INTRODUCING THE USE OF CORPUS IN ENGLISH LANGUAGE CLASS

Rika Mutiara

Esa Unggul University

Arjuna Utara 9th Street, Kebon Jeruk, Jakarta

rika.mutiara@esaunggul.ac.id

Abstrak

Perkembangan teknologi berdampak pada semua aspek kehidupan termasuk pendidikan bahasa. Dengan adanya perkembangan teknologi, korpus bahasa muncul. Korpus memanfaatkan teknologi komputer dan internet. Bentuk kegiatan pengabdian masyarakat ini adalah pelatihan pemanfaatan korpus untuk guru bahasa Inggris di SMA. Kegiatan ini dilakukan untuk memperkenalkan kumpulan teks yang dikenal dengan nama korpus. Seiring dengan perkembangan jaman, teknologi yang memadai digunakan untuk menyiapkan pembelajaran yang lebih baik. Dengan berbasiskan kumpulan teks yang sangat besar, korpus dapat memberikan informasi bahasa mulai dari tingkatan kata sampai dengan diskursus. Investigasi kolokasi dan baris konkordansi merupakan fungsi dasar korpus yang dipelajari dalam pelatihan ini. Dengan menguasai hal ini, guru bahasa Inggris akan mampu mengamati berbagai aspek dalam teks seperti tata bahasa, tindak tutur, dan manajemen informasi. Penguasaan aspek tersebut dapat membuat guru mengembangkan rencana pembelajaran dan evaluasi teks karya siswa dengan lebih baik. Dalam pelatihan ini, guru menyadari manfaat yang bisa diberikan oleh korpus. Selain itu, mereka juga berpartisipasi aktif dalam pelatihan.

Kata kunci : korpus; pembelajaran bahasa

Abstract

The use of technology is increasing in all aspects of life including language education. This condition makes the existence of corpus possible. By using a corpus, English language teachers can prepare the lessons better by providing learning materials that are suitable to the real life. Having seen the importance of a corpus in a language class, this community service aims at giving a training to the English teachers at senior high schools in Jakarta. It introduces a corpus to the English teachers. A corpus consists of texts in a large number. It can give several information from word to discourse levels. There are many features of a corpus. The most frequently used features are investigations of collocations and concordance lines. Being able to use these features makes English teachers will be able to observe several aspects in texts such as grammar, expressions, and information management. It makes them able to develop lesson plans and evaluate students' works more accurately. The teachers realized the benefits that a corpus offers in English language teaching. They also actively participated in the training.

Keywords: corpus; English language teaching

Introduction

English Teacher Working Grop or *Musyawarah Guru Mata Pelajaran* (MGMP) *Bahasa Inggris* is an association where English language teachers can discuss several aspects related to English language teaching such as developing syllabus and instruments of English test. MGMP works on education level. MGMP that collaborates in this community service program is MGMP for senior high school teachers in the region of East Jakarta 2. It sometimes conducts trainings and workshops for the teachers. It expects the teachers to be able to work independently to make innovation. Being able to innovate creatively makes the teachers develop learning materials and activities that are interesting. Mostly, the teachers use textbooks. Using a textbook as the only reference makes the students easily feels bored. Thus, they do not want to involve in the lessons. The teachers are required to be creative to make every lesson different. Moreover, in the textbooks, most materials were developed for the purposes of teaching English. Those texts are different from the ones used for communicating in the real life. Providing real life texts can be one of the ways to develop an interesting lesson and it can be done by studying the language patterns in the text Fligelstone (1993).

MGMP supports the English teachers to develop their competencies that help them do

their duties by conducting trainings. The topics of the trainings mostly are related to the components of instruction. The trainings focus on how to develop syllabus and lesson plans, doing assessment, and using ICT in English class. A training that deals with developing teachers' knowledge in linguistics was rarely conducted. Indeed, having good knowledge in linguistics influences several aspects in teachers' duties. With this knowledge, they will be able to develop syllabus and lesson plans, select learning materials, and conduct a lesson. Having considered all aspects, MGMP is willing to conduct a training to update teachers' knowledge on linguistics. MGMP expects the training to provide up to date knowledge and useful skills for the English teachers. Based on the discussion with MGMP team, it was found that the teachers only know doing text investigation manually. They do not know how to explore a collection of texts with the use of technology.

The development of technology and the massive use of the internet gives benefits on the investigation of texts. By applying a technology, a collection of texts in a large number can be stored electronically. The collection may consist of millions of words. Such collection is called as a corpus. The plural form of a corpus is called corpora. Most corpora are displayed in the websites. Usually, some universities are the owners of the corpora. There are some corpora to be used as references in English language teaching. In the reference corpus the speakers are not English language learners. However, some of them are not free. The users must register and pay before using the corpora. Some of the corpora that can be accessed freely are COCA (Corpus of Contemporary American English) and BNC (British National Corpus). The texts available there can be adapted or adopted to be used to develop learning materials in English class.

