

# Can corpora help EFL learners improve their knowledge of collocations?

**Daskalovska Nina, Ivanovska Biljana**

**Abstract:** As collocation is one of the eight types of word knowledge that students need to learn and many experts agree that collocational knowledge enables learners to use the language fluently, language teachers need to devote more time to this aspect of word knowledge and to supplement the exercises in the textbooks with other activities that will help learners gain a better knowledge of collocations. Collocations can be learned effectively with the help of corpora because they can provide valuable information about collocations as well as examples of their use in authentic contexts.

**Key words:** collocations, corpora, concordance lines, corpus-based activities, foreign language learning.

## Introduction

One of the aspects of knowing a word is knowledge of collocations. McCarty (1990: 12) describes the relationship of collocations as a marriage contract between words and says that some words are more firmly married to each other than others. Pawley and Syder (1983) claim that one of the characteristics of native English speakers is their use of a large number of memorized strings of words, which enables them to speak fluently. In the same line, Lewis (2002: 3) maintains that language does not consist of grammar and vocabulary but of multi-word prefabricated chunks.

According to Sinclair (1991: 108), most of the words do not have independent meanings but are components of a rich repertoire of multi-word patterns that make up texts. He says that words combine with each other on the basis of two principles: the open-choice principle and the idiom principle. According to the first principle, language text is the result of a very large number of complex choices and at each point where a unit is completed (a word, phrase, or clause) a large range of choice opens up and the only restraint is the grammaticalness (ibid.: 109). The principle of idiom, on the other hand, 'is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analyzable into segments' (ibid.: 110). He emphasizes that the

idiom principle is far more pervasive than it has been allowed so far, but its importance has been largely neglected.

Schmitt (2000: 77) points to two key factors in relation to collocations. The first one is that words combine with each other, and the second is that these relationships have different degrees of exclusivity. He exemplifies it with the word *blonde* which is used almost exclusively with the word *hair*, and sometimes with the words *woman* and *lady*, but it is never used with the words *paint* and *wallpaper* although they fit together semantically. Collocations can be grammatical/syntactic (*access to*, *acquainted with*) and semantic/lexical (*spend money*, *cheerful expression*). In addition to these basic kinds of collocations, there are collocations that are not based on grammatical or semantic patterning, such as *at six o'clock* or *on Monday*, for which there is no simple explanation because there does not seem to be any logical reason why these prepositions are used in these phrases (ibid.).

Nation (2001: 318) summarizes the discussions about the importance of collocations with the following arguments:

1. Language knowledge is collocational knowledge.
2. All fluent and appropriate language use requires collocational knowledge.
3. Many words are used in a limited set of collocations and knowing these is part of what is involved in knowing the words.

### **The importance of corpora**

How can corpora help learners improve their knowledge of collocations? When learners have a question, they can analyze corpora with the help of a concordancer in order to come to a conclusion. Talking about the importance of concordancers in language learning, Stevens (1995) says: "First, it interjects authenticity (of text, purpose, and activity) into the learning process. Second, learners assume control of that process. And third, the predominant metaphor for learning becomes the research metaphor, as embodied in the concept of data-driven learning (DDL), which builds learners' competence by giving them access to the facts of linguistic performance'.

The basic idea of data-driven learning is that language learning methodology should focus on authentic materials, exploratory tasks and activities that are learner-centered rather than the traditional teacher-led exercises and activities. As Johns (1991: 2) says, "we simply provide the evidence needed to answer the learner's

questions, and rely on the learner's intelligence to find answers... Language-learner is also, essentially, a research worker whose learning needs to be driven by access to linguistic data". Thus, if we provide language learners "with concordance software and a corpus of natural English, [they] can short-cut the process of acquiring competence in the target language, because the computer is able to help students organize huge amounts of language data so that patterns are more easily discerned (Stevens, 1993:11). "What the concordancer does," according to Tribble (1990:11), "is make the invisible visible."

Software programs called concordancers isolate common patterns in authentic language samples, so that language learners can easily focus on the main item of study and they can actively and consciously explore the target language. Such activities help to develop learners' language awareness and understanding of the structure and functionality of the target language, which is 'a new form of grammatical consciousness-raising' (Rutherford 1987). Consciousness raising is defined by Rutherford and Smith (1988: 107) as "the deliberate attempt to draw the learner's attention specifically to the formal properties of the target language". Thus, by engaging in such activities students become active participants in the learning process rather than passive recipients of knowledge.

