

Instructional Efficacy of Corpus-Based Tools in Teaching Collocations to Iranian University Students with Different Majors

Javad Akbari

Department of English language, Isfahan (khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Hamidreza Haghverdi

Department of English language, Isfahan (khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Reza Biria

Department of English language, Isfahan (khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Abstract

Coxhead's (2000) Academic Word List (AWL) has been frequently utilized in EAP classrooms and reconsidered in light of various domain-specific corpora. Although well-received, the AWL has been criticized for ignoring the fact that words tend to show irregular distributions and be used in different ways across disciplines. Such difference concerns collocations. The present study investigated the instructional efficacy of corpus-based tools in teaching collocations to Iranian university students with different specialized domains. After administering the Solution Placement Test, participants were divided into control and experimental groups. Each group included seventy-five students with three specialized domains, namely; law, dentistry, and physical education. The experimental group was taught the collocations and lexical chunks through using corpus-based tools and the control group was taught through traditional method in which they did not receive any instructional tools, innovative materials, and instruments. The findings indicated that teaching collocations through corpus-based tools was of significant help to the students' retention and learning of collocations. Furthermore, the results showed that the experimental group outperformed the control group in both post-test and delayed post-test. The findings of the present study may offer implications to EFL teachers, materials developers, and curriculum designers.

Keywords: corpus-based tools, specialized domains, chunks, retention, collocations

INTRODUCTION

A collocation is a pair or group of words that are often used together. These combinations sound natural to native speakers all the time, but students of English have to make a special effort to learn them because they are often difficult to guess. Some collocations just sound wrong to native speakers of English. For example, the adjective

fast collocates with *cars*, not with a *glance* or the noun *mistake* collocates with the verb *make*, not with *do* or the adjective *asleep* collocates with the adverbs *sound* and *fast* not with *heavily*.

Collocations are two or more words which have a strong tendency to co-occur in a language as a prefabricated combination of two or more words in a particular context (Halliday, McIntosh, and Strevens 1968). They are one of the difficulties that second language learners, in particular, adult second language learners, have to deal with in the process of learning English. These difficulties depend on a variety of variables such as students' native language (L1) background, age, and personality.

Learning collocations will not only be influential in speaking more fluently and confidently but also lead to the following significant ramifications:

- I. Give you the most natural way to say something: smoking is strictly forbidden than smoking is strongly forbidden.
- II. Give you alternative ways of saying something, which may be more colorful, expressive or precise: instead of repeating it was very cold and very dark, we can say it was bitterly cold and pitch dark.
- III. Improve your style in writing: instead of saying poverty causes crime, you can say poverty breeds crime; instead of saying a big meal you can say a substantial meal. You may not need or want to use them in informal conversations, but in writing they can give your text more variety and make it read better.

As far as the term collocation is concerned, most researchers have defined collocations from the aspect of partnership or co-occurrence of words. Halliday and Hasan (1976) classified collocations from the aspect of discourse. Benson et al (1986) proposed that the classification of collocations is based on two major categories: lexical collocations and grammatical collocations.

In fact, there are varieties of techniques to teach and present vocabulary in a course such as physical demonstration, verbal explanation, providing the students with synonyms and antonyms, translation, using visual aids, asking learners to check the meaning in the dictionary, exemplification and presenting words in the context (Hedge, 2008; Nash & Snowling, 2006). However, it is evident that some of these propounded and ultimate techniques cannot bring about a great deal of vocabulary retention since learners are not able to make use of presented words in performing academic tasks and communicative activities (Hedge, 2008).

It has been indicated that EFL learners not only have difficulties communicating and interacting fluently and confidently but they also have trouble writing appropriately and organizing reliable paragraphs in their English language programs. To that end, it has been extensively claimed that it is mainly due to learners' lack of collocational competence in English.

The purpose of the present study is to examine the effectiveness of Corpus-Based tools in teaching collocations to Iranian university students with different specialized domains. Having considered the significant role of vocabulary in English language programs, specifically, collocations have gained great emphasis over the last few years.

