**cogent**
education

CrossMark

STUDENT LEARNING, CHILDHOOD & VOICES | RESEARCH ARTICLE

# The effect of data-driven approach to teaching vocabulary on Iranian students' learning of English vocabulary

Elyas Barabadi[1] and Yaser Khajavi[2]*

*Corresponding author: Yaser Khajavi, TEFL, Farhangian University, Khuzestan, Iran
E-mails: yaserkhajavi@gmail.com, yaserkhajavi@yahoo.com

**Abstract:** Corpus-based data-driven learning (DDL) is an innovation in teaching and learning new vocabulary for EFL students. Using teacher-prepared materials obtained from COCA corpus, the goal of the present study is to compare DDL and traditional methods of teaching vocabulary like consultation of dictionary or a grammar book. As such, two intact classes ($N = 42$) and one intact class ($N = 20$) who were studying English for Certificate for Advanced English (CAE) comprised the experimental and control group, respectively. Nation's standardized vocabulary size test was administered as a pre-test to make sure that the participants were at the same level of vocabulary knowledge. During the semester, the learners were exposed to teacher-prepared corpus-based materials from COCA such as display list, synonym, keywords in context (KWICs), and collocate. Moreover, they were asked to do similar searches on their own as homework. A post-test based on the reading passages of their course book (Ready for CAE) was designed and administered at the end of the semester. The results of the study indicated that the learners in the experimental group outperformed their counterparts in the control group. The better performance of the former group can be attributed to the fact that the learners could take a more active role in the learning process in which self-discovery, inductive, and bottom-up processes were emphasized.

## ABOUT THE AUTHORS

Elyas Barabadi is an assistant professor in TEFL at the University of Bojnord, Iran. He is interested in psychology of language learning and teacher education.

Yaser Khajavi holds a PhD in TEFL from Shiraz University, Iran. He is interested in cognitive second language acquisition (SLA) and writing pedagogy.

Elyas Barabadi

## PUBLIC INTEREST STATEMENT

Although vocabulary plays an important role in learning a second language, it has not received careful attention in some Asian countries (Catalan, 2003; Fan, 2003). Learning and teaching vocabulary largely is incidental in a sense that teachers do not approach teaching vocabulary in a systematic way; instead, whenever learners are confronted with some difficult words, they are provided with the dictionary definition of those words by their instructors. In spite of the important role that vocabulary can play in helping language students learn a language, there seems not to be a systematic approach to teaching vocabulary, like DDL, with due attention to corpus linguistics. This motivated the researchers to carry out the present study in order to examine the effectiveness of DDL approach to teaching vocabulary for Iranian EFL students with moderate proficiency.

**cogent** · oa

## 1. Introduction

Today, the critical importance of vocabulary learning in developing second language learning is appreciated by all L2 teachers and researchers. No language teacher or learner contests the importance of vocabulary and the lexical dimension in learning that language. Indeed, second language (L2) acquisition depends to a large extent on the development of a strong vocabulary (Schmitt, 2000; Singleton, 1999).

Recent changes in the conceptualization of language learning have given rise to the importance of learning vocabulary. More specifically, the emergence and recognition of "bottom-up processing" skills focusing on learning lexicon in language learning instead of the "top-down processing" skills focusing on grammar learning has acknowledged the key role of vocabulary in developing a second language (Ellis, 1997; Nattinger & DeCarrico, 1992). Lewis (1993, 1997) notes that it is more appropriate to say "grammaticalised lexis," not "lexicalized grammar." Likewise, other scholars (Meara, 2002; Nation, 2001; Schmitt, 2000) suggest that vocabulary be systematically integrated into any course. However, teachers often overestimate their students' vocabulary size, and therefore fail to teach at a level of comprehensible input (Folse, 2004). Knowing a word is no longer limited to the definition of that word; rather a comprehensive understanding of a word includes meta-linguistic awareness about that word (Nation, 2001). Thus, in addition to the dictionary definition of a particular word, learners are expected to know other aspects of that word like spelling, morphology, parts of speech, pronunciation, variant meanings, collocations, specific uses, and register-related contexts of use (Koda, 2000; Nation, 2001).

Due to the nature of corpora, vocabulary learning and corpus analysis are closely related to each other (Read, 2010). Some scholars (Nation, 2001; Schmitt, 2000) hold the belief that vocabulary learning and vocabulary instruction have been extensively affected by corpus linguistics. Corpus-based materials and activities are just one sample of data-driven learning (DDL) as put forth by Johns (1991). Generally speaking, DDL refers to an approach to learning vocabulary that presents learners with linguistic data and has them find the rules and patterns from the examples. The use of linguistic data including online corpora, teacher-prepared written corpora, use of dictionaries and other forms of DDL allows the learners to become familiar with various aspects of vocabulary such as grammar, idioms, and other phrases, and knowing what a word means. Thus, the use of DDL and corpora in teaching vocabulary seems vital. DDL methods in the classroom allow learners to learn some of these issues in an effective way. As mentioned earlier, some aspects of meta-linguistic awareness, like register and collocation, lend themselves easily to instruction via corpora, while some other aspects like parts of speech and specific uses of words can be learned through other sources such as dictionaries. In sum, the use of DDL approach to learning vocabulary seems to be effective in gaining a full understanding of the lexicon of the second language; an understanding that goes beyond the dictionary definition of words by adding some aspects of meta-linguistics awareness about that word.

