Investigating Corpus Usage by Sudanese Undergraduates in Learning English Collocations

IBRAHIM ALSAFI ABDULDAFI
English Language Department
Bayan College of Science and Technology
ABDALLA YASSIN ABDALLA
English Language Department
College of Languages
Sudan University of Science and Technology

Abstract

This paper aims to investigate the use of corpora by Sudanese undergraduates in learning English collocations. To do so, the quantitative approach is adopted and then second year students were chosen as the study subjects. These subjects study Biomedical Engineering at Bayan College of Science and Technology, Khartoum. To do the experiment, the students were divided into two groups; the control group in which students were taught collocations (delexical verbs) through the traditional strategy, and the experimental group in which students were taught how to use corpora and then taught the same collocations using British National Corpus (BNC). After the experiment, results of the post-test revealed that while the performance of the two groups was improved in the post-test, the performance of the experimental group was significantly better than that of the control group. It is thus concluded that corpora have the potentiality of enhancing learners' collocational knowledge and hence their language proficiency. This strengthens the role of corpora in EFL/ESL context.

Key words: collocational competence, language proficiency, corpora, context, language learning.

It is of no doubt that English language has become the world's most essential language. This essentiality has increased the

demand for effective learning of English language and hence the demand to effective and life-changing strategies to English language learning and teaching. It is believed that one of the most crucial factors to achieving natural and native- like language production is mastering collocations. Firth (1957) is responsible for bringing up the term collocation into existence. According to him, "*the meaning of a word should be known by the company it keeps*". This notion substantiates the fact that words should not be studied in isolation and away from context, rather; they have to be contextualized and treated accordingly. Similarly, Lewis (1993:204), provided another view about the nature of collocations claiming that "*instead of words we consciously try to think of collocations, and to present these in expressions rather than trying to break things into ever smaller pieces. There is a conscious effort to see things in larger and more holistic ways*". In another context, emphasizing the significance of collocations, Nasir (2011:1) stated that *"it is important for language learners to master collocations so that their language sounds more natural and easily understood by native speakers"*. Despite that, collocations have received small fortune in language research and English language learners, even at advanced levels, still struggle to master them. This is especially true for Arab learners of English. In this regard, Nasir (2011:2) claimed that *"few studies have investigated the collocational knowledge of Arab speaking learners of English such as El-khatib, (1984); Al-Zahrani, (1998); Mahmoud, (2005)"*. These studies proved that Arab learners of English face difficulties when it comes to dealing with English collocations. In the Sudanese context, for instance, Nasir (2011:68) conducted a study on assessing English collocational knowledge among Sudanese EFL university students. She stated that *"the lower percentages are more than the higher ones, which are clear evidence that the students' performance on collocation tests is very low and collocations are a problem for the students in language learning"*. This suggests that collocations are

problematic part of word knowledge that requires more research and attention.

It has recently been acknowledged that collocations are considered to be essential for achieving communicative competence, and therefore should be integrated into second and foreign language learning and teaching (e.g., Cowie, 1992; Bahns, 1993; Wei, 1999; Lewis, 2000). Today, language learning and teaching, especially collocations have been revolutionized by the use of corpora due to their inclusion of a wide range of natural language data. This body of language has been proved to be fruitful in familiarizing learners as well as teachers with how language is actually used by native speakers in actual contexts. To fully evaluate this, the present paper focuses on corpus usage by Sudanese university students in learning collocations, delexical verbs in particular. More specifically, the paper attempts to answer the following research questions:

- To what extent are English collocations challenging for undergraduates in terms of usage?
- To what extent is there a significant difference between those who study through corpora and those who do not?

