

Designing an English vocabulary workbook based on corpus-based approach: what actual learning task to incorporate target vocabularies into speaking

Andi Rizki Fauzi

Akademi Parivisata STIPARY Yogyakarta
fauziandirizki2@gmail.com

ABSTRACT

Corpus linguistics is a recent approach that many researchers use to conduct a study on language learning and acquisition. Through Corpus, the language learners can discover how a certain word is used authentically in the context. Corpus also allows the learners to find what the most frequent words or phrases emerge in the corpus so it can be one of the indicators that the word and the phrases are generally used by the general speakers. Moreover, it is possible to search and observe the collocation of a certain word so that learners can speak or write natural-sounding English. Therefore, several studies have already addressed how corpus usage will improve vocabulary mastery. The problem is that after knowing the important words and collocation, the learners need to be able to use them in productive skills such as speaking and writing. In the context of hospitality students, speaking is the most significant skill for their potential career as they can use verbal language dominantly in their job. Therefore, the goal of using corpus linguistics for hospitality students should be more than gaining vocabulary. Some exercise activities that enable the students to be able to use the target vocabularies taken from the corpus are very important to be designed. Based on the problem above, this article aimed to review some recent literature regarding the applications of the corpus in the area of language learning and teaching so that some proper ways of the task to help the students incorporate their vocabulary into speaking can lead to the development of English vocabulary workbook model based on the Corpus-based approach.

Keywords: *Corpus linguistics, vocabulary, speaking, workbook*

INTRODUCTION

The use of technology today significantly contributes to the teaching and learning process. In the educational environment of English language lesson, technology provides many benefits for the teachers such as to enhance and to stimulate the interest of students, activate their thinking, the effectiveness of learning, individualize instruction, increase speed of presentation and assimilation of information, and rapid adjustments knowledge if necessary (Kerimbaeva et al. 2017). From this explanation, it can be said that the use of

technology does not only make it easier for teachers to explain learning material but also activate the students thinking. The wealth of resources available on the internet develops inquiry skills, provides access to the world of knowledge, and requires learners' critical thinking, to decide which information to accept and absorb and which to reject (Giavrimis, 2011).

Based on the statement above, the existence of the internet makes a big alternation in the way of teaching English. For instance, the use of a printed dictionary is currently left behind. The reason for this is that the students prefer to use the easier learning sources offered by the online dictionary. By using the online dictionary the students can search for a certain word they are looking for instantly even it is provided in an application embedded in their mobile phone so that they can bring and use it easily every time and everywhere.

Another way of using technology in teaching English is by using the corpus. A corpus is a body of written text or transcribed speech that can serve as a basis for linguistic analysis and description (Kennedy, 2014). A corpus is a collection of texts which is stored on a computer. The texts can be written or spoken language. Corpus is not only useful for a researcher but also for language learners as it essentially tells us what language is like, and the main argument in favor of using a corpus is that it is a more reliable guide to language use than native speaker intuition is (Hunston, 2018). The use of corpus enables both teachers and learners to get exposure to the language used by native speakers. It is very essential especially for learners who use English as a foreign language and those who learn English for a specific purpose. As stated by Donesch-Jezo (2013), students should be exposed to an ample of authentic texts. This enables them to meet vocabulary in a real context, which facilitates its consolidation and memorization.

In connection with vocabulary learning, the use of corpus was proven as a very helpful approach by many researchers. One of the benefits of using corpus linguistics is that the corpus software can analyze the most frequent words that appear in certain corpus data. it gives a picture regarding what vocabularies should be learned in the priority list. The interest of many researchers regarding the contribution of linguistic corpus to improving vocabulary mastery cannot be separated from the importance of vocabulary in language learning. According to Thornbury (2002), teaching words is a crucial aspect of learning a language as languages are based on words.

However, there are not many studies related to how linguistic corpus contributes to improving speaking skills. This is very reasonable considering that speaking ability is not only measured through the use of vocabulary and grammar where the contribution of the linguistic corpus is very tangible. Therefore, it takes several tasks that can be integrated into learning activities so that the words that have been learned through linguistic corpus can be applied in speaking. This article is aimed to propose suitable tasks to incorporate the vocabulary learned through corpus into speaking activities which will be used as an input to design an English vocabulary workbook based on a corpus-based approach. A literature review is carried out to know the contributions of the tasks proposed.