Collocations are indispensable in second language or foreign language learning contexts. Collocations should be taught not only for competence, but also for accuracy and fluency (Wray, 2002).

Research on collocations has interested plenty of researchers for the last two decades; however, majority of studies done on collocations are corpus-based and mostly descriptive in nature. A handful of researchers did classroom-based research and addressed the question of how to teach collocations most effectively (Nesselhauf, 2003). More empirical and classroom based studies are needed to determine how to teach collocations at different proficiency levels.

However, widespread use of 'corpus linguistics' does not mean that the term or its findings are necessarily fully or widely understood in the context of language pedagogy. In addition, many important developments in the field of corpus linguistics are not always communicated or usefully mediated in terms of their implications for language teaching. This is possibly because corpus linguists are very often not language teachers and spend a lot of time talking with one another rather than with teachers.

Most studies regarding lexical bundles take a corpus-based approach and focus on determining the bundles used in a given discipline and on the variability or the similarity in the type and frequency of the lexical bundles used across different disciplinary fields, registers, genres, and different degrees of writing expertise (Biber et al., 1999).

There are also studies suggesting different ways of introducing students to the use of frequently recurring word combinations. Lewis (2000) introduces many innovative ways of teaching collocations. Pang (2010) also describes several strategies and techniques that will enable second language (L2) students to expand their repertoire of academic rhetorical features to include these multi-word sequences.

Very few studies have dealt with explicit instruction and practical applications of lexical bundles or with students' attitudes toward the helpfulness of these multi-word sequences in their writing ability (Cortes, 2006).

Sinclair (1997) examined the potential impact of computer-processed language data on language teaching and indicated why language teachers should pay attention to developments in corpus linguistics. A corpus approach supports the use of examples of real language in the classroom (as opposed to the invented ones) and corpus data can provide language teachers and learners with illuminating (and often counter-intuitive) guidance as to frequent collocations and other language patterns. Form-meaning links can be taught in order to minimize the learning load and the language learner can use corpus evidence to help develop individual creativity in language use.

In another study, Varley (2009) found that students generally had a positive response to corpus consultation and were able to identify benefits clearly, particularly in the areas of vocabulary acquisition and increased awareness of syntactic patterns. Most of the participants of his study indicated they are likely to use concordancers in the future and this interest is strongest amongst those students who have clear goals for their language learning. All assignments produced by students represented an increased awareness of lexico-grammatical usage, particularly with regard to vocabulary use, phrases and collocational patterns.

Nowadays, more and more researchers have accepted corpus analysis as a way to justify their research, using percentages and frequencies to analyze language use. The importance of corpora analysis and its application to applied linguistics is beyond doubt, as recent studies can confirm (Holmes, 1994; Kourilova, 1996; Ceirano and Rodriguez, 1997; Biber, Conrad and Reppen, 1998; de Monnik, 1998; Martí Guinovart, 1999 ;Oostdijk, 2000; Cortese, 2002; Hornero, Luzón and Murillo, 2006).

The study of collocations, especially with available corpora now on-line like British National Corpus, Collins Cobild Bank of English, Multimodel Learning Corpus Exchange, etc., has increased (Shin & Nation, 2007). Though most of the research done is based on corpora, and remains descriptive except maybe Nesselhauf (2003) corpora study with implications for teaching, like advising teachers to teach all possible combinations including articles and prepositions that form the collocations. Still corpus-based research is also very valuable because researchers have opportunities to study the interlanguage of learners, and the most commonly made mistakes with collocations. Moreover ,investigating native corpora also yields information on most used collocations, and shed light for second language and foreign language instruction on what kind of collocations should be taught to learners.

Zahedi et al. (2010) have emphasized the importance of drawing second language learners' attention to standard multiword expressions such as collocations and idioms. In other studies done in the field of lexical collocation Mounya,(2010) point to the impact of lexical collocation awareness on other specific aspects of general English proficiency.