Having considered the condition and teachers' need, the topic of the training is how to get benefits from corpus for English language class particularly at senior high school level. If the teachers master the basic principles in corpus linguistics, they will be able to analyze the text independently. The skill of analyzing texts is useful for developing syllabus and lesson plans and designing instruments of tests.

To do the analysis based on the texts on BNC and COCA, a computer with a good internet connection must be available. The English teachers have this requirement. Therefore, they can participate in the training. Having this training does not only increase teachers' knowledge on linguistics but also give more insight on how to integrate technology in English language teaching.

The topic on corpus linguistics and its application in English language teaching is rarely discussed in teacher's education in Indonesia. Curriculum in the English language study program hardly covers this topic. Therefore, most English teachers in Indonesia are not familiar with corpus linguistics. They neither get knowledge about the use of corpus in English language teaching nor skills on applying the corpus. Therefore, this topic is necessary to be introduced to the English teachers.

Principles on corpus linguistics that is urgent to be presented is counting frequency of words. In a corpus study, searching for a particular feature of a language by considering its frequency is a must. Then, the use of the feature in a sentence is observed. To do this, the context of the feature must be scrutinized (Thornbury, 2010). To observe the context, concordance lines are examined in the level of a sentence to a discourse. Whether the feature occurs in a spoken or written language also must be noticed. There are some characteristics that influence the nature of spoken and written language.

Next, the principle of investigating collocation must be delivered. Learning a word cannot be separated from its collocation. Collocation is not about the co-occurrences of words but also the meaning. Collocation has a pivotal role in language acquisition. The language learners must master the collocations to achieve high proficiency level (O'Keeffe, McCarthy, & Carter, 2009). Collocation shows some words tend to occur together and their frequencies are significant. Collocations may appear in all text or only in particular text types (Hunston, 2011). Here, the English teachers need to consider what text types are suitable to their students.

Methodology

There are two stages in conducting the community service program. The first stage is having a discussion with the committee of MGMP to know the condition of the teachers and what they need. Then, the purpose and the topic of the events were set up. Based on the information of MGMP committee, trainings on learning methods, assessment, and technology in English language teaching were conducted several times. They hardly have a discussion on linguistic topics. A topic on linguistic issue particularly corpus linguistics is necessary to be given. The teachers, moreover, do not have any information about corpus in English language teaching. Mostly, they learnt linguistic only when they were university students. Some studied linguistics in their graduate study program. It can be said they learn any topics in linguistics only in formal education.

Based on the purpose of the training, the materials were prepared. The materials were obtained by studying literature review and doing corpus exploration. The materials cover two issues namely exploring collocation and concordance lines. How to introduce the corpus to the teachers in a simple way also became the concern of material preparation. The MGMP committee asked the teachers to prepare a laptop that has a good internet connection. The MGMP committee socialized the training to their members. The members are teachers from state and private schools. The participants must fill the registration link before joining the training.

The next stage is conducting the webinar with Google meet. The link of the meeting was given to those who registered. The duration is 150 minutes. In the beginning, the MGMP committee led the opening ceremony. Next, the community service team and the participants introduced themselves and greet each other. It was done to create a warm atmosphere. Having such atmosphere increases participants' willingness to participate actively. Therefore, they actively asked and gave responses in the training. In the session of presentation, the participants were given opportunities to try the corpus. It had a practicum session where the participants explore the corpus using their laptops. They might ask questions directly if they had difficulties. The first part in practicum session is registering in BNC or COCA.

Results and Discussion

The training was held on June 27th 2022 at 7.30-10.00. There are 20 English teachers participate in the trainings. They are the members of MGMP in East Jakarta region 1. The rundown is as follows: opening, introducing the speaker, presenting the materials, and closing.

The presentation begins by encouraging the participants to involve actively in the training by asking questions, sharing their experiences related to the discussion, and practicing investigation of features of a corpus. Then, the next session is asking the participants to tell what they know about corpus. As what has been predicted, most participants do not know it. Only one participant mentioned that a corpus is about linguistics. It shows that participants' knowledge about a corpus is very limited.

To introduce the concept of the use of corpus and collocation, the participants were asked whether they have ever got confused in choosing what words to use. An illustration of *conduct research* or *do research* was given to the participants. They were asked to share their cases. They also raised a topic of the interference of the first language that happens in this case. The participants got the concept of collocation. An emphasis was given to the point that in some cases, the characteristic of collocation is arbitrary. If it is arbitrary, why two words collocate cannot be explained. This discussion led to the issue of naturalness in using a language. The English language learners can achieve high proficiency level if they can use English naturally.

The next session is introducing the corpus. The participants did not face serious problems here. They were asked to register on COCA and BNC. For BNC, there are several websites that can be used. However, the trainer recommended registration on website managed by Lancaster University because it gives access to a larger context of the texts. There are some questions regarding the registration on COCA and BNC. They could handle the registration better. One participant asked a question which corpus is better. The response to this question is there is no better corpus. All depends on the purposes on using the corpus. COCA does not provide wider context. Due to this condition,

COCA is not used for investigation in the level of discourse. BNC managed by Lancaster University provides a larger context.