Although vocabulary plays an important role in learning a second language, it has not received careful attention in some Asian countries (Catalan, 2003; Fan, 2003). Learning and teaching vocabulary to a large extent is incidental in a sense that teachers do not approach teaching vocabulary in a systematic way; instead, whenever, learners are confronted with some difficult words, they are provided with the dictionary definition of those words by their instructors. Likewise, vocabulary teaching in Iranian English classrooms is largely incidental (Kafipour, Yazdi, Soori, & Shokrpour, 2011). In spite of the important role, that vocabulary can play in helping language students learn a language, there seems not to be a systematic approach to teaching vocabulary, like DDL with due attention to corpus linguistics. This motivated the researcher to carry out the present study in order to examine the effectiveness of DDL approach to teaching vocabulary for Iranian EFL students with moderate

proficiency. Although there have been some studies in recent years examining vocabulary teaching and learning from different aspects (Hamzah, Kafipour, & Abdullah, 2009; Kafipour et al., 2011), few if any study have investigated the role of DDL approach and specifically the role of corpora in developing second language vocabulary in Iran. To contribute to the related literature, this study aims at probing the effect of DDL on vocabulary learning of Iranian learners.

## 2. Data-driven learning and corpora (the theoretical background)

Language education is continually influenced by corpus linguistics at a rapid pace (O'Keeffe, McCarthy, & Carter, 2007; Sinclair, 2004). In recent years, language teaching has begun to benefit from the applications of corpus linguistics. According to Romer (2009), "corpus linguistics can make a difference for language learning and teaching and that it has an immense potential to improve pedagogical practice" (p. 84). Bennett (2010) believes that corpora can be used in language teaching in three ways: corpus-influenced materials, corpus-cited texts, and corpus-designed activities. Those materials which are developed based on the patterns and frequency information obtained from corpora are referred to as corpus-influenced materials. Those resource books and materials like dictionaries and grammar books are called corpus-cited texts. Corpus-designed activities involve those in which the learner takes on an active role by exploring and analyzing the data in order to induce patterns from the corpora. Indeed, learners get involved in DDL using the activities designed based on corpora.

The notion of "data driven learning (DDL)" was first introduced by Johns (1990) to describe how language learners themselves could explore the language and discover some rules and regularities for themselves. Indeed, language learners are viewed as detectives. The main characteristic of this approach is that the learners themselves have an active role in teasing out the grammatical patterns for themselves after they have been exposed to samples of authentic language (Hadley, 2002). According to Johns (1991), language learner act like a researcher who attempts to analyze the target language data to which he/she is exposed. In fact, language learners should have access to linguistic data in order to directly get involved in examining it. The result of such analysis is that the learner eventually discovers the underlying rules of a linguistic system in an inductive way by getting familiar with the language through the regularities and consistencies encountered.

Currently, DDL is not viewed in a purely discovery-based approach to learning as originally proposed by Johns (1991). Instead, many researchers (Basanta & Martin, 2007; Boulton, 2010; Clifton & Phillips, 2006; Hadley, 2002) acknowledge the usefulness of teacher-guided searches of known rules with other corpus features. Indeed, one of the three common ways for teachers to provide learners with hands-on corpus activities is the use of teacher-prepared corpus material (Reppen, 2011). Instead of asking students to explore the linguistic data (corpus), the teachers will do the exploration and will bring the results into the classroom. Then, the students are expected to analyze and examine the teacher-prepared material. Reppen (2011) argues that the use of such materials better guarantees the appropriateness of the content for the learners in terms of difficulty level. Additionally, it should be noted that DDL is no longer limited to only concordance lines, or incomplete sentences that focus on a common word, or even to corpus in general (Boulton, 2009; Davies, 2008). In his discussion of what constitutes data-driven learning, Boulton (2011) notes that DDL is "not an all-or-nothing affair: its boundaries are fuzzy, and any identifiable cut-off point will necessarily be arbitrary" (p. 575). The point here is that DDL should not be limited to only Corpus-based learning; rather whatever source that constitutes language data regardless of whether it is online corpus, written corpus, dictionaries and even the internet and search engines such as Google can be used as framework for DDL.

Although the use of corpora in language teaching has been extensively advocated (Sinclair, 2004), they have not been used by language teachers and learners in the classroom (Aijmer, 2009). According to McCarthy (2008), the language teachers' unfamiliarity with the corpus and their lack of

awareness on how to develop and design activities based on the corpora has been the main reason for being neglected in the classroom.

