely authenticity, learner autonomy, and data-driven learning. The corpus provides quantitative evidence of word usage in real-life contexts.

Data-driven learning is considered as one implementation of how to use corpora in language learning. Data-driven learning refers to a learning approach that allows students to explore the corpus data in order to understand the language. DDL helps the teachers to familiarize the language to the authentic use of the vocabulary. I in this research, the researcher used corpora in language teaching as an approach which is data-driven learning. Data-driven learning has 4 characteristics. The first characteristic in DDL is the teacher Presents word in context. Second, DDL promotes autonomous learning. DDL Focuses on real, exploratory tasks and activities rather than traditional drill exercises so it is considered as Learner-centred activities. Next, DDL makes corpora as portable teacher as it allows students to Use and exploit of tools rather than ready-made. Finally, learners can detect language patterns through multiple samples. DDL is a

part of implementations of how to use new technology to enhance language learning. Focus on exploitation of authentic materials even for tasks such as acquisition of grammatical structures and lexical items. Based on the explanation above, the researcher is interested in conducting the study that focused on the use of DDL in vocabulary learning. The objective of this research is to compare DDL approach and traditional methods of teaching vocabulary like consultation of course book or dictionary

METHODOLOGY

The research took place in English Department of Universitas Indraprasta PGRI Jakarta where the researcher teaches. The researcher used quasi experimental method. There were 2 intact classes from 3rd semesters. 1 became the control class and the other became the treatment class. Nation's standardized vocabulary size test was administered in order to assess the level of proficiency of the students for this study. This test includes 140 items and can determine the learners' vocabulary size in a range of 0–14,000. Each item in the test has a score value of 100. Having confirmed the fact that the participants possessed almost the same level of vocabulary knowledge, the researchers used a corpus-based approach in the experimental group. More specifically, COCA so that the participants could learn the right collocations, prepositions, and the differences between similar words.

As mentioned earlier, this research aimed to find out whether a corpus-based approach toward vocabulary is more successful than a traditional one in which students learn vocabulary by consulting to dictionary or textbook. For this reason, a vocabulary test, consisting of the key words that the learners in both experimental and control group had encountered in their textbook, was administered to both the experimental and the control group. The test consisted of 20 multiple-choice items and 20 matching items. The items incorporated knowledge of collocation as well as synonyms of those words that were taught during the 3-week treatment period. Indeed, the words were chosen from the course book (internal used only module) that the students were taught in during the semester. Attempts were made to make the content of the test as representative of the content of the textbook as possible. After the 3-week treatment, the test was given to the participants in both experimental and control group

Below are the steps of how the researcher implement DDL in the treatment class.

- For the sake of having a corpus-based approach, in the first session of the course, the researcher introduced COCA corpus to the students in the treatment group. Particularly, three important features of COCA corpus were emphasized; namely, Word List including synonym, collocations and KWICs (keywords in context). In the beginning of each class, the students in the experimental group were given the output of the first 10 hits of the

target words in the list display of the COCA. Attempt was made to include those concordance lines that were relevant in terms of part of speech, collocation, and clarity of the sentence in rendering the meaning of the target word.

- second, students are given target words. Having gone through these example sentences for each target item, the learners were asked to make guesses about the possible meanings and part of speech of each word.
- After that, students are allowed to explore COCA. due to its built-in thesaurus, COCA makes it possible to perform synonym searches easily.
- In order to make the learning process more effective, students were required to discuss each target word in small groups.
- As a result, the students themselves were asked to enter the target words into a query box at the top of the screen and insert an equal sign before the word

FINDINGS AND DISCUSSION

After 3 weeks treatment in teaching vocabulary by using DDL approach in treatment class and using textbook and dictionary in control class, the researcher conduct the t-test. In order to find out whether there is a significant difference or not in the vocabulary mastery of the control class and the experimental class, the results of the calculation using the t-test the results showed

Then determine the value of *ttabel* in the t distribution table with a significance level of 5%.

It turns out that the value of *thitung* is 2.85 and *ttabel* is 1.672 at the level $\alpha = 0.05$. So the value of *thitung* > *ttabel*, so H_0 is rejected means it is significant. So it can be concluded that. Thus it can be concluded that there is a significant difference between students' English vocabulary mastery taught by conventional media and the class taught by data-driven learning.