WHAT ARE COLLOCATIONS

The term collocation is defined by Oxford Advanced Learner's Dictionary as a combination of words in a language that happens very often and more frequently than would happen by chance (2007:281). Speaking of the nature of collocations, O'Keeffe et al. (2007) stated that collocations are not absolute or deterministic, but are probabilistic events, resulting from repeated combinations used and encountered by the speakers of any language. For instance, we say *bitterly disappointed* in preference to (but not the absolute prohibition of) *sourly disappointed*; tea is usually *strong* but cars are *powerful*, and so on (p.59). Researchers seem not to be in agreement when it comes to defining collocation, rather; the term is governed by

the context. For instance, in corpus linguistics, collocation refers to a sequence of words that co-occur more often than would be expected by chance; in phraseology, collocation is a sub-type of phraseme, and so on.

According to the present paper, collocation is a matter of statistical preference that could powerfully be demonstrated by the utilization of corpus data. Hence, the paper employs corpora in an attempt to solve collocation learning difficulties that face EFL/ESL university students.

CLASSIFICATION OF COLLOCATIONS

Collocations have differently been categorized into various categories by different scholars and researchers. The commonest classification is Firthian's (1957) in his semantic theory. According to him, collocations are classified into lexical collocations and grammatical collocations, but they are primarily semantically (not grammatically) based (cited in Bussmann, 1998). Collocations fall into two major groups: grammatical collocations and lexical collocations (Benson et al., 1986, p. ix).

As highlighted by Benson et al. (1986), English collocations are categorized into lexical collocations and grammatical collocations (with 7 and 8 types respectively). To begin with, lexical collocations are characterized by the fact that their meanings mostly reflect the meaning of their lexical constituents. As Benson, Benson and Ilson (1986) put it, lexical collocations are only made up of verbs, adjectives, nouns and verbs in different possible combinations. As demonstrated in the following.

(**1**) verb + noun (investigate a problem).
(**2**) adjective + noun (intolerable situation).
(**3**) noun + verb (volcanoes erupt).
(**4**) noun1 + of + noun2 (a system of life).

(**5**) adverb + adjective (easily acquired).

(**6**) verb + adverb (drive carelessly).

Grammatical collocations, on the other hand, can be defined as phrases made up of a dominant word (verb, noun, and adjective) and a preposition or grammatical structure like infinitive or clause. As shown in Benson, Benson and Ilson (1986) in their introduction to their BBI Combinatory Dictionary of English, grammatical collocations fall into the following combinations: Noun + preposition, noun + to-infinitive, noun + that-clause, preposition + noun, adjective + preposition, predicate adjective + to-infinitive, adjective + that-clause, and the English 19 verb patterns (cited in Ali, 2017, p. 12).

SIGNIFICANCE OF COLLOCATIONS TO EFL/ESL LEARNERS

Collocations are considered to be of a great significance to achieving vocabulary knowledge of EFL/ESL learners and consequently their communicative competence. El-Dakhs (2015) confirmed that it is important to learn collocations due to their frequency of occurrence and their facilitatory effect on language processing (p. 71). Hill (2000) estimates that "collocations are found in up to 70% of everything we say, hear, read, or write" (p. 53). Similarly, Lewis (2000) considers collocations as the most common and most representative of English multi-word expressions. A language phenomenon with such high frequency of occurrence deserves considerable attention. Besides, the use of pre-fabricated chunks reduces processing effort and enhances language comprehension and production (cited in El-Dakhs 2015, p. 71). What makes collocations challenging for EFL/ESL learners to fully acquire, is the fact that each and every word in English language has to some extent collocational relationship with other words. This fact urges language learners and teachers to fully consider developing their knowledge of

