Concordancing as a Tool in Learning Collocations: The Case of Iranian EFL Learners

Shiela Kheirzadeh*., Seyyedeh Susan Marandi

**Alzahra University, English Language Department, Vanak 1993893973, Tehran, Iran
**Sobhe-Sadegh Institute of Higher Education, English Department, Beheshti, Isfahan, Iran

Abstract

Corpus use is one of the promising areas that can act as a reference tool to EFL learners for language problems especially vocabulary. The present study is on concordancing as a tool in learning collocations and it attempted to find the answers to two questions. One on the potential benefits of corpora and concordancing in learning collocations and the other on the type of the collocational combinations searched more by the students. To find the answer to the research questions, the students were taught how to use corpora independently in the hope of becoming autonomous users in future. The results of the study indicated an overall satisfaction with the use of the concordancers in learning collocations among which verb + noun collocations were searched more frequently as they were more problematic for EFL learners. In line with the results of the study some of the implications of using concordancing in learning and teaching are presented at the end of the paper.

© 2014 The Authors. Published by Elsevier Ltd. Open access under CC BY-NC-ND license. Selection and peer-review under responsibility of Urmia University, Iran.

Keywords: corpora; concordancing; collocation

1. Introduction

Corpus study is an emerging and flourishing field in second language acquisition that owns most of its progress to the popularized and widespread use of technology. Gavioli (2005) proposes that language corpora and the methods used for their exploitation have great potential for language learning. Corpora provide systematic access to naturally occurring language, and corpus-linguistic methods support exploratory learning and are well-suited to

* Corresponding author.
E-mail address: sh.kheirzadeh@yahoo.com
encourage autonomous learning and teaching. Contrary to the fact that corpus studies have progressed and expanded enormously in recent years, and involved various aspects of applied linguistics, including language teaching, learners’ language and translation teaching, the use of corpora as a tool in the hands of teachers and learners is moving on very unsteadily (Gavioli, 2005). As claimed by Kennedy (2004), three decades of research on electronic corpora have had surprisingly little influence on language curriculum contents since the amount and type of data provided by corpus concordancing appear, at first glance, unintelligible to teachers and learners, leading them to wonder whether working with such data is really worth the effort. Furthermore, to Granger (2009), learner corpus research has not yet fully realized its stated ambition in that it has given rise to relatively few concrete pedagogical applications. But what exactly corpora are and what are their benefits?

To put simply, a corpus is usually defined as a principled collection of naturally occurring texts for linguistic study (Biber, Conrad, & Reppen, 1998) and a concordancer is a computer software program that rapidly searches a corpus and produces a list of incidences of a given linguistic item (a word or phrase). Yoon (2011) mentions some potential benefits of corpora and concordancing for language learners. First, corpora provide a rich exposure to typical patterns learners face in the real world. Second, they are useful in helping students learn grammar in a real life manner as opposed to the contrived grammar books. Third, they show the function of words in different words and contexts. Forth, they help finding patterns and forms not obvious in other sources like dictionaries. And finally corpora foster inductive learning and this type of learning facilitates long-term acquisition of language and helps learners, in the process, develop cognitive and metacognitive skills such as noticing, hypothesizing, and verifying (O’Sullivan, 2007). Moreover, Hunston (2002) emphasizes that concordancing:

> supports learning, partly because students are motivated to remember what they have worked to find out. In addition, because corpus data can reveal previously unnoticed patterns, a student may well notice something that a teacher has overlooked, or that no textbook covers. (p. 170)

As it was mentioned above, concordancing is a valuable asset in different fields such as translation, linguistics, material preparation and teaching different parts of the language namely grammar and vocabulary. As for vocabulary, the corpus-based perspective states that a lexically based grammar should be emphasized in language teaching. In other words, the approach assumes that “vocabulary should lie at the heart of the teaching process, impinging on, or even guiding the teaching of all areas of language, including grammar” (Owen, 1993, p. 171). Related to the idea of teaching vocabulary is teaching collocations. Collocation is the patterns of the co-occurrence of linguistic features in texts, that is, which words commonly appear together. The term “collocation” is traced back to Firth, who proposed “the meaning of a word was much a matter of how the word combined textually with other words” (McCarthy, 2001, p. 62). As an example of collocation, McCarthy (2001) points out that the combination of “happy marriage” is often observed rather than that of “content marriage”. A crucial notion of corpus linguistics is that it looks at words in typical surrounding contexts of occurrence rather than in isolation which is quite in line with the new methods of language teaching focusing on contextualization.