In many textbooks, the exercises that deal with collocations ask students to match words that usually go together. However, such exercises do not enable students to gain deeper knowledge of the collocations. They only show them one possible collocate of the given words, and since they are given out of context students cannot see how these collocations are used. Moreover, only matching the words that go together does not ensure that students will learn and remember the collocations because students do not need a lot of time and effort to do the exercise, so there is no depth of processing, which is one of the prerequisites for successful learning and remembering the language input. A study that compared corpus-based and traditional activities for learning verb-adverb collocations (Daskalovska, 2009) showed that the participants who were engaged in the corpus-based activity did better in all four parts of the test that was conducted at the end of the experiment ( $F=7,73$ ,  $p<0.05$ ).

Since collocational knowledge is a very important aspect of knowing a word, English language teachers should include activities that will help learners gain a better knowledge of collocations. By using corpora and software programmes, learners will have an opportunity to analyze authentic examples of the language and learn how

certain words and phrases are typically used. The necessary precondition is that students are trained how to use concordance software and how to approach the task of analyzing concordance lines and drawing conclusions. At the beginning these activities can be done in class, but students should also be encouraged to do them on their own at home because they will help them gain valuable knowledge of some important and more frequent words.

### **A corpus-based activity for learning collocations**

The aim of this activity is to find the most frequent collocates of the words *market*, *brand* and *management*. Students will use the on-line concordancer VIEW (<http://corpus.byu.edu/bnc/>). It will allow them to get a list of all the possible noun and adjective collocates of the words *market*, *brand* and *management*, frequency information on the most frequent collocates as well as concordance lines which they can use to see how these collocations are used in context. As the students will need a lot of time to find the information about all the words, they can work in three groups, so that each group can search for information about one word. After they have completed the task, each group will present their findings to the class.

#### *Instructions:*

1. Find the first ten most frequent adjective collocates of the word *market/brand/management* in the British National Corpus. Write them down together with the frequency information (TOT) and the information about the strength of collocations (MI).

2. Now do the same search with the noun collocates of the same word. Check if the noun collocates are used before or after the target word.

3. Generate concordance lines for the collocations.

- a) If you don't know or you are not sure about the meaning of some of the collocations, try to infer their meaning from the contexts.

- b) Look at the words that precede or follow the collocation and identify the most frequently used words or phrases. Can you discern a general grammatical pattern in which the collocations occur?

- c) Choose several concordance lines that clearly exemplify the meaning and use of the collocations.

4. Summarize your results and be prepared to present them to the class.

Some of the possible results that the students will get by searching the corpus are given in the appendices. They show that students will get information about the most frequent noun and adjective collocates of the given words. They can learn whether the collocates are used before or after the target word, how frequent these collocations are, as well as the strength of the collocations. Moreover, they can see authentic examples of these collocations in context, so that they can learn their meaning and how they are usually used. Since students will spend some time exploring all this information by themselves, they will get a better knowledge of the collocations and will be able to form their own conclusions.

### **Conclusion**

This corpus-based activity has several advantages. First, the students will get detailed information about the collocations. Second, the activity is motivational and engaging as students will discover the required information by themselves. Third, because students spend some time analyzing and interpreting the information, which means there is a depth of processing, they will learn and remember the information better. Fourth, this type of activity promotes autonomous learning as it prepares students for self-directed learning. Fifth, it also promotes co-operative learning as students work in groups to come to a solution to a problem.

### **References**

- Daskalovska, N.** (2009). Using corpora to teach collocations. In Beaven, B. (Ed.). *IATEFL 2009 Cardiff Conference Selections*. Kent: IATEFL.
- Johns, T.** (1991). Should you be persuaded: Two examples of data driven learning. In Johns, Tim, and Philip King (Eds.). *Classroom concordancing, ELR Journal 4*. University of Birmingham.
- Lewis, M.** (2002). *Implementing the Lexical Approach: Putting Theory into Practice*. Boston: Thomson.
- McCarthy, M.** (1990). *Vocabulary*. Oxford: Oxford University Press.
- Nation, I. S. P.** (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

**Pawley, A. & Syder, F. H.** (1983). *Two puzzles for linguistic theory: native like selection and native like fluency*. In Richards, J.C. & Schmidt, R.W. (Eds.) *Language and Communication*. London: Longman.