collocations. These facts justify the need to fully integrate and represent these clusters of language into syllabus design, language learning, and language teaching. Similarly, O'Keeffe et al (2007) asserted that vocabulary syllabus for the basic level is incomplete without due attention being paid to the most frequent chunks, since many of them are as frequent as or more frequent than single items which everyone would agree must be taught (p. 46). O'Keeffe, McCarthy & Carter (2007:60) highlighted that the notion of collocation therefore shifts the emphasis from the single word to pairs of words as integrated chunks of meaning and usage, and collocation has now become an accepted aspect of vocabulary description and pedagogy (e.g. Lewis 2000; McCarthy and O'Dell 2005). More recently, Salih (2016) maintained that collocational knowledge, which is stored in chunks, is held by some to be the foundation of language learning, use, and knowledge. James (1998) confirmed that "adherence to the collocational conventions of a foreign language contributes greatly to one's idiomaticity and native-likeness, and not doing so announces one's foreignness" (p. 152). Emphasizing the significance of collocations to all uses of language, Nation (2001) concludes that collocational knowledge is required to all fluent and appropriate language use (cited in El-Dakhs, 2015). It is assumed by this paper that the best way to properly acquire collocations is via utilizing corpora as they provide chunks of language within context. So they are believed to be effective in learning and teaching collocations.

COMMON DELEXICAL VERBS

Delexical verbs are not so many in English language, the commonest among them are *do*, *make*, *take*, and *have*. These light verbs present some challenges to English language learners when it comes to what go with them, especially their right noun collocates i.e. their delexical structure. Most learners of English language, especially Arabs, mix do with

make and take with have. It is believed that this delexical structure poses some difficulties to EFL/ESL learners because they are given small attention in the process of learning as well as teaching. Similarly, O'Keefe et al. (2007) highlighted that delexical verbs are problematic since they occasionally occur with low-frequency items and learners have difficulty understanding the meaning of the whole expression. Hence, the present paper hypothesizes that corpora have the potential for learning this word combination as these verbs require representative source of language data. A source that enables learners to study them within context.

WHAT IS CORPUS

Oxford Concise Dictionary of Linguistics (2014), defined corpus as any systematic collection of speech or writing in a language or variety of a language. Thus, in particular, of large online collections, tagged and searchable for research purpose (p. 82). Another definition is provided by O'Keeffe, McCarthy and Carter, according to them, a corpus is a collection of texts, written or spoken, which is stored on a computer. In the past, the term was more associated with a body of work. For example, all the writings of one author (2007, p.1). According to the researcher, corpus can be defined as a collection of linguistic data meant to represent a language or part of a language. This collection of data is used for language analysis and investigation. It is believed that corpus linguistics has gained popularity after the arrival of computers due to the fact that computers have the ability to store millions of texts and the ability to analyze these texts using software programs. Because of the advent of computers, these huge amounts of information are ready for quantitative and qualitative analysis of the language. This claim is supported by Bibber, Conard and Reppen (1998) claiming that: a corpus is a principled collection of texts available for quantitative and qualitative analysis.

TYPES OF CORPORA

There are many different types of corpora available for different uses and purposes. They can contain written or spoken (transcribed) language, old texts or modern texts, and they can also contain texts of whole books, newspapers, etc. or parts of texts. Which type of corpus we are investigating depends mainly on the purpose (s) of creating that corpus. Corpora can be balanced/representative, monitor, parallel, comparable, diachronic, specialized, multimedia, etcetera.

REVIEW OF PREVIOUS STUDIES

This part reviews some previous studies similar or related to the present one. It deals with studies carried out both in the Sudanese context and outside Sudan so that a global picture of the problem under investigation could be provided. Rabie (2017), for example, investigated the problems of understanding and using idiomatic expressions and collocations among Sudanese university EFL students majoring in English language. In this context, Rabie (2017) claimed that: "Idiomatic expressions put non-native speakers in hot water both in written and oral contexts. In addition to that, material writers and language teachers find it difficult to develop effective materials and teach them effectively so that they can have a subsidiary place in syllabus" (p. 2). The study adopted a descriptive analytical method of research. A questionnaire was designed by the researcher for English language teachers as well as a test on idiomatic expressions and collocations was designed for students. The study findings revealed that students have problems in understanding and using idiomatic expressions and collocations, and mastering these language patterns enhances students' ability of expression and develops their receptive and productive skills. Furthermore, it is also

showed that idiomatic expressions and collocations are significant to learning a foreign language.

