

## Corpus-based Method in Language Learning and Teaching

Ameen Ali Mohammed Al-Gamal<sup>1</sup> & Eassa Ali Mohammed Ali<sup>2</sup>

<sup>1</sup> PhD Research Scholar, Department of Linguistics, University of Kerala  
English Language Instructor, Department of English Language, Faculty of Education,  
Thamar University, Republic of Yemen

<sup>2</sup> PhD Research Scholar, Department of Linguistics, University of Kerala,  
English Language Instructor, College of Languages, Sana'a University, Republic of Yemen

Received: February 21, 2019

Accepted: March 27, 2019

**ABSTRACT:** *A corpus is a large collection of written or spoken texts that can be developed manually or electronically and be representative of the authentic language data. The corpus has numerous applications in the field of applied linguistics such as development of dictionaries, construction of grammar books, enhancement of language learning and teaching, study of dialects, machine translation, etc. The present paper is meant to review some recent literature regarding the applications of corpus in the area of language learning and teaching, whether a first, second or foreign language. The review focuses on some empirical findings from experiments conducted on learners of various levels of language proficiency in different societies. Based on the empirical findings probed from previous studies, it was indicated that corpus-based method of learning and teaching a language is effective and learners get direct access to data, hence, enhancing their language learning in various linguistic skills and develop more autonomy.*

**Key Words:** *corpus-based method, explicit teaching, language learning*

### Introduction

Using corpora in language learning and teaching dates back to the 1980s when corpora emerged in their modern form as systematically organised electronic collections of texts (Vyatkin and Boulton 2017: 1). Corpus linguistics has taken its place rapidly in language teaching and research, especially with the advent of computer technology. There is an agreement among researchers that corpus data enrich the process of learning languages and is considered an essential resource in learning and teaching (Huang 2011: 482). As cited in Dash (2008: 28), Crystal (1997) defines corpus as ‘a large collection of linguistic data, either written texts or transcription of recorded speech, which can be used as a starting point of linguistic description or as a means of verifying hypotheses about a language’. In this definition, the corpus is looked at as a collection of written as well as spoken texts. These texts can be stored in machines-readable forms from which they give access to all sorts of linguistic description and analysis. For Kennedy (1998: 3), who opposes the definition offered by Crystal, the corpus is ‘a large collection of texts assumed to be representative of a given language, dialect, or other subset of a language, to be used for linguistic analysis’. Commenting on the above definition by Kennedy, Dash (2008: 28) has the view that the definition may encompass the socio-linguistic aspects induced within a language, but the definition miserably fails to divert attention to the linguistic criteria considered necessary for designing a corpus.

Corpus is considered as a relatively new trend in language teaching, therefore, some hold back from incorporating them in the process of learning and teaching. Some argue that they are technically challenging and time-consuming and some raise the issue that some types corpus software do not have free access for learners and when accessing the free online corpus, students encounter large amount of data and therefore get confused. However, the researchers argue that these complaints can be regulated and resolved by having a good guidance from well-trained teachers.

Today, corpora are found in classrooms and they contribute in learning languages whether a first, second or foreign language. The corpora can be utilised in developing study materials for beginner and advanced levels in all language skills; be it reading, writing, vocabulary, listening or speaking. They help in developing the ability of the learners and improve their linguistic skills including synonyms, antonyms, polysemy, homonymy, etc. As Barlow (1996:2) rightly states:

“Corpora can reveal not only the range of patterns of a language that the learner must assimilate, but also their frequency, which is an important factor in materials development and syllabus design. Moreover, the use of corpora has the potential to radically alter the field of second language learning, and perhaps Linguistics as well”.

From the above quotation, textbooks in language teaching that are built on corpora offer more common and frequent examples used in real-life than those which are of less usages. Dash (2008: 119) supplements the above quotation by stating that 'corpus-based teaching materials are reliable and authentic because they give attention to the common choices of usage than those that are less common'.