**Rutherford, W.** (1987). *Second Language Grammar: Learning and Teaching*. New York: Longman.

**Rutherford, W. and Smith, M.** (1988). Consciousness raising and universal grammar. in Rutherford, W. and Smith, M. (Eds.) *Grammar and Second Language Teacher: A Book of Readings*. (Pp. 107-116). Boston: Heinle and Heinle.

**Schmitt, N.** (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

**Sinclair, J.** (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

**Stevens, V.** (1993). Concordances as enhancements to language competence. *TESOL Matters* 2, 6:11.

**Stevens, V.** (1995). Concordancing with Language Learners: Why? When? What? *CAELL Journal*, 6, 2: 2-10.

**Tribble, C.** (1990). Concordancing and an EAP writing programme. *CAELL Journal* 1, 2:10-15.

## Appendix 1

*Results for the words management, brand and market.*

MANAGEMENT						
	NOUNS			ADJECTIVES		
	CONTEXT	TOTAL	MI	CONTEXT	TOTAL	MI
<b>1</b>	SYSTEM	454	4.51	SENIOR	431	6.85
<b>2</b>	TEAM	422	5.64	FINANCIAL	240	5.01
<b>3</b>	SYSTEMS	392	5.66	TOP	192	4.19
<b>4</b>	INFORMATION	294	4.09	LOCAL	143	2.78
<b>5</b>	COMMITTEE	223	4.71	NEW	136	1.28
<b>6</b>	GROUP	208	3.48	GOOD	127	1.81
<b>7</b>	NETWORK	178	5.77	GENERAL	111	2.85
<b>8</b>	QUALITY	176	4.59	EFFECTIVE	85	4.25
<b>9</b>	PROJECT	171	4.64	MIDDLE	81	3.83
<b>10</b>	DATABASE	154	6.68	ECONOMIC	80	2.92

BRAND						
	NOUNS			ADJECTIVES		
	CONTEXT	TOTAL	MI	CONTEXT	TOTAL	MI
1	NAME	120	7.20	PARTICULAR	40	5.81
2	NAMES	80	8.36	NEW	23	2.54
3	LOYALTY	23	8.81	SPECIAL	14	4.32
4	IMAGE	16	6.09	INTERNATIONAL	12	4.08
5	LEADER	14	5.58	DIFFERENT	12	2.98
6	MANAGER	14	4.99	UNIQUE	10	6.18
7	AWARENESS	7	5.95	LEADING	10	4.80
8	VALUES	7	4.85	WELL-KNOWN	8	7.37
9	IDENTITY	6	5.58	STRONG	6	3.57
10	SHARES	6	4.48	CERTAIN	5	2.86

MARKET						
	NOUNS			ADJECTIVES		
	CONTEXT	TOTAL	MI	CONTEXT	TOTAL	MI
1	LABOUR	1041	5.96	SINGLE	482	5.42
2	SHARE	666	6.33	FREE	411	5.01
3	STOCK	662	7.06	COMMON	364	5.02
4	RESEARCH	491	4.89	EUROPEAN	326	4.70
5	PLACE	490	4.07	INTERNAL	276	6.06
6	ECONOMY	408	5.98	OPEN	245	3.75
7	FORCES	381	5.72	DOMESTIC	181	5.42
8	VALUE	363	5.05	BLACK	158	3.42
9	HOUSING	348	5.90	COMPETITIVE	143	5.97
10	MONEY	312	3.78	NEW	120	0.65

## Appendix 2

*Concordance lines for the the word **market**.*

...figures in the next paper illustrate a **labour market** situation in Shropshire in which ...

...I would have thought the **labour market** conditions are also a vital ingredient...

...a discussion of women's part in the **labour market** in the twentieth century which ends...

...our businesses have improved their **market share** during this recession and I'll mention ...

...for increased sales and increased **market share** looking into this year and into the last...

...we've im improved our **market share** in that area too...

...has been done on the impact of the **single market** on voluntary organisations...

...opportunities that will come with the **Single Market**, is regarded as particularly important..

...introduced as part of the **single market** legislation, will generally cost UK...

**Mail/contact address:**

Name: Nina Daskalovska, Biljana Ivanovska

Faculty: Faculty of Philology

University: Goce Delcev University, Stip, Republic of Macedonia

Address: Nevena Stojkova 17-1/4, 2400 Strumica, Republic of Macedonia

E-mail: [nina.daskalovska@ugd.edu.mk](mailto:nina.daskalovska@ugd.edu.mk)