Corpus-based approach

The use of a corpus-based approach is impossible without computer software used to process data so that the results of the analysis can be understood. Two of the most common software tools are *MonoConc Pro* and *WordSmith Tools*. Another popular software for corpus data analysis is *AntConc*. It is a freeware program that makes it suitable for people, schools, or colleges with a small budget, and runs on both Windows and Linux/Unix systems. (Anthony, 2004). The useful tool in all of the software is a concordancer tool. Here is an example of a concordance display from *AntConc* software.

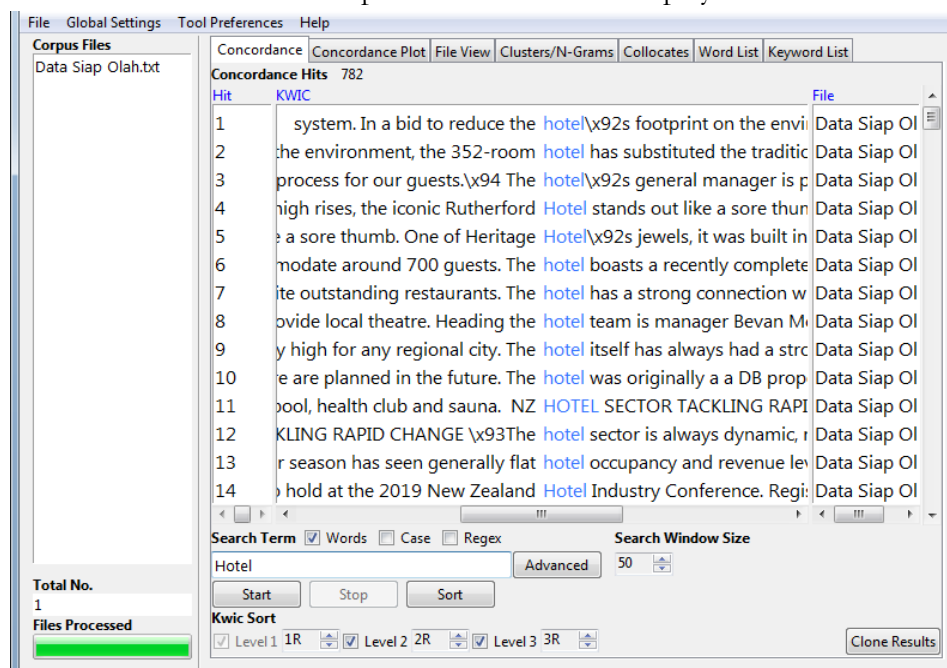


Figure 1. Concordance display from *AntConc* software.

The corpus software processes the data and presents some information such as frequency, phraseology, and collocation. The word frequency list can be seen as the result of the analysis using the software. It is one of the important information as the word frequency can indicate what the most frequent word is used in the corpus data. For instance, Coniam (2004) analyzed two styles of academic writing (Academic A is interested in language evaluation and computational analysis, Academic B with teaching and discourse analysis) and the result showed that the two sets of most common terms were very different. Academic A's high-frequency words include words pertaining to language analysis and parsing (verb, noun, party, parsing, etc.) and to language testing (test, evaluation, marker, assessment, proficiency) (test, examination, marker, assessment, proficiency). Academic B's high-frequency words contain words relating to teacher education (teacher, classroom, curriculum, learners) and discourse analysis (utterance, discourse, illocutionary, elicitation).

The result of this research showed that the word frequency as the result of software tools can give a picture of what words are most frequently used in a certain register. Hongyan (2018) explained that based on the data of the use frequency, the corpus can tell us the real high-frequency words in the whole vocabulary from a scientific perspective,

which is rather important in vocabulary teaching, as it decides the teaching contents in EFL vocabulary teaching. It can be said that the corpus analysis results in word frequency can help the teacher to consider what vocabulary should be learned and become more priority in teaching.