## 3. DDL and vocabulary instruction

The uses and the benefits of DDL for teaching and learning vocabulary have been studied and confirmed by many researchers in recent years (Boulton, 2010, 2011; Cobb, 1999, 2007; Jafarpour & Kousha, 2006; Pickard, 1994; Stevens, 1991). These studies have provided teachers with new ideas and insights for creating activities and tasks based on various corpora.

The widespread use of the Internet, especially Googling the Internet, is a good example of DDL by many learners around the world. Boulton (2011) believes that although this type of DDL is encouraged by many teachers, it remains invisible in the DDL research literature. He further argues that for DDL to reach to a wider audience, it needs not be viewed as a radical or revolutionary practice. Instead, by viewing DDL as an ordinary practice, like the process of Googling the Internet, it will be attracted by larger audiences. Boulton' (2011) comments about DDL were mentioned at the beginning of the literature review to make it clear that DDL is frequently used by language learners around the world. Yet, this frequent use of DDL remains invisible in the literature of DDL.

Stevens (1991) found that vocabulary exercises based on concordance lines are considered easier for learners than traditional gap-filler exercises. Thus, such concordance-based activities should be used "if the purpose of the exercise is to reinforce the vocabulary, as opposed to testing, and if the proclivity of the teacher is to engender a sense of confidence and well-being in the students" (p. 55). Cobb's (1997) study compared a corpus-based approach to teaching vocabulary (DDL) with the traditional approach to teaching vocabulary. In the former approach, the learners were presented with multiple concordance lines, while the latter group was presented only a single sentence accompanied by a short definition of the word. The results of the study showed that the concordance-based approach was more effective. Indeed, viewing concordance lines facilitated the acquisition of transferable word knowledge, supported by the fact that these students were able to apply their knowledge of the word in novel activities and contexts. Jafarpour and Kousha (2006) in their study investigated whether concordancing materials presented through DDL approach affected how students learn collocations of prepositions. The participants in the experimental group in their study were taught through the DDL approach that was based on concordancing, while the control group underwent a conventional-based treatment on prepositions and their collocational patterns. The results of their study indicated that the DDL approach was more effective in teaching and learning collocations of prepositions.

Huei Lin (2016) used a blended approach to assess the pedagogical appropriateness of DDL in Taiwan's EFL grammar classrooms. The study also investigated the effects of DDL compared with that of a traditional deductive approach on the learning motivation and self-efficacy of first-year EFL students. In addition, it examined a group of teachers' hands-on experience of teaching DDL to these students. The findings showed that the students who received DDL treatment improved their learning attitudes. However, qualitative part of the study revealed that, despite technical problems and increased workload, teachers found their DDL teaching experience innovative and interesting, believed in its effectiveness in grammar learning, which caused to transform Taiwanese students' grammar learning patterns from passivity to active engagement.

### 3.1. Research aims

The aim of the present study is to compare DDL approach and traditional methods of teaching vocabulary like consultation of dictionary or a grammar book.

cogent ·· education

## 4. Methods of the study

### 4.1. Participants

Sixty-two CAE (Certificate for Advanced English) students served as the participants of this study. They were all Persian native speakers, from three intact CAE-level classes at two English institutes in Mashhad, Iran. Twenty-eight participants were female and the rest were male. The age of the participants ranged from 17 to 26. In this study, neither gender nor age was a variable. Since language learners were assigned by the coordinators of the institutes, it was practically impossible to disregard their schedule. However, to contain the extraneous variables and selection bias, the three intact classes were randomly assigned to two treatment groups and one control group. Forty-two participants were assigned to treatment group. It should be noted that students in the treatment groups were in two classes in one language institute. As for the control group, there were 20 students who were in the same level of English proficiency but in another institute. The researcher of the current study taught these classes. In order to determine the participants' level of vocabulary knowledge, Nation's (2001) Vocabulary Size Test was administered.

### 4.2. Instruments

As mentioned earlier, Nation's standardized vocabulary size test was administered in order to assess the level of proficiency of the students for this study. This test includes 140 items, and can determine the learners' vocabulary size in a range of 0–14,000. Each item in the test has a score value of 100. Having confirmed the fact that the participants possessed almost the same level of vocabulary knowledge, the researchers used a corpus-based approach in the experimental group. More specifically, Longman corpus of examples, Longman corpus of collocations, and Longman language activator were used so that the participants could learn the right collocations, prepositions, and the differences between similar words.

As mentioned earlier, this research aimed to find out whether a corpus-based approach toward vocabulary is more successful than a traditional one in which students only come across the new words in the context. For this reason, a vocabulary test, consisting of the key words that the learners in both experimental and control group had encountered in their textbook, was administered to both the experimental and the control group. The test consisted of 14 multiple-choice items and 16 matching items. The items incorporated knowledge of collocation as well as synonyms of those words that were taught during the 7-week treatment period. Indeed, the words were chosen from the course book (*Interchange 3*) that the students were taught in during the semester. Attempts were made to make the content of the test as representative of the content of the textbook as possible. The reliability of the test was estimated to be 0.74 using Kr-21 formula. After the 7-week treatment, the test was given to the participants in both experimental and control group.

