

3 ICLICE 2016 8 Mohammad Javad Rezai

Teaching Confusable Words to Iranian EFL Learners: A Corpus Based Approach

Mohammad Javad Rezai*, Elham Ezatabadipour
English department, Yazd University,
Safaieyeh, Yazd University, Iran
Corresponding author: mreza@yazd.ac.ir

ABSTRACT

Confusable words such as *confident* and *confidant* refer to those lexical items which sound or look alike, but frequently cause confusion. The purpose of the current study, using a corpus-based approach to teaching confusable words, was to study the role of corpus in learning and retaining confusable words, to probe into the role of frequency in learning confusable words and to find out the learners' feedback and their attitude toward using corpus in learning confusable words. To this end, 43 male and female freshmen Iranian EFL learners, divided into two groups, served as the participants of this study. After 5 weeks of treatment using BNC corpus for the experimental group and traditional method for the control group, a post-test was conducted followed by a one-month delayed post-test. The results indicated that the use of corpus as a new trend was significantly helpful in teaching confusable words. The results also indicated that the use of corpus can be effective in retaining lexical items. However, the frequency effect was not significant. The findings of the study offer practical implications for the inclusion of corpora in the process of second language acquisition.

Keywords: Corpus, Confusable Words, Retention, Acquisition

Introduction

One of the influential features of every second language teaching and learning environment is vocabulary. It is crystal clear that knowing the more number of words can lead to better understanding of what has been heard and read, and better ability in saying what is intended when we are speaking or writing. The importance of vocabulary for second and foreign language learning is not deniable. This area of teaching is so vast and finding new strategies for teaching vocabularies is the concern of most of the language learning programs.

By the advancement of educational technology, new strategies of teaching vocabularies have been proposed by the investigators. One of the new trends in vocabulary teaching is the use of corpus-based approach. According to Spolsky and Francis (2008) corpus linguistics can be used as an approach for investigating the use of language. The device that this approach uses is computer assisted techniques for analyzing large collections of writing and transcribed speech. The purpose of such investigations is to understand the choices that speakers and writers make in particular conditions, whether typical choices or unusual ones.

English is among those languages containing words which sound similar but possess different meanings which are easily confused. There is a category of words in English that sound or look alike which frequently cause confusion. A word like *apprise* which means "telling someone about something" can be easily confused with the word *appraise* which means "examining someone or something to judge the qualities" since their pronunciation and spelling seem similar.

Hence, choosing the correct word among the sets of confusable words is a problematic area which needs a closer look. The difference between these words can include the difference in one vowel or in one or two consonants. Example of confusion sets include: *principal* and *principle* or the difference between *allusion* and *illusion* (Banko & Brill, 2011).

The attitude of linguistics has been changed by the passage of time and a shift in ELT from grammar to lexis mirrors this change completely (Liu & Jiang, 2009). In the past, the most important concern of linguistics was grammar. However, the advances in corpus linguistics have introduced lexis as a new trend to this field which is worth doing new investigations. Indeed, the role of corpora in the acquisition of lexico-grammatical elements of language is undeniable (Csomay & Petrovic, 2012).

Nowadays, educational theories and practices have been influenced by the use of corpora and the effect of corpus linguistics is mostly apparent in the areas of vocabulary analysis. Using large electronic collections of texts sampled from actual language use is the main concern of corpus-based approach which may let the researchers have a better classification and identification of vocabulary items (Spolsky & Francis, 2008).

Learning confusable words needs more practice using novel methods, and there is a need for more investigation in order to find a new strategy and effective approach for teaching confusable words. This research is going to declare the potency of using a corpus-based approach in teaching confusable words.

The current research sought to study the role of corpus in teaching confusable words. The effective nature of each method can be identified when the students can remember what they have learned after a period of time. So, through this study, the influence of using corpus on word's retention was also investigated. The role of vocabulary frequency in vocabulary learning was another important purpose of this study.

Since teaching via corpus is the concern of the present investigation, the current research can determine whether students are eager to learn words using corpus. In other words, based on the findings of this research, the learners will be able to decide whether they want to continue being exposed to the corpus-based approach or not.

In line with the above-mentioned objectives, the current research addressed the following questions:

- 1) What is the effect of corpus-based approach in teaching confusable words?
- 2) What is the effect of vocabulary learning via corpus on the retention of lexical items?
- 3) What is the role of word frequency on learning confusable words?

Review of the Related Literature

Using Corpus in Teaching

The main concern of the current study was using corpus in teaching vocabulary. There are a number of studies using corpus as the means for teaching and also studying different aspects of language. But using corpus as the means for teaching a specific category of vocabularies called confusable words is something avant-garde.