Focusing on the collocations, the present paper aims at finding how corpora and concordancing help the EFL learners learn collocations and which types of the collocations are mostly checked, searched and learnt in corpora. In other words, this study is to answer the following questions:

1. What are the benefits of concordancing in learning collocations for Iranian EFL learners?
2. Which types of collocational combinations are mostly searched in the concordancers?

2. The review of the literature

According to Granger (2009), learner corpus research has not yet fully realized its stated ambition in that it has given rise to relatively few concrete pedagogical applications. Johns (1991a) suggests that concordances derived from electronic corpora can be used as a source of learning materials for language students. Her main point is that
students should have access to concordance materials in order to examine meanings and functions of words in their authentic contexts as access to linguistic data may improve students’ analytical skills and help their inductive ability. Higgins and Johns (1984), focusing on learners, suggest that rather than making the machine more intelligent in answering learner generated questions, pedagogy of the foreign language should be concerned with making the learner “more intelligent” in posing questions and interpreting responses. In this perspective, the machine should be seen as a tool in the hands of the language learners, rather than a teaching guide.

As Johns (1991a) claims, corpora are the source of learning materials for learners. These learning materials can foster the acquisition of different parts of the language. In the view of the fact that the focus of the present study is on the collocations and corpora, a more detailed look at the idea of corpora and the benefit of corpora and concordancing in learning this somehow burdensome aspect of language worth attending.

2.1. Collocation

Collocations in learner language appear to be a worthwhile subject of study, as collocations are, on the one hand, pervasive in language, and, on the other, difficult even for advanced learners of English. In the literature on the topic, statements such as the following abound: “any analysis of students’ speech or writing shows a lack of collocational competence” (Hill, 2000, p. 49). Collocation can be a key to differentiate natives from non-natives. Likewise, correct uses of collocation are the symbol of advanced level English and writing proficiency; therefore, collocation is the benchmark for total fluency in L2 learning.

The way of looking at collocations in this paper is in line with Nesselhauf (2005) who claims collocations are defined in a phraseological rather than in a frequency-based sense. This means that collocations are not defined as frequently co-occurring words, but as combinations which are not entirely predictable semantically but are arbitrary to some degree. According to this approach, on the one hand, a combination such as drink tea would not be considered a collocation, although the two words are likely to frequently co-occur. On the other hand, combinations such as make a decision, take something into consideration, get in touch with somebody would count as collocations, as part of the reason why these particular verbs and nouns are combined is arbitrary convention.

Hill (2000), who emphasizes the importance of collocational knowledge in L2 pedagogy, addresses a fundamental question of what it means to know a language. He claims that “[students] do not really ‘know’ or ‘own’ a word unless they also know how that word is used [emphasis in original], which means knowing something about its collocational field” (p. 60). Wray (1999) proposes two functions of collocations as: first, they play an essential role in language learning, as they seem to be the basis for the development of creative language. And second, they are essential for fluency in both spoken and written language. Nesselhauf and Tschichold (2002) claim that we are still far from the development of a coherent methodology and even further far from a wide-spread and systematic treatment of collocations in language teaching materials and syllabuses. Granger (cited in Nesselhauf, 2005) in a study of the collocations in the writings of elementary and intermediate L1 German students come to the conclusion that the reason for using collocations is the desire of many learners (often supported by teaching practices) to be original and expressive.