### 4.3. Procedure

The purpose of this study was to compare a corpus-based approach toward teaching vocabulary to a common communicative one at the C1 level. Accordingly, while corpus-based approach was used for teaching vocabulary for language learners in the experimental classes, the control class was taught using traditional method in which either the dictionary definition of the words were sought out or the students were required to infer the meaning of new words from the context. At the beginning of the semester, Vocabulary Size Test, made by the Victorian University of Wellington, was administered to all participants; and the results showed that the three classes had relatively the same level (C1) in English.

For the sake of having a corpus-based approach, in the first session of the course, the researcher introduced COCA corpus to the students in the treatment group. Particularly, three important features of COCA corpus were emphasized; namely, Word List including synonym, collocations and KWICs (keywords in context). In the beginning of each class, the students in the experimental group were given the output of the first 10 hits of the target words in the list display of the COCA. Attempt was made to include those concordance lines that were relevant in terms of part of speech, collocation,

cogent ·· education

| Table 1. List display of the word "erosion" from COCA output |
| --- |
| : Compost enables the soil to do a better job of retaining water and preventing **erosion**. This characteristic is especially valuable in sandy soils where rapid water loss is common |
| of fresh foliage. # Cover! Empty garden beds are an invitation to soil **erosion** and weeds. Cover the surface with an overwintering, soil-building cover crop like annual |
| of the things that our plants do: produce oxygen, build topsoil, prevent **erosion** and flooding, sequester carbon dioxide, buffer extreme weather, clean our water and |
| OUR PRETTY GARDEN perennials may be the key to protecting America's precious topsoil from **erosion**. # Before Europeans came to North America, vast parts of the Midwest were |
| Prairie Strips (STRIPS). And some of the most effective agents for preventing **erosion** are probably growing in our flower gardens right now. # I got into an |
| if just 10% of cropland is planted with native prairie perennials and grasses, **erosion** is reduced by 95% compared with fields planted entirely with row crops. To |
| together even as they make channels for rainwater to enter. # Reduction of soil **erosion** isn't the only benefit of planting prairie strips, though. # In the |
| absurd. # But beyond that, we don't subtract externalized costs. Soil **erosion**, polluted water and nutrient-deficient food never show up on the negative side of our |
| bare dirt is unhealthy. Removing the soil's vegetative cover makes it vulnerable to **erosion** and shuts down biological activity. That activity is important because the work of that |
| the environment, it means more carbon in the atmosphere, more floods, more **erosion**, more dying streams and lakes, more cruelty. Push that number to |

and clarity of the sentence in rendering the meaning of the target word. Having gone through these example sentences for each target item, the learners were asked to make guesses about the possible meanings and part of speech of each word. For example, the output for the word "erosion" taken from chapter 2 (times change; A reading passage about the Great Walls) is shown in Table 1.

In order to make the learning process more effective, students were required to discuss each target word in small groups. Particularly, they were asked to discuss the following questions among themselves:

(1) Identify the part of speech (noun, adjective, verb, adverb, prepositions, etc.) of the bolded word in the following concordance lines?

(2) What do you guess about the meaning of the target word?

(3) What words does the target word go together, or what are the collocates of the target word?

(4) Write one original sentence for the target word?

Similarly, the results of synonym searches in COCA were brought to the classroom by the teacher in the form of teacher-prepared material. Synonym searches were used at this stage in order to enable learners to discriminate between nearly synonymous words. Due to its built-in thesaurus, COCA makes it possible to perform synonym searches easily. As a result, the students themselves were asked to enter the target words into a query box at the top of the screen and insert an equal sign before the word (e.g. =intrigue) as homework. Delving into synonym searches in COCA as a home assignment enables language learners to grasp the multiple meanings of words. For example, the learners in the experimental group were given the following paragraph with some bolded words to be replaced by a synonym.

In the centuries following its **abandonment** around 400 AD, its stones were used by local people to build houses, walls, and even churches. Nevertheless, **spectacular** stretches of the wall remain and a number of forts and museums along its length can be visited, providing a **fascinating** glimpse into the lives of the Roman soldiers who **patrolled** it. Although, built of stone, the wall itself is **vulnerable** to **erosion** and visitors are discouraged from walking on it. **Designated** a UNESCO World Heritage in 1987, Hadrian's wall ranks alongside some of the more famous architectural treasures in the world.

| Table 2. Synonyms and example sentences from COCA | | | |
|---|---|---|---|
| **Words (bolded in the paragraph)** | **Synonym (learners' own guess)** | **Synonym from COCA** | **Example sentence** |
| Abandonment | | | |
| Spectacular | | | |
| Fascinating | | | |
| Patrolled | | | |
| Vulnerable | | | |
| Erosion | | | |
| Designated | | | |

Having found the synonym for each word, students were asked to take a close look at the sample sentences for that synonym in order to make sure that the best synonym has been chosen for each word (see Table 2).

