CORPUS-BASED RESEARCH IN VOCABULARY LEARNING

Rizka Safriyani

Universitas Islam Negeri Sunan Ampel, Jl. A. Yani 117 Surabaya, Indonesia

Article Info

Article History:
Received June 2020
Accepted August 2020
Published September 2020

Keywords:
research-based learning, corpus-based research, vocabulary

Abstract

In the university, corpus-based research is commonly done for writing a thesis. However, corpus-based research can also be introduced for the first year of EFL students to build their critical thinking and vocabulary mastery. Less research discusses the practice of corpus-based research for the first year EFL student. Therefore, it is essential to investigate the benefit and the challenges of corpus-based research in the Indonesian EFL Setting. This study aims to examine the benefits and the challenges of corpus-based research in the Indonesian EFL Setting. Students did corpus-based research in English for the Islamic Studies course. Students tried to structure an English glossary from online Islamic articles, Islamic journals, and Islamic blogs. Forty-four students were chosen as the subject of the research. The survey was done to the students to gather the data about the benefits and the challenges of corpus-based research. The results showed that corpus-based research benefits increase vocabulary, increase students' understanding of research, improve students' accuracy in writing, develop critical thinking, and develop collaboration. Students faced several challenges in implementing corpus-based research. The finding shows students have difficulties in understanding new vocabulary. Besides, they have problems classifying data into specific topics, allocating time, and writing their reports.
INTRODUCTION

Research-based learning is systematized learning that starts from facts, concepts, and principles to procedures, which are very close in nature to research activities (Haryati et al., 2020). By applying research-based learning, learners are more competent in both science and research and are adept at identifying problems and solving them correctly using scientific methods. Studies prove that the application of research-based learning has changed the paradigm of educators who have considered themselves as users of knowledge to become inventors of new knowledge. This shows that current educators are encouraged to discover new things to deliver to their learners. As a constructivist learning method, research-based learning is believed to be very valuable for developing learners' research and language competence (Sjøberg, 2010). Research-based learning is also believed to improve learners' vocabulary mastery because they learn to understand the meaning of words and use them in an acceptable context. Furthermore, research-based learning allows the learners to practice writing scientific papers before writing their thesis at the end of their college (Attard et al., 2010).

However, preliminary research showed that the English Language Education Department learners of UIN Sunan Ampel Surabaya lack research experience before their thesis writing. Several courses are offered to the learners in equipping them to conduct research and write the thesis; they are Qualitative Research in English Language Teaching, Quantitative Research in English Language Teaching, English for Islamic Studies, Research in Applied Linguistics, and Collaborative Classroom Action Research. The English for Islamic Studies course aims to improve the learners' English proficiency at the intermediate level and is offered to first-year learners in the second semester. This course uses Islamic texts and videos as the primary teaching materials and integrates corpus-based research as the teaching strategy to improve the learners' English proficiency. A corpus is a set of texts which is put together for some purposes. A corpus usually consists of written texts, transcriptions of spoken material, or both. Corpus-based research was used because learners must read many texts written in English to grab the corpora and enhance their vocabulary mastery (Krieger, 2003). It is believed that reading a lot of English texts is beneficial and can be done easily for first-year learners. The result of the corpus-based research usually produces a glossary or dictionary. This research would investigate the benefits and challenges of corpus-based research in vocabulary learning (Thompson & Sealey, 2007).

Researches have been conducted to investigate how Islamic themes texts and corpus-based research benefit learners in developing their vocabulary mastery (Atar & Erdem, 2019; Lin-
Fang, 2013). Munir and Hartono (2016) examined how Islamic themed articles in a magazine could enhance learners' reading skills and build their character values. The study reported that the use of Islamic-themed articles could improve learners' reading skills and contribute to the learners' character building positively. Furthermore, the research shows that the use of Islamic-themed articles can reduce student boredom in reading. (2020) investigated the effectiveness of the research-based learning model to improve the Educational Psychology course's learning outcome. Using the stages of orientation, exploration, concept formation, application, and closure, the results of this study show that there were significant differences in the learning outcomes before and after the treatment in using the research-based learning model (Jamalipour, 2015). Despite the success of implementing Islamic themed articles and corpus-based learning model in vocabulary learning, the benefits and challenges of corpus-based learning have yet to be clearly described in the literature (Purwanto, 2009). Therefore, this paper reported the benefits and challenges of corpus-based learning in the English for Islamic Studies course in the English Language Department of UIN Sunan Ampel Surabaya. This paper would portray the learners' beliefs about the benefits and the challenges of corpus-based research in vocabulary learning.