It is clearly demonstrated by above study that Sudanese EFL learners (even those majoring in English language) find it difficult to deal with idiomatic expressions and collocations. The main factors behind these difficulties are attributed to the teaching materials and the techniques adopted. It is assumed that corpora are vital tools for teaching and learning collocations and idiomatic expressions and other clusters of the language, because dealing with such language units is a matter of probabilistic and context preference rather than prohibition. This notion is demonstrated by O'Keeffe, McCarthy and Carter (2007) claiming that, common verbs such as *get*, *go*, *turn* and so on display distinct preferences for what they combine with. Things *turn* or *go grey*, *brown*, *white*, people *go* (but not *turn*) *mad*, *insane*, *bold*, *blind* (p. 59). So that dealing with these verbs is a matter of quantity which corpora are very effective at. Therefore, the current study adopts corpora as an effective strategy to learning collocations.

Another study was carried out in the Sudanese context by Ali (2017) examining Sudanese EFL undergraduates' knowledge and use of collocations. The study sought to explore the knowledge and use of lexical collocations and their potential effect upon EFL learners as well as their impact on learners' language proficiency at university level. Participants of the study were English language teachers and EFL students from College of Languages, Sudan University of Science and Technology. The study adopted the descriptive and analytic method and two data collection tools were used; a completion test for students and a questionnaire for teachers. Results of the study proved that students possess limited knowledge and awareness of English language collocations. Besides, the deviant answers of students' tests indicated learners' poor knowledge of lexical and grammatical collocations usage.

The study above also proved that Sudanese EFL learners face some difficulties in dealing with collocations, especially lexical collocations which are considered to be essential for achieving high level of language proficiency. Unlike the above study, it is assumed by the present paper that these difficulties arise due to the fact that students have not been exposed to corpus-based syllabuses which represent real language used by native speakers. So it is strongly believed that unlike dictionaries as suggested by Ali (2017), using corpora in language learning and teaching has been proved to be effective because corpora are wide language data collected from different contexts/registers. These collections provide different senses and uses of words in different contexts. Therefore, the present paper attempts to employ corpora as a learning and teaching approach for solving such collocation learning difficulties encountered by EFL/ESL learners.

In the Jordanian context, Alzi'abi (2017) conducted a study entitled: Guessing verb-adverb collocations: Arab EFL learners use of electronic dictionaries. The study explored the way advanced Arab EFL learners handle verb-adverb collocations using two learner's dictionaries. Alzi'abi (2017) stated that collocational studies have recently attracted a great deal of interest. He added that, to date, hardly any study has tackled Arab EFL learners' competency in the use of verb-adverb collocations. The sample of the study was 88 M.A students majoring in English at five Jordanian universities. The study results showed that even advanced EFL learners had considerable difficulty in providing extra adverb collocates of both frequent and infrequent verbs. Dictionary use was effective, the subjects performed significantly better especially with infrequent verbs. As shown above, collocations pose some challenges for EFL learners, even those of advanced levels. According to the present paper, these difficulties are attributed to a number of factors; prime among them is the teaching syllabuses students have been exposed to. It is suggested that

learners' performance would be much better if they are exposed to corpus-based materials. Hence, the prime aim of this paper is using corpora in tackling these difficulties.

In the Taiwanese context, Agnes (2012) conducted a study examining the effectiveness of a corpus-based instruction in deepening EFL learners' knowledge of periphrastic causatives. The study aimed to investigate whether a corpus-based instruction could deepen EFL learners' knowledge of periphrastic causatives: *make*, *cause*, and *let*. 47 Taiwanese undergraduates from two intact classes were selected as the subjects of the study. One class was chosen as the experimental group received three-month corpus-based instructions, while the other group which is the control group had no instruction on English causatives. Results of the study showed that the experimental group improved and outperformed the control group significantly in the post-test.