However, in order to have consistency during the class hour and also to have some materials as a frame of reference, teacher-prepared materials were used in the classroom, and the students' home searches were only used as a preparatory activity for class activities. Having gone through this exercise, once at home and once in the classroom, helped learners discriminate between almost synonym words and hence increase their semantic sensitivity of target words.

During the class time, first, the learners were asked for their ideas of synonyms in order to encourage their own guesses and intuitions of each word. Then, using a data projector connected to a computer, the teacher demonstrated to the learners how to find the best synonym for each of the bolded words in the paragraph. For example, the word "designate" was put into the corpus with the formula of (=designate) and the output of the corpus yielded 19 synonyms for the word. Because the corpus is not semantically sensitive, synonyms for all meanings of a word were given by the corpus and the learners needed guidance as how to choose the right synonym for the given word. Here, the five hits out of 19 for the word "designate" is indicated:

(1) Call
(2) Define
(3) Label
(4) Choose
(5) Assign

First based on their guesses and intuitions, the learners were asked to choose the right word. Most of the learners chose the word "Assign." With the help and guidance of their teacher, the class discussed the possible matches and came to the conclusion that "Assign" cannot be a good synonym for "Designate" and they agreed on the word "Choose."

Aside from working out the meaning of target words through list display and synonym section of the corpus, several words from each lesson were chosen and students were asked to check up the collocations of those specific word using the collocation and KWIC sections of the COCA corpus. For example, from the above paragraph, the word "Vulnerable" was put into the query box of collocate and then the option "prep.All" was selected so that all the prepositions that collocate with the word "Vulnerable" would appear. Here, my especial emphasis was on learning appropriate prepositions for the target word so the Command (prep.All) was selected. For other words in other lessons, other parts of speech were selected depending on the textual context and my own understanding of the learners' struggle with some words and their collocations. The following are the first five prepositions that collocate with the word "Vulnerable."

**cogent** education

| Table 3. The most frequent collocate of the word "vulnerable" |
|---|
| British epidemiologist named David Barker found that he could predict which populations would be most **vulnerable to** heart disease in middle age by looking at rates of infant mortality and low |
| rates STDs are most widespread among African-Americans, and can make your immune system more **vulnerable to** HIV. # High frequency of HIV Because HIV rates are greater for Black people |
| # HOLE 13 # " Azalea " # PAR 5 # " This hole is **vulnerable to** birdies and eagles, so you feel you have **to** make 4 just **to** |
| . Francis was taken **to** Spain as the emperor's prisoner. France was left **vulnerable to** its enemies and **to** internal dissent, which Louise de Savoie, as her |
| habit in an alternating pattern. # best practice for planting blueberries # Blueberries are **vulnerable to** numerous pests and diseases, which can cause a blueberry hedge **to** b |

| Table 4. Concordance output (KWICs) for the word "patrol" as a noun and as a verb |
|---|
| Oil was stealing into Mobile Bay, He did not **patrol** on Sundays. In the morning, he came down to |
| A deputy in a four-wheel-drive vehicle was sent to **patrol** the area but saw no sign of the raft or the Stadlers |
| The policemen that they were members of the civilian air **patrol** and attorneys. The officer then |
| The night at his apartment. A siren sounds. A **patrol** car passes them on the shoulder |
| High up in the Sierra Nevada, A civil air **patrol** pilot had given me the GPS coordinates of an |
| controversial question: Just how many officers are needed to **patrol** the District adequately 24 h a day ? # What was uncovered |
| 4736 hundreds of migrants nabbed by the border **patrol** after illegally crossing the US-Mexico |
| have the most influence, or the money to fence or **patrol** it, no matter what the documents say. Foreign governments and |

(1) To
(2) Among
(3) During
(4) Because
(5) Against

Moreover, in order to enhance their learning of specific collocation, five sample sentences of the most frequent collocation (e.g. *To*) were presented to the learners as follows (see Table 3).

The researchers obtained the concordance output of the target words from the online corpus of COCA. These concordance outputs which are called KWICs can be used for different activities by the teacher. Table 4 presents the concordance lines obtained from COCA corpus demonstrating the word "patrol" both as a noun and as a verb. In this exercise, the learners were asked to differentiate between two parts of speech of the target word.

## 5. Results

As indicated in Tables 5 and 6, an independent sample *t*-test was conducted to compare the pre-test scores for experimental and control group on vocabulary size test. The results of *t*-test indicated that there was no significant difference in scores on this vocabulary test for experimental group (*M* = 4,547, SD = 468.64) and control group (*M* = 4,575, SD = 359.64); $t(60) = 0.231$, $p = 0.81$ two-tailed). The magnitude of the differences in the means (mean difference = −27.38, 95% CI: −264.90 to 210.14) was very small (eta squared =).