Nowadays, computers come to combine with the traditional methods of language teaching and formed what is called Computer-Assisted Language Learning (CALL) or Computer-Assisted Language Teaching (CALT). Learners now can have direct access to corpus, without even having previous experience, and generate examples and exercises for learning purposes and subsequently improve their learning skills. Corpora provide learners with multiple examples either for learning single words, phrases, clauses, or even complete sentences. For one single activity, it provides learners with numerous demonstrations which show the variation in grammar. It can be helpful in learning synonyms, vocabulary, collocations, idioms, proverbs, etc. Several demonstrations for one single activity are expected to have positive impact on the learner to comprehend and retain what they are learning, in addition to, developing their skills in research.

Based on the researchers' experience in language learning, especially English as a foreign language (EFL hereafter) when they were students at schools, and also their experience as teachers of English at schools and universities, their observation is that the problem encountered by the learners of EFL is about the inability of the learners to formulate complete meaningful sentences or write up multiple sentences that convey one sense. Therefore, they believe that corpus-based learning does help improve learners' language forms and function. Aijmer (2009) ascertains that the frequency information obtained from the corpus supports the views on the acquisition of languages. Here, the learner tracks the various lexico-grammatical patterns helped by the repetition and the established data of form-meaning links.

### **Findings and discussion of reviewed studies**

As Huang (2011: 483) has rightly stated 'learners should be provided with opportunities to develop a 'feel' for the language through corpus data'. He stresses on the aspect of discovery learning and that learners should be given the opportunity to get a direct access to the data and build on their own profile of meaning and uses through the use of corpora.

#### **Study 1**

Boontam and Phoocharoensil (2018) carried out a study on 30 young Thai learners of grade 4 with the aim to examine the effectiveness of learning three English *prepositions* (during, among and between) through data-driven learning (DDL henceforth) method. Another aim of the study was to examine the attitude held by the learners towards learning through the DDL activities. The participants were learning EFL (Intensive English Programme) at a private school in Nakhonsawan province, Thailand. The researchers were part of this school and they adopted the convenience sampling method to select the participants. The study took a six-week period in which participants attended 50 minutes (once a week) as extra activity class. Participants received two tests: pre-test and post-test. Both tests had the same test items which consisted of 15 items of gap-filling, 15 items of grammaticality judgment test and 5 items of sentence building. When compared to the pre-test answers, the results of the questionnaire and interview showed that there was a noticeable improvement in the post-test responses and the researchers relate the reason of improvement to the effect of DDL activities they received for 6 weeks. Similarly, participants showed positive attitudes towards learning through the DDL method and they described it as helpful, fun and challenging.

#### **Study 2**

The experimental study carried out by Çelik and Elkatmiş (2013) compared two methods of learning Turkish *punctuation marks*; that is, corpus-based method and lecture-based method. The study aimed at exploring the impact of exposing language learners to corpus-assisted language teaching in learning the punctuation marks. A total of 171 university level learners from Primary Education Department at Kirikkale University, Turkey, constituted the sample of the study. The skills meant to be tested were: full stop, comma, semi-colon, colon, triple dot, exclamation mark, quotation mark, apostrophe and hyphen. The study followed the research design of pre-test-post-test control group model. The study comprised two experimental and two control groups. The instruments utilised in data collection were an achievement test on punctuation marks, a questionnaire on the corpus-assisted language learning approach and semi-structured interview forms (conducted with 10 participants in the experiment groups). Post-test results revealed that the experimental groups who were subject to the corpus-assisted language learning approach performed better than the control groups that studied on lecture-based method. Based on the findings, majority of the learners in the experimental groups had positive perception and were pleased to study punctuation marks through concordance outputs and that corpus-assisted language learning is an effective method in learning the Turkish language. These pre-service teachers showed their desire to apply corpus-assisted language learning activities and materials once they go to classroom after graduation.