The next data presented in the software is phraseology. The term phraseology encompasses what has been referred to by various scholars as formulaic sequences, lexical bundles, lexical chunks, and collocations that is, the occurrence of two or more words within a short space of each other in a text (Ackerley, 2017). The result of phraseology data presented in *AntConc* can be seen in the example below:

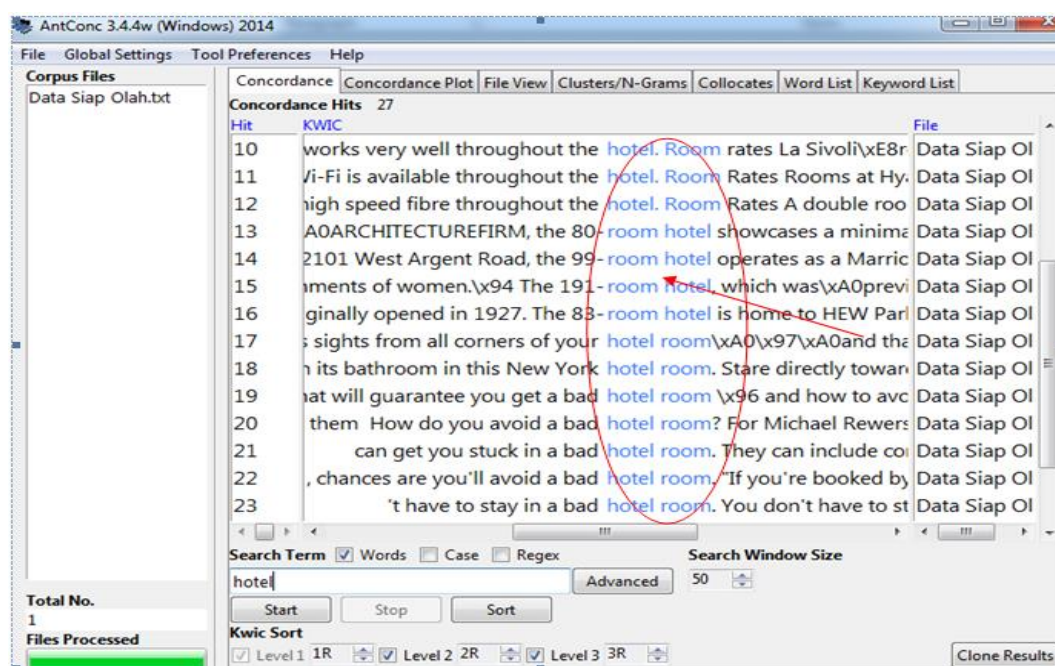


Figure 2. Phraseology display from *AntConc* software.

Based on the picture above, the use of software helps both teachers and students in finding information related to collocations and how many collocations appear in the data corpus.

However, there are several debates related to the use of the corpus, especially about how big the size of the corpus can be used so that the data from the analysis can be reliable. Although it is generally accepted that the bigger the corpus, the more reliable it is for making generalizations about the language use (Sinclair, 1991), it is also believed that a smaller corpus is generally more useful for classroom purposes. One issue that sometimes occurs, particularly when smaller corpora are used, is that of observed zero-frequency i.e. when a particular linguistic expression is not observed at all. If the corpus that is made is small, the problem that arises is that the results of data analysis from the corpus only provide a small picture of what is being studied (Anthony, 2013).

The use of corpus in teaching and learning English

Most recently, the use of linguistic corpus has been performed in the process of teaching and learning English. The use of corpus Corpus-based pedagogy is commonly associated with a 'discovery-based', inductive approach to language learning in which learners extrapolate rules based on their scrutiny of the concordance output. Traditional materials, on the other hand, tend to emphasize rule-based learning (Flowerdew,2015). Corpora provide several examples of an item searched in its context of use, but it does not give the meaning of the word or phrase on the screen. The meaning of a new item has to be deduced through the examples generated (Q'Keeffe, McCarthy, and Carter, 2007). Based on these studies, students must actively participate in the learning process using the corpus and it will automatically activate their thinking. For instance, in learning new vocabulary, the students will not consult the dictionary once they find an unfamiliar word. They should guess the meaning at first. Guessing the meaning in the context is very important to make the students retain the word in a long term memory.

The application of corpus linguistics in teaching can be divided into two aspects: the direct one: taking the relevant knowledge of corpus linguistics: means of developing a linguistic corpus, and applications of the linguistic corpus as the teaching materials; and the indirect one: based on both corpus and computer technology, including compiling a corpus-based dictionary, editing a grammatical reference, textbook, developing multimedia courseware, language learning software, or evaluating or testing tools (Sanyun,2005).