Plenty of other studies have been conducted concerning the acquisition and use of different types of lexical chunks and collocational phrase. In line with this, In the 1970s and 1980s, studies on the acquisition and use of different types of frequent word combinations and lexical chunks have drawn on a number of research methods and techniques as follows: ethnography (Fillmore, 1979), conversational analysis (Manes & Wolfson, 1981) and quantitative research on the use of multi-word expressions (Altenberg, 1993), among others, just to name a few.

In electronic lexicography or Automatic Term Recognition (ATR), a number of studies have investigated how to multiword terminology from texts (e.g. Collier et al., 2002). Basically, those studies identified candidate patterns of words (e.g. noun-noun or adjective-noun combinations) from texts and used various frequency-based or

association-based measures to determine the term hood of those candidates. Other ATR studies took more sophisticated approaches. Wermter and Hahn (2005), for example, distinguished domain-specific from non-domain-specific multiword term on the basis of paradigmatic modifiability degrees. The assumption behind this approach was that the component words of a multiword term had stronger association strength and thus any component of it was less likely to be substituted by other words. However, although the identification of multiword has been an active field of research, few studies have explored ways of making the terminology accessible to EAP students. To our knowledge, Barriere's (2009) TerminoWeb has been the only way addressing this issue in the literature. Below we describe Barriere's platform.

TerminoWeb, as its name suggests, was created with an aim to help learners of different professional areas explore and learn domain-specific knowledge from the web. To get access to the knowledge, a user had to follow different steps. The starting point was to upload a technical paper to the platform.

Techcollo, furthermore, offers an interface which allows users to compare collocations in two different specialized domains or in a specialized and general-purpose corpus. These convenient search functions will more effectively enable EAP learners to discover and explore specialized collocational knowledge online.

METHODOLOGY

Participants

In regard to settings and contexts, the sampling of the study was conducted in Islamic Azad University, Khorasgan Branch. First, a placement test was given to some students with three specialized domains namely; law, dentistry, and physical education and a total of 150 male and female students were chosen based on their test results with intermediate English proficiency level. All the participants' native language was Persian and they were generally studying English both for specific and academic purposes, to be more precise. Then, participants were divided into two main experimental and control groups each 75 students. In this respect, the experimental group was divided into three law, dentistry, and physical education groups each 25 students and so did the control group.

Pilot Study

The design employed in this study is a mixed research method in which both the qualitative and quantitative research types are exploited. It consists of an experimental research method. Moreover, participants are divided into two main groups; an experimental and a control group. To figure out the reliability and validity of the pretest and posttest a pilot study was done. A group of 10 learners who had roughly the same proficiency level as the main participants of the present study participated in the pilot study. The KR-21 formula was used to determine the reliability of the pretest and posttest. As their reliability scores were higher than 0.75, it was concluded that the tests were reliable. To determine the validity of the contents of pretest and posttest, some

experienced English language teachers checked the pretest and posttest. Therefore, they concluded that the tests are valid and can be administered.

Instrumentation

A solution placement test (Edward, 2007) was applied to determine the homogeneous participants for the study. In sum, this test contains 50 multiple-choice questions which assess students' knowledge of key vocabulary and grammar from elementary to upper intermediate levels, a reading text with 10 graded comprehension questions, and an optional writing task to assess students' ability to produce the language. The 50 multiple choice questions and the reading questions are designed to be done together in a 45-minute session. The writing task can be done in the following session and should take approximately 20 minutes.

Before conducting the main research study a pretest was used to find out students' knowledge about collocations and lexical chunks. To that end, thirty multiple choice items were selected for this purpose. The items were chosen from the book *English collocations in use* (Michael McCarthy & Felicity O'Dell, 2002). In order to make sure about the content validity of the pretest, it was consulted with some experienced professors. Furthermore, the reliability of the pretest was 0.84 (calculated by KR-21 formula) which was applied on a pilot group of 10 students.