Picture 1
Introducing COCA and BNC

The next activity is a demonstration of searching words to find the frequencies and collocations in the corpus. The participants did not find difficulties in this session. Some participants tried to have their own word search. In this session, the participants were told that they might make limitations of the text types when they used the query box, for instance, they focus only on spoken text only. In COCA, they may limit the searching only on fiction, academic, or news language. One of the participants asked whether the idioms and their meanings are available in the corpus. The trainer, then, explained that there is no information about the meanings of the idioms. The users are only able to check whether the idioms are used by English native speakers or not. A corpus is different from a dictionary that can give words meanings.

Exploring the data in the corpus makes the participants realize that teaching variations of vocabulary and expressions in English class is important. There are several ways available in English but sometimes the teachers only focus on one way. Let's take an example. Most teachers use only a phrase of *have lunch*. The phrase *eat lunch* also exists in English. Therefore, the teachers need to introduce *eat lunch* after teaching the students the phrase *have lunch*. The participants were asked to check whether both phrases are available in COCA and BNC. Both phrases are used by English native speakers in the United States of America and the United Kingdom.

In observing linguistic item that is related to grammar, the participants did not face difficulties. The trainer gave a demonstration of searching the word *moreover* and observed its position in a clause. Based on this example, the participants could give a related example. They explained about the case of the word *thus*. The participants realize that variation can be created through the positions of conjunction or connector in a clause. Grammar is not only about the rules but it might be about the position of lexical items in a clause.

The next demonstration is an expression. The trainer demonstrated how to investigate the expression of *how are you* in discourse level. This expression was chosen because it is commonly taught in English class in Indonesia context. The participants shared their thoughts here. They mentioned they used limited expressions to carry a particular language function. There are many variations other than *how are you*. Furthermore, they mentioned they taught limited responses to reply a thank. They just taught you are welcome. Other responses such as *my pleasure*, *any time*, *no problem* are rarely used. It shows that the participants are able to make connection of the principle of corpus to their language teaching practices.

At the end of presentation, there is a question concerns on whether the principles in corpus linguistics can be applied to teach English by applying integrated skills. The trainer explained that it is possible to use a corpus specifically in a particular type of text, for instance, the word search only focuses on spoken text. However, it does not mean that language skills must be taught separately if the teachers use a corpus. The findings based on corpus investigation does not influence the methods of teaching. The other question deals with how the use of corpus can support the program of *Merdeka Belajar*. The trainer responded that corpus helps the English teachers work creatively. Creativity is one of the issues on *Merdeka Belajar*. Thus, the use of corpus is in line with the program. The last part is giving assignment to the participants. By doing this assignment, the participants try to use the corpus and observe any language phenomena

that they are interested in. The participants were assigned to use BNC because they need to explore in discourse level.

To evaluate the program, observation during the training was conducted. In addition, a discussion with MGMP committee was done. In the observation, participants' attitudes were noticed. It can be concluded that the participants actively took a part. They listened to the trainers, gave relevant responses, shared thoughts, and asked questions. They also gave responses to other participants' statements. The interaction during the training is good because the participants and the trainer gave responses to each other well. The participants were also able to raise discussion on the relation of the topic to the implementation in the English class.

The MGMP committee gave a good evaluation towards the training. The knowledge given is very useful for the participants. They also invited the trainer to give another session on corpus linguistics. Participants' enthusiasm is high.

Picture 2
The discussion activity

Conclusion

The community service program makes the participants understand basic concepts on corpus linguistics and its implication to English language learners at senior high school levels. They are aware that there are many variations in Englsih to be explored. The results of exploration can be the inputs to syllabus and lesson plan development, designing test instruments, and conducting teaching activities in the class. The participants also understand the importance of using natural language in teaching activities. It helps students' to be fluent language users. Besides, learning variations of words and expressions also can create creative lessons so the students do not get bored. Outside the classroom, the students get exposed with various sources of English language use. They

can use those sources to know how English is used for communication. If the English lessons do not represent the use of language in real life, the students might wonder why they need to learn "the classroom English". They will not see the use of learning English in the class to the real life. And finally, they will be demotivated.

References

Fligelstone, S. (1993). Some reflections on teaching from a corpus linguistics perspective. *ICAME Journal, 17*, 97-109.

Hunston, S. (2011). *Corpus approaches to evaluation: Phraseology and evaluative language*. Routledge.

O'Keeffe, A., McCarthy, M., & Carter, R. (2009). *From corpus to classroom*. Cambridge University Press.

Thornbury, S. (2010). What can a corpus tell us about discourse?. In . A. O'Keeffe & M. McCarthy, *The Routledge Handbook of Corpus Linguistics* (pp. 271-287), Routledge.