The results of the vocabulary size test indicated that the receptive vocabulary knowledge of the two groups for reading comprehension is the same. As indicated in Table 5, the mean score of the control group is a little higher. However, it should be noted that this difference is not statistically significant as indicated by independent sample *t*-test. Making sure that the two groups (though in

cogent **education**

| Table 5. Group statistics for vocabulary size test | | | | | |
|---|---|---|---|---|---|
| | **Groups** | **N** | **Mean** | **Std. deviation** | **Std. error mean** |
| Pre-test | Experimental | 42 | 4,547.6190 | 468.64522 | 72.31353 |
| | Control | 20 | 4,575.0000 | 359.64163 | 80.41831 |

| Table 6. Independent samples test for vocabulary size test of the two groups as pre-test | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Levene's test for equality of variances** | | **t-test for equality of means** | | | | | |
| | | **F** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean difference** | **Std. error difference** | **95% confidence interval of the difference** |
| | | | | | | | | | **Lower** / **Upper** |
| Pre-test | Equal variances assumed | 1.413 | 0.239 | −0.231 | 60 | 0.818 | −27.38095 | 118.74511 | −264.90654 / 210.14464 |

| Table 7. Group statistics | | | | | |
|---|---|---|---|---|---|
| | **Groups** | **N** | **Mean** | **Std. deviation** | **Std. error mean** |
| Post-total | Experimental | 42 | 25.9524 | 2.16347 | 0.33383 |
| | Control | 20 | 20.8500 | 2.00722 | 0.44883 |

| Table 8. Independent samples test | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Levene's test for equality of variances** | | **t-test for equality of means** | | | | | |
| | | **F** | **Sig.** | **t** | **df** | **Sig. (2-tailed)** | **Mean difference** | **Std. error difference** | **95% confidence interval of the difference** |
| | | | | | | | | | **Lower** / **Upper** |
| Post-Total | Equal variances assumed | 0.002 | 0.961 | 8.879 | 60 | 0.000 | 5.10238 | 0.57467 | 3.95288 / 6.25189 |

three classes: two experimental groups and one control group) have the same level of vocabulary knowledge, the teacher introduced one common textbook for three classes (*Ready for CAE*) as their course book during the semester. It was based on the reading passages of this textbook that the final test was designed and administered to all the participants in the study as the post-test. The results of the post-test were also subjected to independent sample *t*-test. The results of the study indicated that there was a significant difference in scores for learners in the experimental group ($M = 25.95$, SD = 2.16) and control group (20.85, SD = 2.00; $t(60) = 8.87$, $p = 0.000$, two-tailed). For a graphical representation of the results of the post-test, Tables 7 and 8 are presented.

## 6. Discussion

The results of the study are confirmed by other studies (Boulton, 2008; Chambers, 2005) which found that corpus-based data-driven approach to teaching and learning vocabulary is more effective than traditional methods like consultation of a course book or reference book such as a dictionary or grammar book. One possible reason why learners in the experimental group outperformed learners in the control group may be the fact that the former group could establish the connection between form and meaning as they analyzed concordance lines of the target words in terms of part of speech (Chambers, 2005). Similarly, it can be argued that the learners in the experimental group had a

chance to see different forms of the word and notice different senses of the target word. As Boulton (2008) argues, the corpus-based method leads to more profound leaning in which more concrete examples of the target words are given in the context, and this highlights the usage of the word in a given context. In other words, the learners are exposed to several authentic contexts for a word at one time. These multiple contexts in turn increase the likelihood of guessing the meaning of new words by the learners. All these presumed advantages of DDL facilitate the process of autonomous language learning and hence increase the joy and interest of language learning.

The data-driven activities used in this study can be characterized as one strand of communicative language classroom; namely, language or form-focused instruction (Nation, 2001). The other strands are comprehensible meaning-focused input, meaning-focused output, and fluency development. Data-driven instruction can be complementary to other activities in a communicative classroom in which comprehensible input, meaningful output, and fluency development are emphasized. In other words, inductive and exploratory teaching activities for new vocabulary should be supplemented with meaning-focused and fluency-building tasks in which learners attempt to reach a communicative outcome. The notable contribution of corpus-based DDL activities, especially when monitored and guided by the teacher is that they facilitate language learners' "noticing" (Schmidt, 2001) of some features of the target words in the context that traditional methods fail (Willis, 2011).

As suggested by other researchers (Guan, 2013; Willis, 2011), corpus-based data-driven approach brings about a dramatic shift in teachers and students' roles. The latter group is encouraged to play a more active and enthusiastic role by studying concordance lines and discovering language rules for themselves with proper guidance of their teacher. As for language teachers, they become more of an organizer and facilitator of the learning process than transmitter of knowledge. Additionally, this student-centered exploratory learning mode which is initiated by corpus-based data-driven approach can also enhance learners' self-confidence and interest in English by letting them to take more charge of their own learning. One important implication of the study lies in the effect that this approach can have on strengthening the learners' autonomy. In fact, when students refer to the corpuses of vocabulary, they can understand the context of language. This will lead to a sense of self-sufficiency and independence, which consequently leads to improving their autonomy. Another implication is related to the effect of technology particularly computers which need to be considered in deepening learning vocabulary by language learners. Language teachers can teach students how to learn vocabulary via the Internet and computers which can be of interest to new generation who are digital natives.