After the treatment a posttest was administered to students to understand the instructional efficacy of corpus-based tools in teaching collocations to students with different specialize domains. The posttest comprised 30 multiple choice items and was selected from *English Collocations in Use* book (Michael McCarthy & Felicity O'Dell, 2002). Moreover, in order to determine the effectiveness of the treatment applied in this study, a delayed posttest (as the posttest) was administered three weeks after the posttest. As mentioned in the previous part, in order for the researcher to find out the reliability of the test, KR-21 formula was utilized and the result was 0.82 showing a good score. Additionally, to figure out whether the posttest was valid or not, plenty of experienced and knowledgeable teachers and professors were consulted.

Procedures

The experimental group was taught the collocations and chunks through using corpus-based tools and the control group was taught collocations and chunks through traditional method in which they did not receive any instructional tools and innovative materials and instruments. Corpus-based tools were devised for intermediate and upper intermediate students to assist them in using collocations and lexical chunks more appropriately. Collocations and word partners were brought in a wide variety of different contexts and settings in which vocabularies were learnt in their relations with other words and phrases rather than in isolation. Thousands of collocations are identified and introduced on this program. In line with this, it provided the students with plenty of collocations in bold which were used in texts and exercises that were complementary for the collocations lessons.

Date Analysis

All the data and results gained through pretests, posttests, and delayed posttests were fed into the computer and then analyzed employing SPSS. Technically, both qualitative and quantitative data were analyzed with the help of SPSS program. Participants received 1 point for each correct answer in multiple choice items in the pretests, posttests, and delayed posttests. Descriptive statistics (frequency, mean, and standard deviation) and two-way ANOVA were used for the analysis of the quantitative data. Descriptive statistics were applied to analyze all quantitative data.

RESULTS

Homogeneity of learners with regard to their English proficiency level

In order to ascertain the homogeneity of all groups according to their English proficiency level, a two-way between-groups analysis of variance was run to the scores of participants on placement test. Table 1 shows the results of Levene's Test of Equality of Error Variances for the participants.

Table 1. Levene's Test of Equality of Error Variances for the Placement test

F	df1	df2	Sig.
.435	5	144	.823

Homogeneity of learners in collocational knowledge

In order to establish the homogeneity of groups in terms of collocational knowledge prior to the study, a test comprising 30 multiple-choice items was administered to all groups as the pretest. Then, the performances of the participants on this test were compared and analyzed applying another two-way ANOVA. The sig. value .597 is greater than 0.05 (the critical value). Therefore, it is concluded that all learners were homogenous in their collocational knowledge at the beginning of the treatment.

Table 2. The Results of Levene's Test of Equality of Error Variances for the Pretest

F	df1	df2	Sig.
.738	5	144	.597

Comparing using corpus-based tools and traditional approaches

With regard to means of participants in post-test, all students in experimental groups outperformed the students in control groups. It shows that the significant level calculated for the groups (experimental and control) is .000, which means that treatment was statistically significant ($p < 0.05$). As it is shown in this table the significant level for majors is .039, it means that majors are significant, too. But the interaction between the treatment and major is not significantly influenced the students' performance since it is more than .05.

Table 3. The Results of ANOVA for the Post-test

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	262.300 ^a	5	52.460	5.117	.000	.151
Intercept	57506.460	1	57506.460	5609.474	.000	.975
Group	168.540	1	168.540	16.440	.000	.102
Major	68.040	2	34.020	3.318	.039	.044
Group * Major	25.720	2	12.860	1.254	.288	.017
Error	1476.240	144	10.252			
Total	59245.000	150				
Corrected Total	1738.540	149				

In order to find out the location of the differences, post hoc comparisons of means (Scheffe test) were completed. The results of multiple comparisons and the homogeneous subsets for the post-test scores are illustrated in Tables 4. Inspection of the p-values shows that the scores obtained from dentistry differ significantly ($p < 0.05$) from those of physical education and vice versa.