### Study 3

Mukherjee (2004) conducted test workshops on corpus linguistics for 248 qualified English language teachers at secondary schools in North Rhine-Westphalia in Germany in 2001 and 2002. The intention of holding the workshops was of three-folds: first, to introduce teachers of English at secondary schools to basic *principles and methods in corpus linguistics*. Secondly, to familiarise teachers *with language pedagogical applications and implications of corpus-based research*. Thirdly, to find out *what teachers know about corpus linguistics* before the test workshop, and what they think about the relevance of corpus linguistics to their own classroom practice after the test workshop. Participants were asked to answer some questions before the commencement of the workshop and some other questions after the completion of the workshop. During the workshops delivered at different schools, teachers were introduced to some basic concepts in corpus linguistics such as the corpora of present-day English, the representativeness of generated data, the notion of word-lists and concordances, the applications of corpus data in classrooms, etc. The conclusion reached by Mukherjee was that before conducting the workshops the results of the survey revealed that very few English language teachers in Germany know about corpus linguistics and its application in teaching processes. However, after the workshops, more than 95% of the teachers became well aware of benefits that can be gained from the use of corpora in teaching, especially in English language teaching.

### Study 4

Levchenko (2017) conducted a study on 10 doctoral students (6 males and 4 females) from Samara University, Russia. The study aimed to increase the academic awareness of using corpus-based classroom activities (C-BCA) in *English for academic purposes* (EAP) classes. The study concentrated on the improvement of writing competencies. As a methodological resource, the researcher employed the Academic Word List (AWL) that contains 570 word families (developed by Coxhead in 2000). The study took a fifteen-week period in which participants attended twice a week for two hours. The participants were from various fields such as History, Sociology, Psychology, Pedagogy, Management and Economics. The participants were enrolled in English classes as a preparation to pass their postgraduate (EAP) exam and improve their English skills in order to publish research papers, deliver presentations and for improving their communication skills in their scientific fields. When it came to the learning process, some students worked online and some others used printed copies of corpora. Participants were given three feedback surveys in the form of a questionnaire at three different stages; at the beginning, in the middle and at the end of the intervention study. Both qualitative and quantitative methods of analysis were utilised. Based on the results, participants' perception was that corpus approach was too helpful to the development of EAP skills. They affirmed that C-BCA increased their confidence towards L2 academic writing tasks and judged it as an effective teaching tool. As the chief concentration of the study was on the development of the academic writing skills, participants reflected that after the training they came to realise the high-frequency academic words that they should incorporate in their future writings.

### Study 5

In collaboration with Noor and Amir, Al-Mahbashi (2015) carried out a study on 60 Yemeni female students from the second level in the English Language Department, Faculty of Education, Sana'a University in 2014-2015. The study had the intention to explore the effect of the DDL (corpus-based learning) over the dictionary use on developing *word meaning* and *collocation* of receptive vocabulary knowledge among Yemeni EFL university learners. The participants were divided into two groups: an experimental group, called (DDL group) and a control group, called (dictionary group). Participants responded to three tests: pre-test, post-test and delayed post-test. The sessions were held twice a week for two hours in the reading course classroom. As a methodological resource, the researchers employed the Corpus of Contemporary American English (COCA) developed by Davies (2008) in designing the DDL activities for the experimental group. The pre-test examined the participants' *word meaning knowledge* with two questions whether they know the meaning of the provided words in the test items or not. As for the collocation knowledge, participants were asked to choose the correct collocation of a word from three alternatives. The structure of the post-test and delayed post-test was same as those in the pre-test. Both tests were identical in the number of examined words except for an order change of words in the delayed post-test.

Due to the insufficient availability of computers and lack of high-speed Internet access in the Department of English, where the study was executed, vocabulary corpus materials were developed for the experimental group as printed worksheets in which word meaning and word collocation were included. As for the control group, a dictionary was provided to learn the meaning of the target words. Many participants used Oxford Word Power dictionary and some downloaded Longman dictionary on their mobile phones. All participants

used the Oxford collocation dictionary to learn collocations. The reference dictionary does not supply examples for each collocated word; therefore, four printed materials were prepared to cover one example for each collocation. In addition to that, an activity for each set of the target collocation words was designed.