As mentioned before, each session students were provided with some teacher-prepared concordance lines in order to understand the new vocabulary in context. Having been introduced to the new vocabulary in this way, the learners were required to search COCA corpus at home to seek and explore more examples of the words in context. Working with corpus once in the classroom and once at home allowed language learners to gain greater exposure to various senses of the target word, and hence helped the learners in the experimental group to outperform the learners in the control group. The impetus for using teacher-prepared materials was that the learners were more likely to be exposed to the new vocabulary in such a way that was meaningful and relevant to them. As indicated in the previous section, out of many examples of concordance lines for a particular word like "patrol," only a handful of them that were considered appropriate for the learners were chosen. Attempt was made to bring in examples that would illustrate different parts of speech of the target words. As Reppen (2011) argues, picking out certain concordance lines does not reduce the authenticity of the materials; rather, it only makes the authentic materials more relevant and meaningful for the learners. Specifically, KWICs of a target word like "patrol" initiated lively discussion among students concerning what parts of speech that particular word belonged to.

By introducing the learners to the output lines of KWICs in the form of teacher-prepared materials in the classroom and helping them to discover patterns of language use, the learners did not show much difficulty working with online corpus (COCA) on their own. Moreover, engaging with materials

in this manner is more likely to help learners develop analytical skills and those processes needed for discovering patterns of language use. Once students get equipped with such skills and processes, they are more likely to become autonomous language learners. This autonomy is to some extent the result of many options that are provided for learners when interacting with corpus on the web. Back to our paragraph example in the previous section, the learners were required to obtain the bar graph for the target word "designate" in order to raise their awareness of the differences of language use in speech and writing.

It is worth noting that because of the dual role (teacher and researcher), we are aware of the potential for bias and assure the readers that this was not the case. Experimental and control groups both have received the same "quality" instruction; however, the DDL approach is superior because of the merits of the approach and not because of differences in the quality of instruction.

## 7. Conclusion

Corpus-based data driven approach is an innovation in language teaching which has the potential to make language learners "language investigators in their own right" (Willis, 2011, p. 51). This approach can provide a rich source of authentic instances of language use, and creates a learner-centered learning environment in which they are encouraged to discover and internalize regularities and patterns of language use. Additionally, this learner-centered approach has two main features: first, learners become language researchers on their own; and second, a lot of discussion is generated in the classroom as they report the findings of their own corpus searches to the classroom. With the same token, the teacher takes a more active and critical role by becoming a facilitator who keeps monitoring and guiding students self-discovery of language rules. From a psycholinguistic perspective, data-driven method helps language learners "notice" (Schmidt, 2001) the target language features, and hence raises their awareness of those features. Future studies may investigate the psychological factors at work when using corpus databases. For example, it is worthwhile to see how this approach can influence self-efficacy of language learners. In addition, the way that using corpus databases can affect teachers' teaching is of importance. Another area of research would be investigation of the extent to which learners' autonomy can be affected. Finally, learners' attitudes toward using DDL approach in learning vocabulary can be used in helping teachers with how to apply this approach in their teaching effectively.

## Author details
Elyas Barabadi[1]
E-mail: elyas.ba1364@gmail.com
Yaser Khajavi[2]
E-mails:yaserkhajavi@gmail.com, yaserkhajavi@yahoo.com
[1] TEFL, University of Bojnord, Bojnord, Iran.
[2] TEFL, Farhangian University, khuzestan, Iran.

## References
Aijmer, K. E. (2009). *Corpora and language teaching*. Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl
Basanta, C. P., & Martin, M. E. R. (2007). The application of data-driven learning to a small scale corpus: Using film transcripts for teaching conversational skills. *Language and Computers: Studies in Practical Linguistics, 61*, 141–158.
Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor: University of Michigan Press. http://dx.doi.org/10.3998/mpub.371534
Boulton, A. (2008). DDL: Reaching the parts other teaching can't reach? In A. Frankenburg-Garcia (Ed.), *Proceedings of the 8th Teaching and Language Corpora Conference (TaLc 8)* (pp. 28–44). Amsterdam: Rodopi.
Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics, 35*, 81–106.
Boulton, A. (2010). Data-driven learning: On paper, in practice. In T. Harris, & M. Moreno Jaén (Eds.), *Corpus linguistics in language teaching* (pp. 17–52). Bern: Peter Lang.
Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Gozdz-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563–580). Frankfurt: Peter Lang.
Catalan, R. (2003). Sex differences in L2 vocabulary learning strategies. *International Journal of Applied Linguistics, 13*, 54–77. http://dx.doi.org/10.1111/ijal.2003.13.issue-1
Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology, 9*, 111–125.