Table 4. The Results Post hoc for the Post-test

(I) Major	(J) Major	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Law	Dentistry	-.54	.640	.701	-2.12	1.04
	Physical Education	1.08	.640	.245	-.50	2.66
Dentistry	Law	.54	.640	.701	-1.04	2.12
	Physical Education	1.62*	.640	.044	.04	3.20
Physical Education	Law	-1.08	.640	.245	-2.66	.50
	Dentistry	-1.62*	.640	.044	-3.20	-.04

Influence of corpus-based tools in students' learning and retention

The results of tests of between-subject effects are shown in Table 5. The results of the two-way ANOVA show that the p-value for the F ratio of 18.299 is .000, which is much smaller than the level of significance .05 set for this study, and the p-value for the F ratio of 3.862 is .023, which is much less than the level of significance (0.05). Concerning the interaction between the groups (experimental and control) and majors, the analysis run in SPSS shows that a sig is .002 which is less than the cutoff point. Accordingly, the second null hypothesis of the study is rejected and, the results confirm that the application of corpus-based tools influence university students learning and retention of collocation in specialized domains.

Table 5. The Results ANOVA for the Delayed posttest

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	249.473 ^a	5	49.895	7.742	.000	.212
Intercept	49431.527	1	49431.527	7670.409	.000	.982
Group	117.927	1	117.927	18.299	.000	.113
Major	49.773	2	24.887	3.862	.023	.051
Group * Major	81.773	2	40.887	6.344	.002	.081
Error	928.000	144	6.444			
Total	50609.000	150				
Corrected Total	1177.473	149				

According to Tables 5 and 6, a Post hoc comparison using the Scheffe test was carried out in order to locate the exact differences in the performances of the target groups. This test systematically compares each pairs of majors, and indicates that there is a significant difference in the means of law and physical education since the amount of sig. is .049 that is less than .05.

Table 6. The Results of Post hoc test for the Delayed posttest

(I) Major	(J) Major	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Law	Dentistry	.08	.508	.988	-1.18	1.34
	Physical Education	1.26*	.508	.049	.00	2.52
Dentistry	Law	-.08	.508	.988	-1.34	1.18
	Physical Education	1.18	.508	.071	-.08	2.44
Physical Education	Law	-1.26*	.508	.049	-2.52	.00
	Dentistry	-1.18	.508	.071	-2.44	.08

DISCUSSION

The present study investigated the instructional efficacy of corpus-based tools in teaching collocations to Iranian university students with different specialized domains namely; law, dentistry, and physical education. Teaching collocations through using corpus-based tools helped learners to expose themselves to many contextual and real life situations in which collocations were utilized. Furthermore, the application of corpus-based tools, in particular, helps university students to retain and recall collocations better and, consequently, learn them more accurately.

The quantitative data were gathered after posttest and delayed posttest and showed that using corpus-based tools had a significant influence on teaching collocations to university students with different specialized domains in the short term and long term. Thus, the first null hypothesis (using corpus-based tools is not superior to traditional

approaches in teaching collocations to Iranian university students) was declined. Also, the second null hypothesis (the application of corpus-based tools does not influence university students' learning and retention of collocations used in specialized texts) was rejected.

The first null hypothesis, using corpus-based tools is not pedagogically superior to traditional approaches in teaching collocations to Iranian university students with different specialized domains, was declined. The results indicated that experimental group outperformed the control group in the posttest. Scores in the experimental group were higher than scores in the control group.

In other words, teaching collocations through corpus-based tools in the experimental group is superior to traditional approach in the control group. This finding is in agreement with all of the other studies conducted by Barrier's (2009), Wermter and Hahn (2005). They argued that corpus-based tools can inevitably help learners to reinforce collocations and lexical chunks. Moreover, they asserted that teaching collocations through corpus-based tools can raise students' awareness and interest.

The second null hypothesis, the application of corpus-based tools does not influence university students' learning and retention of collocations used in specialized domains, was rejected. The findings stated that students had a positive attitude towards the corpus-based tools and the application of corpus-based tools had a considerable impact on the students' learning and retention of collocations so that they could remember and recall the collocational phrases and lexical chunks in the long term.

This finding is in agreement with the study conducted by Barrier's (2009) that collocations were better remembered employing corpus-based tools. Accordingly, Techcollo offers an interface which allows users to compare collocations in different specialized domains or in a more specialized and general-purpose corpus. These convenient search functions will more effectively enable EAP learners to discover and explore specialized collocational knowledge online. It can be said that by means of experiencing online tools, students acquire and internalize collocational knowledge more quickly and efficiently.