The following procedures were followed in the training: First, both groups were introduced to the meanings and the types of collocations. Then the pre-test followed. The pre-test results revealed that 28 words out of the 30 were unknown to the students in both groups. Then experimental group got two training sessions working on an activity to observe the context of the target words in the concordance lines. They worked under the support and guidance of the teacher (researcher). The dictionary group was taught the same vocabulary with the same two aspects using only the dictionary and the designed collocation materials. After four weeks of instruction using corpus and dictionary, participants in both groups took the post-test. The participants were then given the delayed post-test one month after the treatment.

The pre-test results revealed that there was no significant difference between the mean scores of the two groups. Though the post-test results revealed that the control group and experimental group achieved a noticeable improvement compared to their pre-test results, the t-test results showed that the experimental group performed significantly better than the control group. That could be due to the effect of the DDL method that led to the improvement of experimental group performance. The performance of the participants was expected to decrease, in the delayed post-test results that took place after a month from the treatment, but the results revealed an improvement in the performance of the two groups. The researchers believe that this improvement could have happened due to the intensive preparation for the mid-term exam in which participants were given the opportunity to review the vocabulary after the treatment; so that, the delayed post-test results for the two groups was not surprising. When analysing the delayed post-test results, it was also found that the gain and retention by the experimental group was higher than that of the control group. The researchers concluded by stating that 'The long and short-term impact of DDL method could be attributed to learning the target words in rich contexts through numerous exposures to them that would subsequently lead to vocabulary acquisition (through what is called discovery learning)'.

## Conclusion

The reviewed studies have shown how corpus-based method could contribute to the field of language learning and teaching regardless of the levels of proficiency of the learners. In the field of learning and teaching, the corpus techniques are especially helpful for both learners and teachers. Corpus answers students' questions and opens up wide doors for the teachers to prepare their teaching aids in an interesting and authentic manner. It offers flexible learning strategies when handled properly. Based on the reviewed studies, results revealed that corpus-based method is effective in learning and teaching languages. In DDL specifically, learners can have a direct access to the data (hands-on) and that results in highly effective and efficient learning outcomes. Corpora can make a difference in language learning and teaching in which learners become linguistic researchers and teachers become facilitators, guides and coordinators of the process.

## References

1. Aijmer, K. (ed.) (2009). *Corpora and language teaching*. Amsterdam: John Benjamins.
2. Al-Mahbashi, A. Noor, N. M. and Amir, Z. (2015). The effect of data driven learning on receptive vocabulary knowledge of Yemeni university learners. *3L: The Southeast Asian Journal of English Language Studies*, 21 (3), 13-24.
3. Barlow, M. (1996). Corpora for theory and practice, *International Journal of Corpus Linguistics*, 1(1), 1-38.
4. Boontam, P and Phoocharoensil, S. (2018). Effectiveness of English preposition learning through Data-Driven Learning (DDL). *3L: The Southeast Asian Journal of English Language Studies - Vol 24(3)*, 125 - 141.
5. Çelik, S. and Elkatmış, M. (2013). The Effect of corpus assisted language teaching on the learners' proper use of punctuation marks, *Educational Sciences: Theory & Practice*, 13(2), 1090-1094.
6. Coxhead, A. (2000). A new AWL. *TESOL Quarterly* 34(2), 213-238.
7. Dash, N. (2008). *Corpus linguistics: An Introduction*. Dorling Kindersley: New Delhi, India.
8. Huang, L. (2011). Corpus-aided language learning. *ELT Journal*, 65(4), 481-484.
9. Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Addison Wesley Longman Inc.
10. Mukherjee, J. (2004). Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In: U. Connor/T. Upton (eds.). *Applied Corpus Linguistics: A Multidimensional Perspective*. Amsterdam: Rodopi. 239-250.
11. Levchenko, V. (2017). Use of Corpus-based classroom activities in developing academic awareness in doctoral students, *The New Educational Review*, 48, 28-40. DOI: 10.15804/tner.2017.48.2.02
12. Vyatkina, N. and Boulto, A. (2017). Corpora in language learning and teaching. *Language Learning & Technology*, 21(3), 1-8.