Clifton, J., & Phillips, D. (2006). Ensuring high surrender value for corporate clients and increasing the authority of the language instructor: The dividents of a data-driven lexical approach to ESP. *The Journal of Language for International Business, 17*, 72–81.

Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System, 25*, 301–315. http://dx.doi.org/10.1016/S0346-251X(97)00024-9

Cobb, T. (1999). Breadth and depth of lexical acquisition with hands-on concordancing. *Computer Assisted Language Learning, 12*, 345–360. http://dx.doi.org/10.1076/call.12.4.345.5699

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology, 11*, 38–63.

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 400+ million words, 1990-present.* Retrieved from http://www.americancorpus.org

Ellis, N. C. (1997). Vocabulary acquisition, word structure, collocation, word-class, and meaning. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 122–139). Cambridge: Cambridge University Press.

Fan, M. Y. (2003). Frequency of use, perceived usefulness, and actual usefulness of second language vocabulary strategies: A study of Hong Kong learners. *The Modern Language Journal, 87*, 222–241. http://dx.doi.org/10.1111/modl.2003.87.issue-2

Folse, K. S. (2004). *Vocabulary myths: Applying second language research to classroom teaching.* Ann Arbor: University of Michigan Press. http://dx.doi.org/10.3998/mpub.23925

Guan, X. (2013). A study on the application of data-driven learning in vocabulary teaching and learning in China's EFL class. *Journal of Language Teaching and Research, 4*, 105–112.

Hadley, G. (2002). An introduction to data-driven learning. *RELC Journal, 33*, 99–124. doi:10.1177/003368820203300205

Hamzah, M. S. G., Kafipour, R., & Abdullah, S. K. (2009). Vocabulary learning strategies of Iranian undergraduate EFL students and its relation to their vocabulary size. *European Journal of Social Sciences, 11*, 39–50.

Huei Lin, M. (2016). Effects of corpus-aided language learning in the EFL grammar classroom: A case study of students' learning attitudes and teachers' perceptions in Taiwan. *TESOL Quarterly, 50*, 871–893. http://dx.doi.org/10.1002/tesq.2016.50.issue-4

Jafarpour, A., & Kousha, M. (2006). Data-driven learning and teaching of collocation of prepositions: the case study of Iranian EFL students. Research on foreign languages. *Journal of Faculty of Letters and Humanities, 49*(200), 1–31.

Johns, T. (1990). From printout to handout: Grammar and vocabulary teaching in the context of datadriven learning. *CALL Austria, 10*, 14–34.

Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *Classroom concordancing. English Language Research Journal, 4*, 1–16.

Kafipour, R., Yazdi, M., Soori, A., & Shokrpour, N. (2011). Vocabulary levels and vocabulary learning strategies of Iranian undergraduate students. *Studies in Literature and Language, 3*, 64–71.

Koda, K. (2000). Cross-linguistic variations in L2 morphological awareness. *Applied Psycholinguistics, 21*, 297–320. http://dx.doi.org/10.1017/S0142716400003015

Lewis, M. (1993). *The lexical approach. The state of ELT and a way forward.* Hove: Language Teaching Publications.

Lewis, M. (1997). *Implementing the lexical approach. Putting theory into practice.* Hove: Language Teaching Publications.

McCarthy, M. (2008). Acessing and interpreting corpus information in the teacher education context. *Language Teaching, 41*, 563–574.

Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research, 18*, 393–407. http://dx.doi.org/10.1191/0267658302sr211xx

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524759

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511497650

Pickard, V. (1994). Producing a concordance-based self-access vocabulary package: Some problems and solutions. In L. Flowerdew & A. K. K. Tong (Eds.), *Entering text* (pp. 215–226). Hong Kong: Language Centre.

Read, J. (2010). Methodological issues in research on vocabulary learning. In Y.-N. Leung (Ed.), *Proceedings from the 19th International Symposium on English Teaching, Taipei, Taiwan* (pp. 63–73).

Reppen, R. (2011). Using corpora in the language classroom. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 35–51). Cambridge: Cambridge University Press.

Romer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 83–98). Philadelphia, PA: John Benjamins.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge, MA: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524780

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge: Cambridge University Press.

Sinclair, J. (2004). New evidence, new priorities, new attitudes. In J. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 271–299). Amsterdam: John Benjamins. http://dx.doi.org/10.1075/scl

Singleton, D. (1999). *Exploring the second language mental Lexicon.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139524636

Stevens, V. (1991). Classroom concordancing: Vocabulary materials derived from relevant, authentic text. *English for Specific Purposes, 10*, 35–46. http://dx.doi.org/10.1016/0889-4906(91)90014-N

Willis, J. (2011). Concordances in the classroom without a computer: Assembling and exploiting concordances of common words. In B. Tomlinson (Ed.), *Materials development in language teaching* (pp. 51–78). Cambridge: Cambridge University Press.

cogent •• education

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at www.CogentOA.com**