REVIEW OF LITERATURE

Constructivist Learning Theory

Constructivist learning is learning that emphasizes processes and freedom in exploring knowledge and efforts in constructing experiences. Constructivist learning provides learners an opportunity to express their ideas in their language and think about their experiences so that they become more creative and imaginative. Sjøberg (2010) states that constructivist learning theory could be applied in foreign language learning. The constructivist approach contains critical concepts in developmental psychology and deals with representing the world more accurately and working on logical operations of representations of the concepts of world reality (Atkinson, 1919; Vernon, 1976). Before acquiring any new things, the constructivist model considers learners as people who have initial skills. Students ought to be capable of thinking.

Furthermore, Pagán (2006) states that constructivist teaching methods are mainly learner-centered. Constructivist models emphasize fewer teachers' participation and, instead, the learners work more to solve the problems. Constructivist learning techniques allow the learners to practice and improve their language skills. In constructivist learning, teachers cannot merely rely on direct assessment, but they must also assess the learners with subjective and context-based techniques (Pannen et al., 1991). The most common of these are essays and open-question
exams. Usually, teachers apply all evaluation types to single students, but they may also adapt essays for small classes to be working. Therefore, constructivist learning could help learners recognize problems, formulate hypotheses, construct mental models, conduct research, test their hypotheses, and reach conclusions with some practical guidance from the teachers.

**Research-Based Learning**

Research-based learning is often associated with problem-based learning because both starts with problems to be resolved through scientific steps. Research-based learning can provide a variety of benefits in the context of the development of metacognition and achievement of competencies to be learned during the process of learning. Research-based learning will involve students in various scientific processes like collecting information, building hypotheses, collecting the data, analyzing the data, and drawing conclusions based on the analyzed data. Widayati et al. (2010) report on several research-based learning strategies developed at Griffith University:

1. **The use of the teachers' research results as a means of enriching teaching materials**
   Research-based learning enables the updating of teaching materials written by teachers. Suppose the teachers conduct and publish research continuously. In that case, the research results can be used as a real example of research as an authentic material to be used in the classroom. By doing so, the learners would understand the procedures and importance of conducting research quickly.

2. **Use of research results in the latest and discuss the history of the development of the research results**
   Research-based learning focused on library research can be optimized by collecting data obtained from the library to be discussed to support the appropriate subject matter. For example, in the Indonesian context, research on the English language curriculum in schools can be used as a new discussion material on the similarities and differences in the curriculum from time to time. Students can understand that the existing curriculum at the moment is the result of policies and practices that have been developed previously. All of this is a dynamic unity of the development of science that is very interesting to study.

3. **Enrich learning activities with contemporary research issues**
   Research-based learning allowing students to identify current issues around them related to learning. In the Indonesian context, for example, online learning is becoming increasingly popular and in demand. Students can be asked to discuss the application of the issue to
resolve real problems in the community. This strategy can be developed in several ways, for example:

a. Comparison reports on research results and reports news that happened in the community.
b. Comparison reports on the research methodology and related arguments with the research findings presented in the journal research.
c. Comparison reports on the development of appropriate knowledge with the subject matter.

4. Teaching material on research methodology in the process of learning

Research-based learning allowing students to learn about the stages of research systematically and by academic ethics. This strategy can be applied by doing the following steps:

a. Introduce the research methodology and provide a complete understanding of the research.
b. Design teaching materials by integrating research methodologies on the subject matter to apply them to solve real research problems.
c. Design teaching materials with various research methodologies that are related to several current research issues so that students can learn to analyze research data

5. Enriching the learning process with small-scale research activities

Research-based learning is done in groups by conducting small-scale research, such as students conducting simple surveys. In this learning process, groups of students conduct research. Thus, the research culture can be more developed because students must fulfill the values of cooperation, honesty, critical thinking, and meeting academic ethics together.