Nesselhauf (2005) in an attempt to find the factors correlating with learners’ difficulties with collocations, compared two groups of V+N collocations called RC1, collocations with verbs that permit only a very limited number of nouns like draw the curtain, and RC2, collocations with verbs that permit a larger number of nouns but where some arbitrary restriction nevertheless holds like make a new start. The results showed that while the RC1 collocations that were produced by the learners were acceptable 75.1% of the time, RC2 collocations were acceptable only 63.3% of the time. It seems, therefore, that if RC2 collocations are produced, they are more likely to be deviant than if RC1 collocations are produced. The reason for the difficulty of RC2 collocations, as claimed by Nesselhauf (2005), is that this group of collocations is on the borderline between collocation and idioms. They are considered as borderline collocations as "the meaning of the noun was judged to be close to one of its meanings
outside a few expressions, but the noun seems less independent than in other collocations, and the meaning of the verb is often restricted to the noun in question” (p. 215). Furthermore, Nesselhauf (2003) presents the most explicit continuum model which focuses on “verb-noun combinations” and presents three major classifications of the collocation: free combinations (both verb and noun are used unrestrictedly, e.g., *want a car*), restricted collocations (the verb is used restrictedly and combined with certain nouns, e.g., *take a picture*), and idioms (both verb and noun are used restrictedly and often seen as a chunk, e.g., *foot the bill*).

Howarth (1998a) also claims that more cohesive collocations are less difficult for advanced learners than less cohesive ones. And collocations with high-frequency and/or light verbs are deemed worthy of particular attention (Hill, 2000). Moreover, Hill argues that the use of fixed expressions can facilitate naturalness, fluency, and effectiveness in language use. According to him, the obstacle for intermediate students to move to the advanced level is the lack of collocational knowledge. As he puts it:

> Spending a lot of class time on traditional EFL grammar condemns learners to remaining on the **intermediate plateau** [emphasis in original]. Helping learners to become ‘advanced’ needs a huge injection of lexis. It is lexis in general, and collocational competence in particular, which allows students to read more widely, understand more quickly, and speak more fluently. (2000, p. 68)

Considering the great concern learners have to learn collocations, especially in EFL contexts where natural exposure to language is absent, and since it is the sign of being advanced in language and attaining naturalness, fluency, and effectiveness in language use, we as the teachers need to provide them ample opportunities to learn. Of the many possible choices through which learners can learn collocations is concordancing, a software program that rapidly searches a corpus and produces a list of incidences of a given linguistic item (a word or phrase) which is the focus of the next section.

### 2.2. Concordancing and collocations

The finding that the collocations the learners produce are frequently not unacceptable *per se* but rather are existing English collocations used inappropriately in the context, probably lead to the conclusion that collocations should not merely be taught in isolation but also in context and that attention should be paid to their usage.

Empirical support for this view of contextualization has come from corpus studies, which have regularly found that most of naturally occurring language, both spoken and written, consists of recurrent patterns, many of which are phraseological (e.g., Altenberg, 1998; Stubbs, 2001). Sinclair (1991) argues that information about collocation is one of the most useful assets that corpus research can provide for language description. Aijmer and Altenberg (1991) also assert that “collocations represent the intersection of lexis and grammar, an area which can be fruitfully studied in corpora” (p. 4). Howarth (1996) claims that corpus studies have also shown that collocations are a frequently occurring type of semi-prefabricated unit. In an analysis of over 5,000 verb-noun combinations in a written 240,000 word corpus, for example, over a third of the combinations were found to be collocations.

Of the limited number of the studies which have specifically addressed the issue of teaching collocations, most are concerned with investigating the use of computer aided language teaching facilities in collocation teaching. Utilizing a concordancer program to examine the relative effectiveness of inductive and deductive approaches to learning grammatical collocations at two levels of difficulty in the classroom, Sun and Wang (2003), conclude that the inductive group improved significantly more than the deductive group in learning collocation as demonstrated by an error correction test. The level of difficulty of collocation was also found to influence the learning outcome with easy collocations being more suitable for an inductive approach. In a follow-up study by Chan and Liou (2005) on the effects of Web-based concordancing on collocation learning in a CALL classroom, they used five Web-based practice units, three of which included the use of a bilingual Chinese–English concordancer to teach verb + noun collocations to EFL students. In line with Sun and Wang’s (2003) results, they also found that explicit online
instruction was effective in promoting EFL learner knowledge of collocation, with results significantly higher for units in which the concordance had been used. Results also showed significant differences in learning between four verb + noun collocation types with concordancers deemed most suitable for use in the instruction of de-lexicalized verbs and L1–L2 noncongruent verb collocations. Lin (2002) examined the effects of employing receptive and productive tasks on verb + noun collocation teaching. The subjects in the study were divided into two groups of lower achiever and higher achiever students. The results indicated that both groups did better on the receptive tests than the productive ones, but lower achiever student had a slightly better performance on the productive tests that the higher achiever students after being exposed to the treatments.