Granger (2012) provides a typology for learner corpora. He introduces different features as the main dimensions of this typology including: time of collection, scope of collection, targeted language, learners' mother tongue, medium and text type.

According to Aston (2001), the use of corpora can provide evidence about linguistic performance which can help finding those aspects of language that needs to be taught. In corpus-based studies, researchers test a hypothesis or question based on corpus data like "British National Corpus".

Widdowson (1978) also offered some comment that can be used in a connection with the issue of corpora. He provided a contrast between genuineness (a quality of texts) and authenticity (a quality of discourse interpretation). Considering these two concepts, corpora which are those texts that occur naturally, provide samples of genuine language. But there is no guarantee for their authenticity when they are reproduced as discourse and part of other contexts. In other words, Widdowson does not believe in authenticity of the materials provided in corpora (cited in Macdonal & Badger, 2000).

Aston (2001) indicated an alternative way for preserving authenticity, which is adaptation of role of observer. So, the learner who uses the corpora should be both participant and observer. Playing the role of observer will be helpful in noticing the strategies of interaction between learners and text, and being participant will help learners to test these strategies (as cited in Aston, 2001).

Linguist (2009) declares that corpus linguistics can be used as a methodology including different methods for investigation of different aspects of language such as sociolinguistics, syntax, semantics, pragmatics, and discourse analysis.

Aston (2001) suggested that corpora are good resources for teachers and material designers. Corpora are helpful for making choice in teaching. It also aids exploring tests, problematizing language and authentication of discourse.

According to Romer (2010), the link between corpus linguistics and language teaching can be considered in different parts including: in the past and in the present.

At present, researchers distinguish two different types of corpora including "General" and "Specialized" Corpora. General corpora possess some unique characteristics. They are fairly large in size and include text types from different varieties and registers of the intended language. Examples of such corpora are COBUILD Bank of English, the British National Corpus (BNC) and the corpus of Contemporary American English (COCA). Specialized corpora are collections of texts from a specific field of study like economics. These texts are produced by a small group of language users (for example, advanced learners of English whose L1 is Swedish). There are different features related to the use of corpus-aided discovery learning such as learner motivation, serendipity, communicative competence, language awareness raising and autonomous learning. The Specialized corpora can be helpful in ESP courses through which the priority is given to teaching words and expressions that learners need for handling that text in a better way. The examples of specialized corpora are Academic Word List (AWL) by Coxhead (2000) and Romer(2010).

According to Granger (2012), linguistic utilization of learners' corpora can happen through two methodological approaches. The first approach is contrastive interlanguage analysis which includes comparison between native and non-native data or different varieties of nonnative data. The second approach is computer aided error analysis. The focus of attention in this approach is on errors in interlanguage by using computer devices for retrieving them.

Previous Studies

In what follows, some of the related studies on the use of corpus in teaching different aspects of language are presented.

Phocharoensil (2012) conducted a study to find out the effect of language corpora for Thai EFL learners, in which the effect of using corpora on learning English grammar was under investigation. The purpose of this study was finding the students' attitude toward corpus based grammar teaching which was obtained through the use of questionnaire. The participants of this study were 17 native Thai students who were graduated in English for careers program at a university in Thailand. The instruments used in this study were,

questionnaire and interview. At the end of the study, most of the students had a very positive attitude towards this method of teaching. In fact, the students were satisfied with not only language learning by means of corpora, but also understanding grammar points rather than memorizing them.

Csomas and Petrovic (2012) used a corpus-based technique to investigate the extent to which incidental vocabulary learning can be facilitated by the use of watching related movies and TV shows. Frequency of distribution of a specific category of technical words (legal vocabularies) was the matter of concern. The corpus in this study consisted of transcription of seven movies and five TV episodes, with a content relevant to legal vocabularies. The results indicated that, the corpus of 130000 words, movies and TV shows possesses a high potential to be used as means of incidental vocabulary learning and they are capable of being utilized in ESP courses.

Liu and Jiang (2009) investigated the effect of using corpus and contextualized lexico- grammar in second language teaching. The British National Corpus (BNC, 2001) and the BNC baby (2005) were used as the corpora in this study. The data for this study were the students' corpus search, grammar exercises and their journals about the use of corpus, the teachers' journal, sample teaching activities, and students and teachers post study questionnaire. The participants of this study were from one university in china and two universities in the United States. The results showed that using corpus and lexico-grammatical approach in second language teaching can be effective leading to some outcomes such as improvement in grammar understanding and progression in learning skills. The study also revealed some features that can be influential in using corpus and corpus analysis such as course content, learning settings and students' learning strategies.