In the teaching and learning process, teachers typically incorporate corpus into their teaching in three ways: first, they collect data from corpus searches; then, they prepare materials on a level basis; and finally, they work with these materials. Also, teachers may use the available online corpora when teaching a particular language pattern or create specialized corpora from authentic texts or student papers and have students analyze the data (Reppen, 2011). From all the steps explained above, the teachers must construct the materials based on their students' level since the corpus presents the authentic text which is hard to comprehend particularly for those learning English as a second language. However, the application of corpus linguistics into EFL vocabulary teaching can almost perfectly deal with the lack of real context that becomes a common problem for language learners it is because the language data in the corpus are all from natural contexts, which can help the learners to use the word accurately and properly (Hongyan, 2018).

To solve the problems of the difficulty of authentic text for students, the teachers can integrate the corpus and online dictionary. According to Fauzi and Suradi (2019), in using corpus, assisting an online dictionary to the students gave positive impacts although for those who had a very low level of vocabulary mastery still needed more supports as the monolingual dictionary (English to English) was not easy enough for them.

A lot of studies was conducted on the use of corpus in many areas of language, such as grammar, writing, and vocabulary and the result proved that it made a positive contribution (Canaria & Ingleses, 2000; Paker,2017; Hanafiyeh& Keshi, 2013; Barabadi & Khajavi,2017; Fauzi,2018). Based on that research, the corpus will take an important role in teaching grammar and vocabulary. In contrast to the conventional approach, grammar is commonly taught by the teacher's examples; however, the corpus may present several

examples in the original text and may also present a comparison between two separate registers. In addition to that, by using corpus in teaching vocabulary, the students will be actively engaged in learning activities that can significantly increase their vocabulary.

The need for vocabulary workbook based on corpus-based approach

Vocabulary is the key to learning a language, including English. Therefore, it is natural that the study on the means of improving the efficiency of vocabulary teaching becomes the foremost of the practical linguists' attention (Hongyan,2018) and in the last decade, the use of linguistic corpus has been shown to have a positive impact on vocabulary learning as stated by Yavuz(2014) that the lexical approach in language teaching mostly depends on teaching words and collocations and concordance is a way of lexical analysis of words and their combinations. However, the vocabulary workbook based on corpus-based approach is very limited particularly for a special field such as tourism. One of the reasons is as stated by Vongpumivitch, et al,(2009) that deciding which words are worth teaching has not been a simple matter. It is very reasonable, as to create material such as a workbook, the teachers should consider some aspects such as the familiarity to the corpus software, collecting the data to build the corpus, analyze the data in the corpus such to determine which words will be selected, and the what activities must be engaged.

According to Sinclair and Renouf (1988: 148), a lexical syllabus would focus on '(a) the commonest word forms in a language; (b) the central patterns of usage; (c) the combinations which they usually form.' According to Xiao & Mcenery (2005), the focus of the lexical approach to language pedagogy is teaching collocations. All of the aspects above are offered by corpus. Through the corpus software, the most frequent words in a topic can be identified even their collocation. The corpus data is also taken from the authentic text and provides an overview of how the words are used.

Yang (2020) proposed two phases in the application of corpus linguistics to create content-based materials for pedagogical purposes: 1.) Identifying Target Words and Deciding When to Introduce Them, 2) developing and revising materials. Several principles were followed for the presentation of the target vocabulary: authentic materials from corpora would be used throughout, target vocabulary would be presented multiple times, materials would appear inconsistent formats for each lesson. The interview result of the research conducted by Yang also showed that the students said the new materials from the application of corpus linguistics help them memorize the vocabulary better.

The possible task to incorporate vocabulary into speaking activities

There is a challenge emerges when the corpus vocabulary workbook is designed particularly for tourism students. The hospitality students learn English vocabulary to be able to use them in verbal communication. the reason behind this is that they generally use English for speaking in their future job. Regarding this, the use of corpus does not only focus on how the students learn the target vocabularies but how the students can use them in speaking.

According to Luisa & Varela(2012), there are two ways of using corpora for the study of vocabulary. Teachers can either encourage students to make use of the relevant

software through hands-on, practical activities (explicit) or base their tasks on corpus data (implicit). In either case, corpus-based activities should be motivating, enlightening, and make sense to learners.