Agne's (2017) study is similar to the present one in terms of the learning approach adopted, which is using corpora in enhancing EFL learners' knowledge. As demonstrated by the above study, corpora have been proved to be an effective approach to English language learning and teaching because of their inclusion of large and authentic language data. These data are believed to be effective in investigating words and their grammatical environments which provide a baseline for understanding collocations of words. It is claimed by this study that understanding the use of individual words (such as periphrastic causatives, delexical verbs, etc.) cannot be achieved without their context of use which can best be provided by corpus-based activities. Consequently, corpora are suggested by the researcher as a new strategy to English language learning and teaching.

In conclusion, this part has reviewed and discussed the previous studies similar/relevant to the present one. These studies are discussed with regard to their designs adopted, the instruments of data collection employed and the results and

findings obtained. Some of these studies have dealt with learners' problems with collocation learning and use, while others have investigated the use of corpora in language learning and teaching. Some of these studies have come to the fact that corpora are effective strategy to language learning in comparison to other strategies but none of them have examined the use of corpora in learning delexical verbs. This has come to be the strength of this paper.

METHODOLOGY

This paper is quasi-experimental research. For achieving its aims, the quantitative approach to data collection has been adopted. Specifically, tests were chosen as an instrument of data collection. In this concern, Creswell (2009), defined quantitative research as a means for testing objective theories by examining the relationship among variables. These variables can be measured typically on instruments such as tests, questionnaires, etc., so that numbered data can be analyzed using statistical procedures. Concerning the population of the study, it is conducted on second year students of Biomedical Engineering (Bachelor) at Bayan College of Science and Technology. These students study English language as university required courses. 60 of them (males and females) were selected as the study sample. Then these 60 students were divided into experimental group (taught the delexical verbs using corpora) and control group (taught the same delexical verbs using traditional strategy). With regard to the instruments, two tests (pre-test and post-test) were employed in the data collection process. The tests were designed on collocations. Specifically the delexical verbs *make*, *do*, *have*, and *take*. Each test consists of thirty questions; the first fifteen for *make* and *do* and the other fifteen are for *take* and *have*. Students were given two choices (verbs) and were asked to choose the appropriate one. These tests were specially designed

by the researcher to measure students' performance on delexical verbs as it is assumed that these verbs are considered to be challenging for undergraduates in terms of usage. Regarding the analysis of data, the pre-test and post-test were collected and marked and then statistically analysed using the SPSS (The Statistical Packages of Social Sciences) so that accurate results could be obtained.

RESULTS

The data were inserted in the SPSS program and the following results were obtained.

Table 1: One-Sample Test

	Test Value = 0					
	T	df	Sig.(2- tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Pre-test	36.222	59	.000	15.80000	17.7614	19.8386
Post-test	44.521	59	.000	21.73333	19.8015	21.6652

The above table shows the performance of the two groups (the experimental group and control group). As demonstrated, there is a significant statistical difference between the two groups which means that the experimental group performed significantly better than the control group (sig less than .5). This result proves the fact that corpora are effective strategy in learning collocations of English language.

DISCUSSION

This paper aims to investigate the effect of corpus usage by Sudanese undergraduates in learning English collocations. The paper findings showed that the experimental group (who study delexical verbs via corpora) outperformed the control group

(who study the same delexical verbs through traditional strategy) which clearly indicates that corpora have the potentiality of enhancing learners' collocational knowledge and hence their language proficiency. In consistency with this study as shown in the previous studies section, a corpus-based study was conducted by Agnes (2012) in the Taiwanese context. Results of the study proved that the experimental group improved and outperformed the control group significantly in the post-test. In addition, the findings of this paper have also indicated that collocations, specifically delexical verbs, pose some difficulties for EFL learners in terms of usage as it is clearly shown in students' performance of the two tests. So it is assumed by this paper that corpora have the capability of providing these verbs within context and thus making it easy to acquire them. In conclusion, it can be implied that EFL teachers need to adopt corpus-based strategies/approaches that represent the real language used by native speakers, so that students' collocation use sounds more natural and native-like. In addition, for teachers to exploit corpora to the fullest, they should be trained on how to practically use them in their classrooms.