In another study, Jafarpour and Koosha (2006) investigated the effect of concordance materials on teaching and learning collocation of prepositions. The participants were 200 senior students majoring in English at three universities of Shahrekord, Iran. The materials of collocations of prepositions were from a corpus called Brown corpus (2005). The teaching materials related to the prepositions were from various grammar books. The results indicated that data driven learning approach is effective in teaching and learning the collocation of prepositions. Secondly, learners' performance on collocations of prepositions has a positive relation with their level of proficiency. The final result of this study was that EFL learners tend to use the same collocation patterns of L1 in their L2 production.

Given the above-mentioned investigations, it seems that there are few studies which have concentrated on teaching confusable words via corpus. There is a need for more investigation aimed at finding new strategies and trends in teaching this neglected category of words which is problematic in the area of second language acquisition. Therefore, the present research intends to fill the above-mentioned gaps by conducting an empirical investigation on the role of corpus-based approach in teaching confusable words.

Methodology

Participants

The participants of this study included 43 male and female freshmen Iranian EFL learners majoring in English language and literature from Yazd University. The students' age ranged from 18 to 34. To measure the students' level of proficiency, the Oxford Quick Placement Test was administered after which the intermediate level participants were selected for the purpose of the study.

Instruments

The main instrument used in the current study for the experimental group was the British National Corpus (BNC) and COBUILD dictionary. The BNC corpus comprises 100 million word collections of samples. These samples include both written and spoken languages collected from different sources.

To offer an example, the distinction between *principle* and *principal* is elaborated here through the BNC corpus.

Principle

- 1) Buck never allowed himself to do anything that went against his **principles**.
- 2) I agree with it in **principle** but I doubt if it will happen in practice.
- 3) These people lack all understanding of scientific **principles**.
- 4) This **principle** will remain central to every element of training at every stage
- 5) He would vote against it on **principle**.

Principal

- 1) Methane is one of the **principal** gases contributing to the greenhouse effect.
- 2) I found myself unable to identify the soup's **principal** ingredient.
- 3) That was my **principal** reason for moving.
- 4) He was **principal** dancer at the Dance Theater of Harlem.
- 5) Donald King is the **principal** of Dartmouth High School.

Procedure

The design of the current investigation was a quasi-experimental one including a pre-test, a treatment, a post-test and delayed post-test applied in this study. The pre-test was conducted before any instruction to determine the extent to which the students were familiar with the selected confusable words. The pre-test included 40 selected confusable words which were scattered randomly. The participants were asked to write the meaning or a word which was synonymous to the given item.

The materials used for instructing confusable words were different between these two groups. For the corpus group, the material was selected from British National Corpus (BNC) and COBUILD dictionary as two sources of authentic language use examples. Four pairs of words were selected for each session and five sentences were chosen as instances of authentic use of that word.

The examples were on sheets of paper distributed among the students, read by the instructor and the synonyms or the meanings of those confusable words were explained.

The materials used for the traditional group were two instances under each word in Oxford Advanced Dictionary. The instructor read each vocabulary, explained the meaning of that word and then read the two sentences as examples to illustrate the meaning of that word.

After the treatment, an immediate post-test was conducted for both corpus and traditional groups. The post-test consisted of a production and comprehension task. This post-test took 30 minutes to administer and the production questions were distributed first. Following the production questions, the comprehension questions were distributed among the participants.

Results

Table 1 presents the means and standard deviations of learners' overall performance on the post-test. The learners in the corpus group outperformed those of the traditional group (Mean difference=8.60)

Table 1
Descriptive statistics of the posttest results

| | Group | N | Mean | Std. Deviation |
|-----------|-------------------|----|-------|----------------|
| Post_test | Traditional Group | 22 | 14.77 | 4.898 |
| | Corpus Group | 21 | 23.38 | 4.295 |

The data obtained from the independent samples t-test indicated that there was a significant difference between the two groups concerning their overall performance on the post-test ($t_{(41)}=6.13$, $P=0.0001$). The magnitude of the difference in the means was large ($\eta^2=0.47$).

In order to have an analysis of the performance of the participants in the delayed post-test, an independent samples t-test was conducted. Table 2 shows the means and standard deviations of two groups' performance. The mean accuracy score of the corpus group was higher than that of the traditional group (Mean difference=2.94).