The research regarding how to design the materials and tasks using a corpus-based approach is very limited. One of the example which suggest the activities based on the leveraging corpus linguistics is given by Gavioli (2001). In the activities, the teachers can ask the students to find certain information in the concordance and discuss their finding with their partner as can be seen below:

WORKSHEET					
<i>handsome</i>	<i>good</i>	<i>attractive</i>	<i>pretty</i>	<i>tall</i>	<i>successful</i>
1. In pairs, choose three of these adjectives and have a look at their concordances.					
• Which words do they most frequently occur with? • Do they occur as part of a series of adjectives?					
2. Discuss your findings with other pairs, then with the rest of the class.					

Figure 2. DDL material initiating learner-centered corpus analysis (Gavioli 2001:120)

By discussion, the students are asked to find the most frequent words that occur in the corpus and finally asked to speak in English through discussion. This process is very essential to encourage them to speak English.

In this article, I propose the use of IELTS speaking part two. In IELTS part 2 (Long turn), the candidate speaks at length for 1 to 2 min on a topic chosen by the examiner. One minute of preparation time is given before the monologue. The examiner ends this part with one or two rounding-off questions (Fernandez, 2018). I modified this task to incorporate the vocabulary that the students learned through corpus into some activities as can be seen below:

Activity 1 :

The students will get a topic card that is related to the topic in the unit. They will be asked to speak for 1-2 minutes. However, before the talk, they will have 3 minutes as preparation time. In preparation, the students can develop their ideas and target vocabulary and its collocation they have learned to be used in the talk and write them down on a piece of paper in form of mind mapping to help them in speaking. For example:

Topic Card given : A hotel you stayed in

Describe a hotel you have stayed in.

You should say

- Where the hotel is

-
- Why you stayed at that particular hotel
 - Give details description of the hotel
 - Explain what makes the hotel special.
-

The target words to say:

five star hotel
luxury hotel
best hotel
professional services
best service.
whole staff

lovely guestrooms
excellent restaurants
suite guest room
standard room
great bars
hospitality
staff work

Figure 3. Example of mind mapping

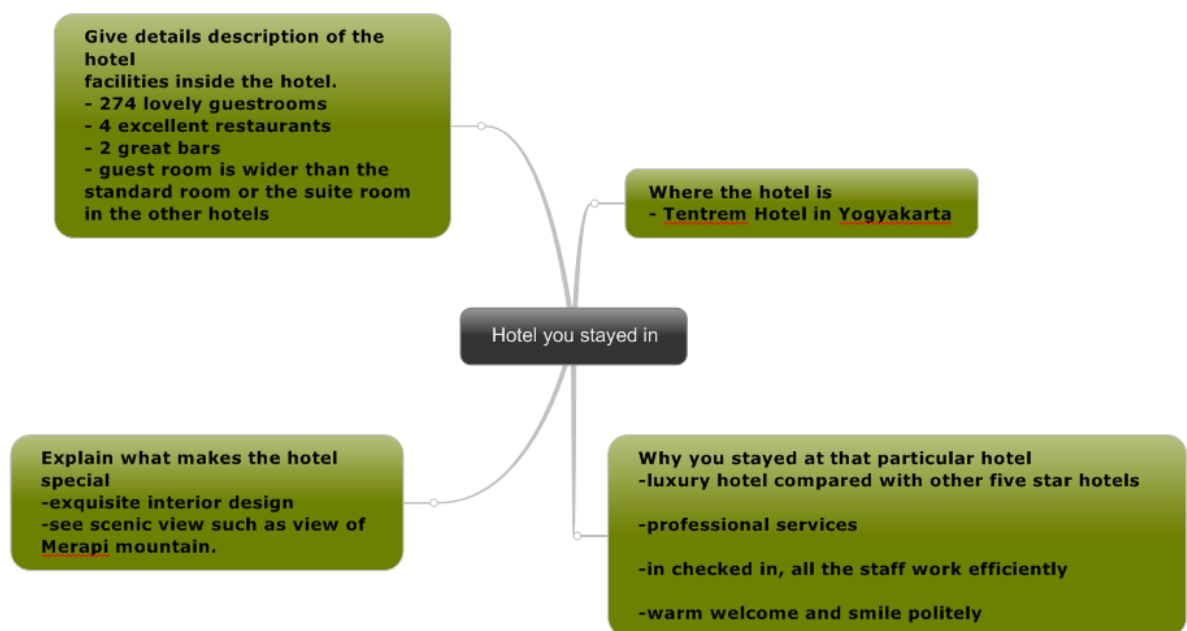


Figure 4. The example of mind mapping for speaking activity

Activity 2:

The students work in pairs in a speaking activity. One takes a role as a speaker and another helps record the voice. Then they will switch roles. They can do this activity as many as they need to make their speaking better. The partners can give inputs and suggestions to improve the quality of speaking in form of word choice, grammar, and pronunciation.