Healey and Jenkins (2006) state that there are four kinds of research-based learning activities. The first activity is called research-led teaching and learning. In this part, student learning materials are taken from the results of research conducted by others. Research-oriented teaching and learning are the second continuum that reflects the activities that develop students' knowledge and abilities to implement research methodologies and methods following their scientific disciplines. Research-based teaching and learning are the third activities that ensure that many students are doing research. It usually engages in the research process and inquiry. The last type is known as research-tutored, where students engage in research discussions. All types above show the element of research in all kinds. Corpus-based research can be done in various kinds of research-based learning activities; however, the researcher
applied Research-Based Teaching and Learning as the approach because students did library research to improve their vocabulary mastery.

**RESULTS AND DISCUSSIONS**

**The Benefits and the Challenges of a Corpus-Based Research**

Corpus linguistics is a field which concentrates on a collection of processes or method for learning words. Krieger (2003) formulates that a corpus consists of a collection of natural documents, collected from writing or registered speech transcriptions. The key goal of corpus linguistics is to discover patterns of accurate language usage by examining real usage (Wilkins, 2015). The collection of text or corpus is called as corpora. A concordance lets readers see texts in context. Some tools can be used to analyze corpora, for example, AntCont, Concordance, CILIC, and CorpKit. Different tools provide the creation of frequency information, such as the language frequency database which lists all words appearing in the corpus and defines how many times each unit happens in that corpus.

Corpus-based research can be done quantitatively and qualitatively. Thompson and Sealey (2007) have performed a corpus-based analysis to examine the children's literature vocabulary. They compared the children's literature corpus to a corpus of adult literature and newspaper articles to determine whether children's written language has different linguistic properties relative to the adult text. Moreover, their finding shows that the vocabulary in children's literature shares many of the language's characteristics in adult fiction but, to a lesser extent, the news text's vocabulary profiles. Moreover, corpus-based research can examine almost any language pattern. It can analyze the lexical, structural, discourse, phonological, and morphological aspects of a language. The discovery of male versus female use of tag queries can also be used as corpus-based research.

Willis (1998) states that corpus-based research could enhance students' ability to identify useful phrases and common collocations, identifying the structure and nature of both spoken and written discourse. Atar and Erdem (2019) also state that corpus-based research provides data for language studies. Corpus-based research includes holistic information about language structures because the data obtained from corpora is commonly analyzed by computer, and evidence is used concretely. Corpus-based research does not focus on single contexts. Multiple contexts were drawn to grasp the general understanding of the word's nature in the real-life data. However, Hunston (2002) discovered the pitfalls of Corpus-Based Research. Corpus-based research is weak in providing a detailed analysis of the user of the languages. Corpus-based research could not give an analysis of the context and structure of the talk among speakers.
Therefore, many people argue that corpus-based research did not portray the actual language used. it focuses on the form rather than the function of the language.

**Vocabulary Learning**

Vocabulary is a set of words that speakers of language use. Nation (2008) states that high-frequency words are words that are frequently used in language production. Those words are classified from the 2,000 most frequent word families—the words used frequently in a formal or informal situation. The words occur in the written and spoken text, such as newspapers, conversations, novels, and academic texts. Nation (2008) also suggests that these words should be considered as the target words in the vocabulary development program for the basic level.

Academic words are words that can be found in academic texts. Academic words are different from the list of 1,000 or 2,000 of high-frequency words. These words occur in the newspaper, children's books, very formal conversation, and academic writing. These words consist of 570-word families and are commonly known as the Academic Word List (AWL). Nation (2008) also adds that university students who use English for their academic tasks should learn Academic vocabulary. However, Nation (1990) also states that academic vocabulary is usually learned after the students mastered the high-frequency words.

Berne and Blachowicz (2008) state that vocabulary learning is an important part of English language learning. The learning of a new word in articles, books, or the internet sound very critical. It is also fundamental to language teaching and is of utmost importance to a language learner. More repetitions of vocabulary within a context could help a learner to learn and acquire vocabulary better. The use of interesting story, interesting articles, or news could help students to recognize the vocabulary. The incidental process of vocabulary learning would facilitate vocabulary mastery.