This study shows that DDL approach has a positive and significant influence on the English vocabulary mastery of second year of university students. This is evident from the results of data analysis calculations, that the results of the English vocabulary test for the treatment class taught obtained an average (mean) of 86.36. Whereas the control class which was taught using textbook obtained an average (mean) of 72.4. From these data it shows that the vocabulary scores of students who are not given treatment using DDL approach have not shown satisfactory results. The following is a comparison of students' vocabulary mastery scores using DDL approach and textbook and dictionary.

Value	Control Class	Treatment Class
Mean	72.4	86.36
Standard Deviation	18.02	19.80
Varians	324.73	392.17

Tabel 1. Comparison of the Mean Value, Standard Deviation and Control Class Variance

The table above shows a statistical summary of the mean, standard deviation and variance of the English vocabulary mastery test results for the control class and the treatment class. The values in the control class which were taught by textbook and dictionary were quite low compared to the values of the treatment class which were taught by DDL approach.

With the results of the analysis above, in order to find out whether there is a significant difference or not in the vocabulary mastery of the control class and the treatment class, the results of the calculation using the t-test show that the *t*count is 2.85. By referring to the distribution of *t*table, it can be seen that degrees of freedom ($60-2=58$) at a significant level of 0.05, the value of *t*table, is 1.67. The fact that the value of *t*table, (1.67) is smaller than the value of *t*count (2.85). It shows that the null hypothesis (H_0) is rejected and the alternative hypothesis (H_1) is accepted. Thus it can be concluded that there are significant differences between the mastery of English vocabulary students who are taught by textbook and dictionary with classes taught by DDL Approach.

The results of the study are confirmed by other studies which found that corpus-based data-driven approach to teaching and learning vocabulary is more effective than traditional methods like consultation of a course book or dictionary (Boulton, 2010; Eichstaedt et al., 2021; Jaidka et al., 2020). DDL approach allowed language learners to gain greater exposure to various senses of the target word, and hence helped the learners in the experimental group to outperform the learners in the control group. DDL provides authentic source for students to learn the vocabulary DDL approach created learner centered environment classroom as it allows students to formulate and answer their own questions, discover grammar rules, and solve language problems by exploring the corpus.

CONCLUSION

Based on the findings and discussion above, we can conclude that data-driven approach to teaching and learning vocabulary is more effective than traditional methods like consultation of a course book or dictionary. Furthermore,

the DDL approach is superior because of the merits of the approach and not because of differences in the quality of instruction.

REFERENCES

- Ali, B. J., & Anwar, G. (2021). Vocabulary Learning Strategies and Foreign Language Acquisition at Private Schools. *Ali, BJ, & Anwar, G.(2021). Vocabulary Learning Strategies and Foreign Language Acquisition at Private Schools. International Journal of English Literature and Social Sciences, 6(3), 163–173.*
- Barabadi, E., & Khajavi, Y. (2017). The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary. *Cogent Education, 4(1), 1283876.*
- Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning, 60(3), 534–572.*
- Catalan, R. M. J. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics, 13(1), 54–77.*
- Eichstaedt, J. C., Kern, M. L., Yaden, D. B., Schwartz, H. A., Giorgi, S., Park, G., Hagan, C. A., Tobolsky, V. A., Smith, L. K., & Buffone, A. (2021). Closed- and open-vocabulary approaches to text analysis: A review, quantitative comparison, and recommendations. *Psychological Methods, 26(4), 398.*
- Ivone, F. M. (2005). Teaching English as a foreign language in Indonesia: The urge to improve classroom vocabulary instruction. *TEFLIN Journal, 16(2), 195–208.*
- Jaidka, K., Giorgi, S., Schwartz, H. A., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2020). Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. *Proceedings of the National Academy of Sciences, 117(19), 10165–10171.*
- Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice.* Cambridge university press.
- Yang, X., Kuo, L.-J., Eslami, Z. R., & Moody, S. M. (2021). Theoretical trends of research on technology and L2 vocabulary learning: A systematic review. *Journal of Computers in Education, 8(4), 465–483.*