Activity 3:

After the students feel quite confident to speak, the students can speak up in front of the class. The teacher and the students can comment on the speakers as valuable input for the speakers regarding their speaking performance. This activity aims to know whether the

students can use the target vocabulary and its collocation they have learned or not and pronounce them correctly in their speaking. Here is an example of the student's speech:

Example of speaking:

I have stayed in Tentrem **Hotel** in Yogyakarta. The reason why I chose this **luxury hotel** from the other **five star hotel** is based on my friend recommendation. He said to me that the **professional services** given by the **whole staff** in Tentrem Hotel is **the best in Yogyakarta**. For example, when I checked in, all the **staff work efficiently** to make **the guest** get the **best service**. I didn't have to wait long for the check-in process. In addition to that, the **hospitality** of the entire **staff** made me believe that this is the **best hotel** I have ever stayed in. They greeted me with a warm welcome and smile politely.

After entering the **hotel**, I was very impressed with the facilities inside **the hotel**. It provides 274 **lovely guestrooms**, 4 **excellent restaurants**, and 2 **great bars**. I also surprised with a **suite guest room** that I chose to stay. It is wider than the **standard room** even the **suite room** in the other **hotels**. The room had exquisite interior design and I could see scenic view surrounding **the hotel** from my **room's balcony** such as view of merapi mountain. All the **services** and the stunning view I enjoyed made the experience staying in Tentrem hotel was very special and unforgettable.

Figure 5: The example of speaking activity

From the activities above, the students get chances to use the vocabulary they learned in their speaking activity. It is expected that the students can use them when they are working and improve their English speaking quality, particularly in the vocabulary aspect.

CONCLUSION

The use of corpus linguistics offers benefits in teaching and learning English especially, in the vocabulary aspect although the use of corpus also has many weaknesses. Some of the researchers conducted proved the effectiveness of using corpus in teaching and learning vocabulary. However, there are very limited materials such as workbooks based on a corpus-based approach that focuses on the performance task i.e speaking. It is already known that knowing the vocabulary is not enough if the students cannot use them in speaking, especially those who will use speaking activity in their future job such as hospitality students. The positive contribution of using corpus in teaching and learning vocabulary should be supported by a teaching and learning material based on the corpus-based approach which is not only focused on learning words but how to retain them in long term memory and use them in a speaking activity. IELTS speaking part 2 designs is proposed as the task given in the workbook. The possibility of the learners to use all the target words based on the corpus analysis makes the task can be carried out. However, the further investigation related to the use of the task needs to be conducted.

ACKNOWLEDGEMENTS

This article is part of research funded by the Ministry of Research and Technology / National Agency for Research and Innovation in 2020.