1. Methodology

1.1. Participants

The participants of the study were 27 the junior students majoring at English Translation in Sobhe Sadegh Institute of Higher Education in Isfahan, Iran. They were both male and female with the age range of 20 to 30. Their level of proficiency was intermediate decided by the judgment of the five instructors, besides the researcher, who were teaching different courses to them. All these students were the students of one class, taught by the researcher, so there was no randomization. The reason was that teaching collocations is normally part of teaching a course that is called "idioms and metaphorical expressions" and the researcher was the instructor of this course.

1.2. Material

From among the concordancing tools and programs available on the net, Compleat lextutor (http://www.lextutor.ca/concordancers/) was selected since it was more manageable to the students in comparison with the other tools and the students were free to choose from among different types of corpora like spoken, learner, law, and medicine based on their own preference. Also it was possible for them to compare their search items in more than one corpus, something that was done by some students.

1.3. Procedure

To know about the potential benefits of learning collocations through corpora, in the first two sessions of the semester the students were introduced to the idea of corpora and concordancing and the possible advantages of its use. Different concordancing tools were introduced to the students and a list of some helpful ones were given to them along with a brief explanation about each. But as it was mentioned, for the present study http://www.lextutor.ca/concordancers/ was selected. In the next five sessions, the students were taken to the computer room and were allowed to work in group or individually, as they preferred, and searched for the collocations they wanted to know more about them and the ones decided by the course instructor. The purpose was to make sure if the learners could make an autonomous use of the concordancer. After these training sessions, the students were asked to do a sort of project for their course. They were asked to search the collocation(s) they felt they needed more and wrote down the results and samples of the search. They were also asked about the potential benefits of corpora in learning more about collocational combinations and whether they had any problem during the search process. Upon the delivery of their projects, they received both comments and credits either through e-mail or personal contact. Moreover, five out of 27 students were randomly selected for the interview to provide more detailed information about the pros and cons of concordancing.

3. Results

The first question was to find about the potential benefits of the concordancing and corpora for the students. The results of the interviews and the assigned projects showed an overall satisfaction and even excitement in using concordancers and corpora to learn collocations. The most frequent benefit claimed by the participants was that as
the collocations are one the most burdensome aspects of language for them to learn, concordancing was of the greatest help in their correct use in both writing and speaking. Along with the "idioms and metaphorical expressions" course, the students had the "essay writing" and "oral presentation" courses. The consensus was that since collocations are indispensible parts in both writing and speaking, concordancing helps them to be more expressive and confident in those courses too. On the same issue, one of the interviewees mentioned that:

Whenever I want to use a collocational combination, I translate from Persian to English and it is wrong mostly. But I could not check everything in the dictionary. It is too hard. Every week I must write an essay and talk about a topic. I can't use a dictionary. It is time consuming. I was depressed. Now by searching in corpora I can find the collocations I need and even more.

In addition to the great help of concordancing in learning and using corpora in writing and speaking courses, two of the students claimed for its help in their job requirements. One of them mentioned that she was a fashion designer so she was interested in the possible collocational combinations that are related to the topics and ideas of art or fashion, as for her assignment, she did an extensive search on the collocations with the word "design" and "fashion". Another one, a photographer, had the same idea since he was supposed to be able to read and write about photography, so he did searches on the words related to his job requirements such as "picture" and "camera". Another student claimed that she was much interested in medical topics since she was very interested in knowing how to help people so she did her search on medical corpora. For her and some others, personal interests were another incentive to search the concordancers owning to the fact that they could access to different field specific corpora as law.