Table 2
Descriptive statistics of the participants' performance in the Delayed Post-test

| | Group | N | Mean | Std. Deviation |
|-------------------|-------------------|----|-------|----------------|
| Delayed_post_test | Traditional Group | 22 | 11.82 | 3.887 |
| | Corpus Group | 21 | 14.76 | 4.614 |

The examination of the learners' overall performance in the delayed post-test revealed that there was a significant difference between the participants in the two groups in favour of the corpus group [$t_{(41)}=2.26$, $p<0.05$). Nonetheless, the magnitude of the difference in the means was moderate ($\eta^2=0.11$).

The last question in the current study was the effect of frequency of words in learning confusable words. In other words, one aspect of this investigation was an effort to find out whether the results were under the influence of high and low frequency words.

In order to address this question, at first the words were categorized into two groups of high and low frequency words. This categorization was based on the count displayed on each word query's page, indicating the frequency for the selected word within the texts making up the BNC. Those words which possess a frequency count less than 1000 were considered as low frequency words and those which possess a frequency number more than 1000 were considered as high frequency words.

The frequency of high and low frequency words was calculated via SPSS. Then a paired samples t-test was conducted in order to compare the high and low frequency words in post-test and to find out whether there was any significant difference in the mean scores under these two different points. Table 3 shows the descriptive statistics of the participants' performance for high and low frequency words. As can be seen from the table, the frequency was not a significant factor in providing responses to confusable items.

Table 3
Descriptive statistics of the participants' performance in terms of frequency

| | Mean | N | Std. Deviation |
|--------------------------|---------|----|----------------|
| High frequency post test | 7.7209 | 43 | 2.44293 |
| Low frequency post test | 11.1860 | 43 | 4.41995 |

The results of the paired-samples t-test indicated a statistically significant difference between the performance of students on high frequency ($M=7.72$, $SD=2.41$) and low frequency words ($M=11.19$, $SD=4.42$). So it can be said that considering the post-test, the participants performed better on low frequency words in comparison with the high frequency words. [$t_{(42)}=6.64$, $p<0.005$ (two-tailed)].

Discussion and Conclusion

The first question addressed the efficacy of corpus-based approach on teaching confusable words. The overall results of the study revealed that the L2 learners in the corpus group ($M=23.38$) outperformed the participants in the traditional group ($M=14.77$). The obtained results indicate the effectiveness of using corpus in the realm of language teaching in general and vocabulary teaching in particular. Evidence from corpora can provide a lot of insights on the functioning of language. Indeed, a vast amount of lexical patterning can be revealed by studying corpora. They offer lexico-grammatical information which can be unique in their nature. The results of the current study lend support to the view that such a computing technology should be incorporated in different areas of language teaching at higher levels of language proficiency.

Exploring the effectiveness of using corpus in teaching different aspects of language was the matter of concern in a number of previous studies. Phocharonesi (2012), Jafarpour and Koosha (2006) and Liu and Jiang (2009) have drawn the researchers' attention to the importance of corpora in teaching English as a second language. So, the use of corpus in teaching different aspects of second language learning is a research demanding area. However, teaching confusable words via corpus is something novel among the previous studies in which corpora were the major source of materials.

The results of current study are in line with Jafarpour and Koosha (2006) for using corpus as the main source of teaching materials and the effectiveness of the treatment by the use of corpus. However, the current study did not take advantage of concordance lines and corpus driven approach, since the main methodology in this study was corpus-based approach.

According to Liu and Jiang (2009), having a lexicogrammatical approach by the use of corpus materials can be a good strategy in teaching a second language like English. The findings of current study are in line with Liu and Jiang (2009) for the effectiveness of using corpus in teaching, especially teaching lexical items. All in all, the use of corpus can be efficient in disambiguating the subtleties involved in the acquisition of confusable words. The traditional methods do not offer any superiority in this regard. The effect of using corpus on the retention of vocabulary items was investigated in the second research question. The overall results show that corpus group's performance on the delayed post-test production task ($M=3.52$) was significantly better than that of the traditional group ($M=2.18$) with a mean difference of 1.34. Furthermore, the students' performance in the delayed post-test comprehension task in the corpus group ($M=11.24$) was better than that of the traditional group ($M=9.64$) although the difference did not turn out to be significant.

A comparison of the participants' performance in the two groups across time was made possible by considering both post and delayed post-test. The overall results of the delayed post-test and also comparing the results of the delayed post-test with the post-test show that both groups had a lower performance on the second administration of the test. Nonetheless, the difference between the two groups was still significant. Given the above, it can be concluded that the use of corpus in teaching confusable words can be an effective strategy which helps in the retention of vocabulary items.

Investigating the role of word frequency was another important factor in the current study. In other words, one part of the present study aimed at finding the difference between learning high and low frequency words.