Chi and Lip (2009) found that the most widely used and effective vocabulary learning techniques are 1) repetitive spelling of the word in the mind, 2) analyzing the word by breaking down the fragments of sound, 3) recalling terms by doing a project, and 4) asking classmates for the meaning of the term. Another research by Subekti and Lawson (2007) mentions the use of a dictionary. Simple contextual use and regular pronunciation of the word and its meaning are commonly used among Indonesian postgraduate students. They best-used those strategies for learning new words. It can be inferred that memory plays an important role in determining the vocabulary mastery (Agustina et al., 2012). The vocabulary learning techniques that are commonly done refers to the repetition of the word. Therefore, a crucial factor in successful vocabulary learning was putting new words into practice that they have only learned. The
repetition could be done through corpus-based research because corpus-based research could expose learners with words from real data.

**English For Islamic Studies**

*English for Islamic Studies* is a course designed to prepare students with the Nation (1990) states that learning vocabulary in context is the primary way for native speakers to expand their vocabulary knowledge. When students read articles in journals, the internet, newspapers, or social media, they have familiarized themselves with the appropriate and acceptable sentence context. In other words, many students may read, increasingly honed the ability of students to use vocabulary in context (Gu, 2017).

Furthermore, based on students' responses, they want to do the corpus-based research in the future. This response shows how reading activities can bring the spirit of reading to find out new things. When students try to formulate a big topic that most often appears in discourses/texts analyzed, they must find their issues relevant to their findings. In formulating and compiling research, students have applied constructivist theories that construct knowledge through the ability to remember and re-express experiences. Based on the results of the questionnaire, there are four benefits obtained by corpus-based research. The first benefit is to develop collaboration. The second benefit is developing critical thinking, the third benefit is to improve students' accuracy in writing and the last benefit is to improve students' understanding of research. The description of each benefit is presented in figure 1 below.

![Figure 1. The benefits of corpus-based research](image)

Figure 1 shows that the most preferable benefit from the research process is that corpus-based research can increase vocabulary. Seventy-eight percent of the students state that corpus-based research can expand their vocabulary. Nine percent of the students say that corpus-based research can improve their understanding of research. Nine percent of the students say that corpus-based research can improve their understanding of research. Moreover, four percent of the students
believe corpus-based research can improve their writing accuracy, develop critical thinking, and build collaboration. However, students also face some challenges in applying corpus-based research. The description of the challenges can be seen in the following figure.

![Figure 2. The challenges of corpus-based research](image)

Figure 2 shows that corpus-based research's biggest challenges are the difficulties in understanding new words or sentences. 82.65 percent of the students state that they have difficulties understanding new words or sentences. There is 9 percent of the students say that classifying the data is very challenging. Furthermore, 4 percent of the student's states have limited time allocated, and they got difficulties in writing reports.

**Students' voice about the benefits of corpus-based research**

Figure 1 indicates that reading activities will help students gain new vocabulary and understand the context of using the vocabulary. Students state that corpus-based research could increase vocabulary. When they read a lot of text, students would get insight into how the vocabulary is used in a context. Corpus-based research has forced students to read a lot and gain a lot of new vocabulary. Since corpus-based research analyzes vocabulary use, the students get a sort of language exposure from a different type of text. The context of discourse is the internal aspects of the discourse and everything that externally covers a discourse. Hymes (1972) states that there are several elements of a discourse that is the speaker, topic, setting, channel, and communication context. Students will become more familiar with the Vocabulary and the context attached to the word. For example, in the word prophecy, this word is quite widely used in scientific articles, and Islamic-style newspapers are rarely found on social media. In further studies, students can explore the functions of the discourse context in research on discourse.
Furthermore, figure 1 indicates that students also feel that corpus-based research adds their knowledge about research. Creswell (2014) states that research is searching for a theory, theory testing, or problem-solving. Research is the application of a scientific approach to the recognition of a problem. In this case, students have been introduced to the science of problem-solving through scientific procedures. This activity would improve critical thinking because students need to think critically to finish their research. The research is aimed at seeking information so that students' direct benefits when doing research are obtaining new information.