The other benefits of corpora claimed by interviewees and other students were:

- Knowing about multiple meanings of the verb collocations
- Distinguishing the differences in meaning of a collocational combination by comparing it in both spoken and written corpus
- How to translate different collocational combinations differently, two of the students presented their collocation project search along with their exact or possible Persian translation
- While knowing some collocational combinations to satisfy their needs, they felt the need to learn more to perform better
- Faster, easier and more enjoyable that working with dictionary
- Knowing about the prepositions of a verb
- Watching films and reading journalistic texts

Over and above all these undeniable benefits, it would be remiss not to mention the cons of concordance search to students as claimed by some. The most salient reaction was that it was time consuming to sort through concordance examples, and sometimes frustrating not to find relevant examples. They also struggled with how to formulate search terms, and they were not able to understand all concordance examples and identify relevant ones. It was shown in many cases in which the interviewees and others mentioned that they were forced to click on the search item and see the bigger context to make head or tail out of it.

The second question was to find which kind of collocational combinations were searched more. The results of two of the searches are shown in tables 1 and 2. The most frequent type of searched collocational combinations, as it was shown in table 1, was V+N collocational combinations. Namely as:

- Take, [→, N], challenge, chance, class
- keep, [→, N], date, group
- give, [→, N], appearance, yell
- get, [→, N], idea, hint
- tell, [→, N], lie, truth
- set, [→, N], table, precedence
In addition to this frequent collocational combination, there were other combinations searched and reported. The below table is reported by one of the students as:

- Adj+ N (table 2')
- V+ prep

**Table 2: Summary of the search results**

<table>
<thead>
<tr>
<th>Adjective</th>
<th>The noun effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Altering</td>
</tr>
<tr>
<td>2</td>
<td>Additional</td>
</tr>
<tr>
<td>3</td>
<td>Adverse</td>
</tr>
<tr>
<td>4</td>
<td>Beneficial</td>
</tr>
<tr>
<td>5</td>
<td>Brass</td>
</tr>
<tr>
<td>6</td>
<td>Cleansing</td>
</tr>
<tr>
<td>7</td>
<td>Decorative</td>
</tr>
<tr>
<td>8</td>
<td>Desired</td>
</tr>
<tr>
<td>9</td>
<td>Economic</td>
</tr>
<tr>
<td>10</td>
<td>Evident</td>
</tr>
<tr>
<td>11</td>
<td>Evil</td>
</tr>
<tr>
<td>12</td>
<td>Lethal</td>
</tr>
<tr>
<td>13</td>
<td>Long-term</td>
</tr>
<tr>
<td>14</td>
<td>Net</td>
</tr>
<tr>
<td>15</td>
<td>Only</td>
</tr>
</tbody>
</table>
4. Discussion and conclusion

In general, the results of the present study showed the participants' enhanced awareness of the importance of collocation and collocational knowledge. Here, it is important to note that when they looked for collocation patterns, they not only focused on forms, but also on meanings, which is clear in their claims about the importance of collocations in speech and writing and also due to the point that some students try to find exact or near equivalents for the collocations in their L1 language, Persian.

As Hyland (2003) suggests, corpora and concordancing in the L2 writing class can serve as research tools that help learners to discover underlying rules and regularities from the given concordance data and to raise their language awareness on the one hand, or as reference tools that learners can consult when they encounter problems in writing along with or instead of traditional reference resources such as dictionaries and grammar books on the other. They recognized that patterns using the same words could differ according to subtle differences in the surrounding text; this was clear as nearly all students clicked on the searched collocation to see it in a bigger context. Therefore, collocation is a combined art that intermingles form with meaning. That is, the corpus approach is not only a mechanical process of looking for a form, but it is also concerned with meaning and subtle nuances.

Some of them reported that they began to pay more attention to the combinations of words in their reading as well as their writing and speaking while they claim that collocations help in reading journalistic texts. In other words, corpus use not only helped them to solve immediate writing and language problems, but also expanded their language awareness and helped them with their language learning. Their excitement with the use of collocations beyond their "idioms and metaphorical expressions" course clearly confirms this fact. Generally, they became independent learners as they transfer the skill to new material and settings. And As Nesselhauf (2005) claims, they try to be original and expressive.