The results of the conducted paired samples t-test on the post-test showed that the students' performance on the low frequency words ($M=11.18$) was slightly better than that of high frequency words ($M=7.72$). Nonetheless, there was no significant effect for the frequency of words in the delayed post-test.

On the whole, the obtained results do not lend support to the facilitative effect of high frequency words in the acquisition of confusable words. This was the case in both immediate and delayed post-tests. The L2 learners could perform better on those words which were less frequent, but this factor was not effective through the time. One possible reason is that the effect of instruction overrides the frequency factors. The learners may also show more awareness and conscious attempt to get to know the meaning of less familiar words

The analyses of the data provided a variety of advantageous effects and some challenges through the experience of having a journey on the world of vocabulary by using a new strategy called corpus-based approach.

The obtained results from different analyses on the data show that the use of corpora is effective in providing a new strategy of vocabulary learning. Since the learners are responsible for inferring the meaning of the given words from the examples. This new strategy can be facilitating in this regard.

The use of corpora is also influential in the improvement of the command of grammatical rules and enhancement of using context for guessing the meaning of vocabulary items among EFL learners. On the other hand, using corpora as a new strategy in teaching vocabularies can be helpful in the retention of lexical items like confusable words.

The use of corpora is also a reliable instrument for both inductive learning and deductive learning. As Aston (2001) declares, through inductive learning the students are confronted with vocabulary and grammar usage in concordance data (corpora) and then they try to generalize what they have observed into special rules and patterns. However; in deductive learning, language learners use corpora for testing rules and patterns that they have in their mind or classifying the corpus data into groups of rules and patterns which they are familiar with.

Thus, the provided motivation by means of corpora and the use of inductive and deductive learning will lead the learners to retain rules and patterns, since they are engaged in a deeper processing of language.

It is worth mentioning that one important role in determining the degree of effectiveness of current strategy is played by the students. It is completely clear that, on the basis of instructor's observations, motivation of students for learning new vocabularies and testing novel strategies in such learning processes is a major factor in the present study. The importance of this factor is revealed by referring back to the answers given to feedback questionnaire. The motivated students, who preferred learning new items and discovering new strategies simultaneously, responded to the questions more positively.

The results of the present study provide different implications for language teachers, language learners and practitioners. Based on the findings of the current research, the instructors can make valuable suggestions for the improvement of teaching vocabulary items. These suggestions can be based on their students' and their own experience in using corpus.

The fact that the use of authentic examples of intended lexical items as context can be an advantageous aid for learning specific types of vocabularies and is an important fact that language instructors can utilize as a new strategy of teaching vocabulary.

Last but not the least, the use of corpora can be helpful for memorization and retention of lexical items. The EFL learners can benefit from corpora in their current and future learning provided that they hold a positive view towards the efficacy of corpus-based approaches in their language acquisition process.

References

- Aston, G. (2001). Enriching reality: language corpora in language pedagogy. *ELT journal*, 55(3), 238-246.
- Banko, M., & Brill, E. (2001). Scaling to very very large corpora for natural language disambiguation. *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, 26-33. Association for Computational Linguistics.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238.
- Csomy, E., & Petrović, M. (2012). “Yes, your honor!”: A corpus-based study of technical vocabulary in discipline-related movies and TV shows. *System*, 40(2), 305-315.
- Granger, S. (2012). Learner corpora. *The Encyclopaedia of Applied Linguistics*. Doi:10.1002/9781405198431.
- Jafarpour, A. A. & Koosha, M. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL journal*, 8(4), 192-209.
- Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.
- Liu, D., & Jiang, P. (2009). Using a Corpus-Based Lexicogrammatical Approach to Grammar Instruction in EFL and ESL Contexts. *The Modern Language Journal*, 93(1), 61-78.
- MacDonald, M., Badger, R., & White, G. (2000). The real thing?: authenticity and academic listening. *English for Specific Purposes*, 19(3), 253-267.
- Phoocharoenil, S. (2012). Language Corpora for EFL Teachers: An Exploration of English Grammar through Concordance Lines. *Procedia-Social and Behavioral Sciences*, 64, 507-514.
- Romer, U. (2010). Using general and specialized corpora in English language teaching: Past, Present and Future. *Corpus-based approaches to English language teaching*, 18-38.
- Spolsky, B & Francis, M.H. (2008). *The handbook of educational linguistics*. London :Blackwell Publishing. 539-556.
- The British National corpus, version 3(BNC XML Edition), Distributed by Oxford University Computing Services on behalf of the BNC Consortium. URL: <http://www.natcorp.ox.ac.uk/>