Another benefit that can be interpreted in figure 1 is that this learning can increase writing accuracy and enhance collaboration. Larsen-freeman and Larsen-freeman (2006) defined writing accuracy as the proportion of error-free in lexical, morphological, and syntactic errors. This finding is interesting because corpus-based research could strengthen the students' understanding of word choices to improve writing accuracy. For example, students can infer that accuracy refers to how the language structure is used properly by learners through their use of vocabulary. If students read a lot and do corpus-based research, they would avoid spelling errors. The accuracy in writing includes the student's ability to identify spelling errors and punctuation. Repetition of article reading could build students awareness about the spelling of the vocabulary. This finding is in line with Willis (1998) that corpus-based analysis enables students to determine the structure and the nature of spoken and written discourse.

The other advantages of doing corpus-based research are related to collaboration. Laudel (2002) suggests that Research collaboration is characterized as a system of research activities by several actors linked functionally and coordinated to achieve a research objective corresponding to those actors' research objectives or interests. By doing a group research project, students worked in groups to collaborate to answer the problems since they have the same research objectives. The students who work in a group can achieve much more than individuals working. A broader range of skills can be applied to practical activities such as sharing and discussing ideas. Group work can play a pivotal role in deepening students' understanding of a particular subject area.

Challenges in Applying Corpus-Based Learning

There are three challenges in applying corpus-based learning. Figure 2 indicates that the first challenge is that students have difficulty in understanding new sentences or words. Students who never read out of their book would lack vocabulary. Through research-based learning, learning new words is a fun activity, and the student even can do with the people around. Most students just read English text if they have got an assignment. Therefore, they
frequently get difficulties in understanding new words due to the limited exposure of the language input.

The second challenge is difficulty in classifying data. Students have difficulty in classifying the data because they rarely read English Islamic text. The questionnaires show that students usually read a popular article in English. They read English Islamic Article only for the assignment. They also find the discourse on the non-fiction text. This finding is in line with the study of (Schmitt, 2008) who states that when reading fiction, readers feel more at home since fictional worlds are self-contained and complete, it is not important for the reader to test and verify facts with the real world as if reading non-fiction.

Successful comprehension of texts depends on the readers’ semantic and syntactic encoding abilities and their vocabulary knowledge in the language. If the students cannot comprehend the text, they will get difficulties in classifying the data. The next challenge is that students have limited time allocation and have difficulty writing their reports. Since this course is given for the third-semester students, they have limited experience writing a research report. The time allocated for the research project was limited due to the national holiday and the university agenda this year. If there is no national holiday and the university agenda, students may finish the project better. Students expected a longer time to complete their reports. They can just submit their writing draft since they need to revise the document in academic writing. Even though students have got the template of the research report, they need much practice. There are some errors in terms of formatting and word choices. Students have got difficulties in terms of linguistics difficulties, organization, and content. This finding is in line with Pratiwi (2016), which states that the writing difficulty related to the language use and vocabulary aspects was the most difficult one compared to the organization, mechanics aspects, and content.

Corpus-based research becomes one of the examples of vocabulary learning activity in English language teaching. According to Özönder (2016), recalling terms by doing a project is a common strategy in vocabulary learning. Corpus-based research has helped students look at specific words repeatedly, forcing them to recall the term in the discourse. Students would master the vocabulary and able to distinguish the use of the vocabulary in a different context. Therefore students must be given many reading activities. The reading process can be done through an extensive reading program or a corpus-based project to make it meaningful. This research provides exciting reading activities because it uses reading resources that can be determined by students. Students have the autonomy to decide the title of discourse.
CONCLUSION

This study aims to investigate the benefits and the challenges of corpus-based research in the Indonesian EFL Setting. The results showed that corpus-based research benefits are increasing vocabulary, increasing students' understanding about research, improving students' accuracy in writing, developing critical thinking, and developing collaboration. Students faced several challenges in implementing corpus-based research. The finding shows students have difficulties in understanding new vocabulary. Besides, they have problems classifying data into specific topics, allocating time, and writing their reports. The study involved only two classes in English for the Islamic Studies Course. It could be replicated with more participants to decide whether the present study's conclusions are similar. Next, research-based learning with different proficiency levels (elementary, pre-intermediate, and high) or different types of research would help gain more insights into the benefits of research-based learning.

REFERENCES