CONCLUSION

This paper mainly aims to investigate corpus usage by Sudanese undergraduates in learning English collocations. To do so, the following research questions have been raised by the researcher:
1. To what extent are English collocations challenging for undergraduates in terms of usage?
2. To what extent is there a significant difference between those who study through corpora and those who do not?

The quantitative approach is employed. Specifically, two tools were used in the data collection process; a pre-test and a post-

test. The population targeted by the study was second year students (Bachelor) of Biomedical Engineering at Bayan College of Science and Technology. The paper findings showed that the experimental group outperformed the control group which clearly indicates that corpora have the potentiality of enhancing learners' collocational competence and thus their language proficiency. In addition to that, the findings have also indicated that collocations, specifically delexical verbs pose some difficulties for EFL learners in terms of usage. So it is assumed by this paper that corpora are capable of providing these verbs within context and hence making it easy to acquire them.

REFERENCES

1. Agnes, L. (2012) *The effectiveness of a corpus-based instruction in deepening EFL learners' knowledge of periphrastic causatives*. Retrieved from https://tesol-international-journal.com/wp-content/uploads/2013/11/A6_V6.pdf [Accessed on 02/04/2018].

2. Ali, A. (2017). *Investigating the Sudanese EFL undergraduates knowledge and use of collocation* (Doctoral dissertation). Retrieved from http://repository.sustech.edu/bitstream/handle/123456789/18590/Investigating%20the%20Sudanese%20EFL....pdf?sequence=1 [Accessed on 11/07/2017].

3. Alzi'abi, S. (2017). *Guessing verb–adverb collocations: Arab EFL learners' use of electronic dictionaries*. http://lexikos.journals.ac.za Retrieved from file:///C:/Users/wn7/AppData/Local/Temp/1394-2237-1-PB.pdf [Accessed on 02/04/2018].

4. Benson, M., Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of English: A guide to word combinations*. Amsterdam: John Benjamins.

5. Bibber, D., Conard, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

6. Bussmann, S. (2006). *Routledge dictionary of language and linguistics*. New York: Routledge.

7. Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed methods approaches*. Singapore. SAGE Publication Asian-Pacific Pte. Ltd.

8. El-Dakhs, S. (2015). Collocational competence in English language teaching. *Arab world English journal*, 6 (1), 68-82.   Available   at https://www.researchgate.net/publication/282364488

9. Hill, J. (2000). Revising priorities: From grammatical failure to collocational success. In M. Lewis (eds.). Teaching collocation: Further developments in the lexical Approach. Hove: LTP, 47-67.

10. Nasir, O. (2011). *Assessing English collocational knowledge among Sudanese EFL university students*. [M.A thesis], Sudan University of Science and Technology.

11. Oxford Advanced Learner's Dictionary. (2007). Seventh ed. Oxford University Press. Firth, J. (1957). *Papers in linguistics*: 1934-1951. London: Oxford University Press

12. Lewis, M. (1993). *The lexical approach*. Hove: LTB Publications

13. James, C. (1998). *Errors in language learning and use: Exploring error analysis*. New York: Addison Wesley Longman Limited.

14. O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. New York: Cambridge University Press.

15. Oxford Concise Dictionary of Linguistics. (2004). Sixth ed. Oxford University Press.

16. Rabie, E. (2017) *An investigation into the problems of understanding and using idiomatic expressions and collocations among university EFL students*

17. *majoring in English language*.[Doctoral dissertation]. Sudan University of Science and Technology. Retrieved from http://repository.sustech.edu/handle/123456789/18656 Accessed on 05/05/2017].

18. Salih, A. F. (2016). *A corpus-based study of academic collocation use and patterns in postgraduate computer science students writing*. (Doctoral dissertation). University of Essex. Available at http://repository.essex.ac.uk/16426/1/FINAL%20AFNAN %20PHD%20HESIS-%20submit.pdf [Accessed on 01/08/2017].