CONCLUSION

The aim of this study was to investigate the instructional efficacy of corpus-based tools in teaching collocations to Iranian university students with different specialized domains. In other words, this study not only attempted to prove that using corpus-based tools is superior to conventional ways in teaching collocations but also endeavored to examine that the application of online tools can influence students' learning and retention of collocations utilized in specialized texts. The results of this research indicated that learners in the experimental group achieved significantly higher scores in the posttest and delayed posttest than those in the control group. These findings were related to effect of online tools in learning as mentioned by Barrier's (2009), due to more exposure to collocations, corpus can increase comprehension and

cognitive retention of collocations and when there is an interdisciplinary relationship between course materials and collocations, it can make information more memorable.

Corpus and online tools are perceived as important elements for learning process and therefore teachers should consider the use of them in language teaching and in their language classrooms. Corpus can be a powerful stimulus in language teaching to motivate students to participate in the classrooms' tasks. By application of corpus-based tools in the class, teachers can provide a combination of learning, retention, awareness, and interesting environment. Students could learn much from online corpus-based tools. Learning another language means how to communicate properly in that language because language is used in authentic and real life situations. On one hand, learning word partners and lexical chunks plays a vital role in language communication. On the other hand, collocations are integral parts of every language and they are used in everyday speech. Consequently, online tools could be used as a stimulus in teaching collocations. In sum, corpus is a powerful factor in retaining and learning collocations and has the potential to be used in ways that can make a positive contribution to classroom language learning. Since collocations are practiced in a wide variety of different contexts, they can be retained and recalled easily. The more collocations are repeated and exposed, the better they are learnt.

REFERENCES

- Barriere, C. (2009). *Finding Domain Specific Collocations and Concordances on the Web*. Proceedings of the Workshop on Natural Language Processing Methods and Corpora in Translation, Lexicography, and Language Learning.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam, the Netherlands: John Benjamins.
- Biber, D. & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, 26, 263-286.
- Collier, N., Nobata, Ch., & Tsujii, J. (2002). Automatic Acquisition and Classification of Terminology Using a Tagged Corpus in the Molecular Biology Domain. *Terminology*, 7(2), 239-257.
- Cortes, V. (2006). Teaching lexical bundles in the disciplines: An example from a writing intensive history class. *Linguistics and Education*, 17, 391-406.
- Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly*, 34(2), 213-238.
- Hedge, T. (2008). *Teaching and Learning in Language Classroom*. Oxford: Oxford University Press.
- Lewis, M. (2000). Learning in the lexical approach. In Lewis, M. (Ed.), *Teaching collocation. Further developments in the lexical approach* (pp. 155-185). Hove: LTP.
- Lewis, M (2006). *Teaching Collocation. Further Development in Lexical Approach*. Hove: Language Teaching Publications.

- Lewis, M. (2000). Language in the lexical approach. In Lewis, M. *Teaching collocation: Further developments in the lexical approach*. (pp. 126-154) Hove: Language Teaching Publications.
- McCarthy, M. (1988). *Vocabulary*. Oxford: Oxford University Press.
- Nash, H. & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language and Communication Disorders, 41(3)*, 335-354.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press, Cambridge, UK.
- Pang, W. (2010). Lexical bundles and the construction of an academic voice: A pedagogical perspective. *Asian EFL Journal, 47*, 1-13.
- Richard, J. C. & Rodgers (2003). *Approaches and Methods in Language Teaching*. Cambridge: Cambridge University Press.
- Sinclair, J. (1997). *Corpus, concordance, collocation*. Oxford. Oxford University Press.
- Varley, S. (2009). I'll just look that up in the concordancer: integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning, 22(2)*, 133.
- Wermter, J. & Hahn, U. (2005). *Paradigmatic Modifiability Statistics for the Extraction of Complex Multi-word Terms*. Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. Association for Computational Linguistics.