REFERENCES

- Ackerley, K. (2017). Effects of corpus-based instruction on phraseology in learner English. *Language Learning & Technology*, 21(3), 195–216. Retrieved from <http://llt.msu.edu/issues/october2017/ackerley.pdf>
- Anthony, L. (2004). IWLeL 2004 : An Interactive Workshop on Language e-Learning. In *AntConc: A Learner and Classroom Friendly, Multi-Platform* (pp. 7–13).
- _____. (2013). A critical look at software tools in corpus linguistics * 1. *Linguistic Research*, 30(2), 141–161.
- Barabadi, E., & Khajavi, Y. (2017). The effect of data-driven approach to teaching vocabulary on Iranian students ' learning of English vocabulary. *Cogent Education*, 6(1), 1–13. <https://doi.org/10.1080/2331186X.2017.1283876>
- Canaria, R., & Ingleses, D. E. (2000). Grammar Teaching in the 21st Century? *, 1999(April), 548–560
- Coniam, D. (2004). Concordancing oneself: Constructing individual textual profiles. *International Journal of Corpus Linguistics*, 9(2), 271–298. <https://doi.org/10.1075/ijcl.9.2.06con>
- Donesch-Jezo, E. (2013). Using language corpus in teaching foreign language vocabulary. *International Multidisciplinary e-Journal*, 2 (1), 11–25.
- Fauzi, A. R. (2018). Building the Students ' English Vocabulary for Tourism through Computer-Based Corpus Approach. *IJIELT*, 4(2), 133–148.
- Fauzi, A. R. & Suradi (2019). Integrating Online Corpus and Online Dictionary to Expand Student ' s English Vocabulary for Tourism 2 . *IJIELT*, 5(1), 13–26.
- Fernandez, C. J. (2018). Behind a spoken performance : test takers ' strategic reactions in a simulated part 3 of the IELTS speaking test. *Language Testing in Asia* 8 (18), 1–20 <https://doi.org/10.1186/s40468-018-0073-4>
- Flowerdew, L. (2015). Language learning theories underpinning corpus-based pedagogy. In F. Formato & A. Hardie (Eds.), *Corpus Linguistics 2015 Abstract Book*.
- Gavioli, L. (2001). *The learner as researcher: Introducing corpus concordancing in the classroom*. In G. Aston (Ed.), *Learning with Corpora*, (pp. 109–137). Houston: Athelstan.
- Giavrimis, P. (2011). Information and Communication Technologies and Development of Learners ' Critical Thinking: Primary School Teachers ' Attitudes, 4(3), 150–160. <https://doi.org/10.5539/ies.v4n3p150>
- Hanafiyeh, M., & Keshi, A. K. (2013). The Effect of Corpus-Based Instruction and Thesaurus-Based Teaching on Iranian EFL Learners' Grammatical Knowledge. *Appl. Sci. Res. Tribbble & Jones*, 3(2), 167–179.
- Hongyan, L. (2018). A Study on Corpus-Based EFL Vocabulary Teaching. *Journal of Intensive Studies on Language, Literature, Art, and Culture*, 2(2), 21–25.
- Hunston, S. (2018). *Introduction to a corpus in use*. Cambridge University Press. <https://doi.org/DOI:https://doi.org/10.1017/CBO9781139524773.002>
- Kennedy, G. (2014). *An Introduction to Corpus Linguistics*. London and New York: Routledge Taylor & Francis Group.

- Kerimbaeva, B. T., Niyazova, G. T., & Kaya, K. (2017). The Role of Computer Technology in Teaching English Language. *RUDN Journal of Informatization of Education*, 14(1), 108–113. <https://doi.org/10.22363/2312-8631-2017-14-1-108-113>
- Luisa, M., & Varela, R. (2012). Corpus Linguistics and Language Teaching: Learning English Vocabulary. *ES. Revista de Filología Inglesa*, 33, 285–300.
- Paker, T. (2017). The Effectiveness of Using Corpus-Based Materials in Vocabulary Teaching *International Journal of Language Academy Bütünce Temelli Etkinliklerin İngilizce Corpus-Based Materials in Vocabulary*, (April). <https://doi.org/10.18033/ijla.3494>
- Q'Keeffe, A., McCarthy, M. J. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press
- Reppen, R. (2011). Using corpora in the language classroom. In B. Tomlinson (Ed.), *Materials Development in Language Teaching (second edition)*. Cambridge: Cambridge University Press.
- Sanyun, L. (2005). *Study on Corpus and Vocabulary Teaching Strategy*. *Computer-assisted Foreign Language Education [J]*, Otc. 2005.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press
- Sinclair, J. and Renouf, A. 1988. 'A lexical syllabus for language learning' in R. Carter and M. McCarthy (eds.) *Vocabulary and Language Teaching*. London: Longman.
- Thorburry, S. (2002). *How to teach vocabulary*. England: Pearson Education Limited. Ur, P., (1998). *A course in language teaching*. Cambridge University Press.
- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). E NGLISH FOR Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33–41. <https://doi.org/10.1016/j.esp.2008.08.003>
- Xiao, Z., & Mcenery, T. (2005). Corpora and language education. Retrieved from http://web.uni-plovdiv.bg/rousni/files/Unit_10_Corpora_and_language_studies_pt2_3.pdf
- Yang, Z. (2020). The Study and Application of Corpus Linguistics to Create Content-based Materials for Pedagogical Purposes. *Journal of Language Teaching and Research*, 11(3), 480–485.
- Yavuz, F. (2014). The Use of Concordancing Programs in {ELT}. *Procedia - Social and Behavioral Sciences*, 116(0), 2312–2315. <https://doi.org/http://dx.doi.org/10.1016/j.sbspro.2014.01.565>