The answer to the second research question was that verb+ noun collocations were among the most frequently searched collocational combinations by the students. This may be the indication of the greater difficulty in dealing with this type of collocational combination since the verb + noun collocations are among the most common collocations in English. In line with the Nesselhauf (2005) claim that if RC2 collocations, collocations with verbs that permit a larger number of nouns but where some arbitrary restriction nevertheless holds, are produced, they are more likely to be deviant than if RC1 collocations are produced, the search items that were mostly done by the students were of RC2 type indicating that verbs collocating with a wide number of words are the greatest source of difficulty as this group of collocations is on the borderline between collocation and idioms. Benson (1989) mentions that a large number of verb + noun collocations are arbitrary and non-predictable. For example one can say: commit a crime and perpetrate a crime, commit a fraud and perpetrate a fraud. However, one can only say commit suicide, not *perpetrate suicide; make an effort, not *do an effort. One can say hold a funeral, but not *hold a burial. Likewise, make an estimate is frequently used, but not *make an estimation According to Benson, Benson, and Ilson (1986) many lexical collocations in English consist of a verb and a noun, such as bring in an acquittal, file a complaint, and put on airs.

Here are several possible reasons for such a high percentage: first, verbs are among the most difficult linguistic elements in the process of second language acquisition, because, as pointed out by Nesselhauf (2005), the verb is arbitrarily restricted in its collocability, the arbitrariness bring much more problems to EFL learners as they do not have the chance to have more exposure to the authentic language use which naturally makes its usage more difficult; second, the negative transfer or literal translation is the most important cause, as many foreign language learners
take it for granted that every linguistic element has its counterpart in their mother tongue, thus they tend to use verbs regardless of their semantic features. This was evident in the present study where some students try to come to a generalization for the translation of for example the verb *take* in different collocational combinations.

5. **Implications of corpora in learning collocations**

Learning collocations must involve some repetition, as suggested by Nesselhauf (2005), and corpora seem to be a chance for repeating them. Another factor that helps learners learn collocations is what Nation (1990) calls the systematicity in teaching collocations. This systematicity can be achieved through different means as essay writing, but essay writing not in the sense of giving a topic to students and asking them to write. The way that essay writing can help acquiring collocations is to provide the needed collocations for the students before the writing process starts and corpora is a place to start finding the needed collocations not only by teachers but by learners themselves.

Contrasting similar collocations in terms of their literal and figurative meaning such as *take measures*– *take measurements* also contrasting collocations with simple verbs are a desirable way of learning collocations. Here, corpus is a means to contrast, compare and discuss the similarities and differences with students. A further important implication for the teaching of collocations that has emerged from the analysis is that it is by no means sufficient to focus on the form of collocations. It is at least as important to focus on the exact meaning and usage of collocations, in particular of verb constructions. Usage includes what has been called ‘semantic prosody’ (e.g., Sinclair, 1999), for example the strong tendency of a collocation to occur only in negative contexts (as *have no intention of*-ing).

Knowledge of the level of style of collocations (e.g., the frequent use of *have a look* when *take a look* would have been more appropriate stylistically) is to some extent essential in producing stylistically acceptable texts, that is knowing which kinds of collocations are used in spoken English, which are used in written English and which are specific to a particular field (e.g., medicine), a fact that has been approved by the students in the present study too. Corpora are great, powerful, easy to use references to teach students these stylistic differences (Aston, 2000).

With respect to corpus use in L2 writing, the participants were favorable overall to its usefulness, despite their various difficulties in using the corpus. They recognized the importance of collocation to successful writing for general purposes, and for formal writing. Further, the participants perceived collocation as the key factor in producing writing that approximates native production. This is consistent with Conrad (2000), who argued that corpora can improve language learning in foreign countries, despite the lack of native speakers in the teaching context. The use of authentic materials can be more beneficial in EFL contexts where natural language use hardly takes place and no target language use is available outside the classroom.

The last but not the least implication of corpora is what Kennedy (1992) notes. He believes that one way in which corpus linguistics can contribute to language teaching is in the selection of those features which seem worth teaching in a given pedagogic context since, in his idea, designing syllabus for ESP and general courses seems to be quite different. While in the case of a course of general English, designing a syllabus may be a very complex operation and may involve a variety of considerations which cannot be limited only to corpus work (McCarthy 2002).

**References**


