

CORPUS LINGUISTICS IN TEACHING ENGLISH AS A FOREIGN LANGUAGE

О.В. Павленко

Національний технічний університет України "КПІ"

Teaching English as a foreign language (EFL) today requires using more and more technology. In recent decades a lot of investigation has been devoted to how computers can facilitate language learning. One of the promising directions is application of corpus linguistics in EFL teaching.

To begin with, corpus linguistics has gained a lot of popularity among different scholars. It should be noted that before the 20th century, it took years for linguists all over the world to collect data about rules of word formation and texts, its functions and interaction with other languages. All data was collected manually. Nowadays, corpus linguistics investigates language as expressed in corpora. In turn, corpora include samples of the text taken from the natural context, and the text now stored and collected virtually. [2]. Current research shows that today we are able to use the so-called "second generation" of English language general corpora, which include the British National Corpus (BNC); Corpus of Contemporary American English (COCA); and COBUILD Bank of English. Modern corpora use do not follow the strict rules of the language, they deal with intuition of great number of speakers along with the linguistic analysis. This gives a higher level of objectivity in the creation of datasets.

In recent years EFL teachers are using the potential of corpora in their teaching. Leech (1997) observed that "a convergence between teaching and language corpora was apparent". The author notes that the convergence has three focuses. He outlines the indirect use of corpora in teaching as the first focus and exemplifies it by such directions as reference publishing, materials development, and language testing. The second focus is direct use of corpora in teaching i.e. teaching, teaching to exploit, and exploiting to teach. The third focus lies in further teaching-oriented corpus development (LSP corpora, L1 developmental corpora and L2 learner corpora)[3].

Taking into account the ideas given above, let us consider advantages of using corpus linguistics in teaching EFL. Typically, all corpora are created and updated by linguists, teachers, translators, and other researchers. Second generation corpora include millions of words which are collected from spoken and written communication, analyze

data from fiction, popular magazines, newspapers, academic and professional text. For example, the COCA's interface allows learners to search for exact words or phrases, wildcards, lemmas, part of speech, or any combinations of these. It is possible to search for surrounding words (collocates) within a ten-word window (e.g. all nouns somewhere near *faint*, all adjectives near *woman*, or all verbs near *feelings*), which often gives you good insight into the meaning and use of a word [3].

Corpora also give teachers and learners a great possibility to limit searches by frequency. The function of the word frequency comparison as well as phrases, and grammatical constructions can be selected according to the following criteria:

- by genre: comparisons between spoken, fiction, popular magazines, newspapers, and academic, or even between sub-genres (or domains), such as movie scripts, sports magazines, newspaper editorial, or scientific journals;
- over time: compare different years from 1990 to the present time.

It is also easy to carry out semantically-based queries of the corpus. For example, it is possible to contrast and compare the collocations of two related words (*little/small*, *democrats/republicans*, *men/women*), to determine the difference in meaning or use between these words. You can find the frequency and distribution of synonyms for nearly 60,000 words and also compare their frequency in different genres, and also use these word lists as part of other queries. Finally, you can easily create your own lists of semantically-related words, and then use them directly as part of the query [4].

Teachers can use corpora for curriculum and materials design to provide students with up-to-date materials. In turn, students can use corpora for reading texts on general and professional topics, for learning vocabulary in natural context and for writing lesson preparation. Corpora allow ESP teachers and learners to easily use concordances to search for collocations in their professional field as well as to find possible usages of terminology. Let us consider the search results for the word *potentiometer* in Electrical Engineering corpus (Table 1):

Table 1

001	ange of 2 to 1000 utilizing a 100-kQ variable resistance (a	POTENTIOMETER	or "pot" for short). Solution It is usually p
002	onent Common symbols 1 fixed resistor 2 variable resistor 3	POTENTIOMETER	4 thermistor 5 LDR 6 capacitor 7 inductor 8 tr
003	or these terms? 1 amplifier 2 video recorder 3 television 4	POTENTIOMETER	5 coaxial cable Task 9 What terms are represen
004	ment that is typically used with general-purpose op amps. A	POTENTIOMETER	is connected between the offset-nulling termin
005	pa(r)/ [11 moving contact in a variable component such as a	POTENTIOMETER	or variable resistor wireless /w aialas/ see r
006	ectrolytic capacitor connected in series with a five-kilohm	POTENTIOMETER	(pot). The positive terminal of the capacitor
007	cuit which divides a voltage into two or more smaller parts	POTENTIOMETER	/pa,tensi omitta(r)/ [11 variable electronic co

008	,stju:dtau/ [7] portable recording studio pot /pot/ [1] see	POTENTIOMETER	potential difference /paten fl 'd ifrans/ see v
009	between the offset- ulling terminals with the wiper of the	POTENTIOMETER	connected to the op- amp negative supply. Movin
010	imeter connected to the op- amp negative supply. Moving the	POTENTIOMETER	wiper introduces an imbalance that counteracts

The results given by the Electrical Engineering corpus show first ten possible combinations of the word *potentiometer* with different parts of speech, positions in the sentence as well as natural and professional contexts. With this table students will be able to determine the potential collocations by themselves, useful phrases (e.g. *a potentiometer or "pot" for short; connected to the op-amp negative supply*), choose the detailed examples and learning more context by pressing the key word in search results [5]. Barlow (1992) suggests that a corpus and concordance can be used to compare language use – student/native speaker, standard English/scientific English, written/spoken [1]. EFL/ESP teachers may benefit from this resource by analyzing the language in books, readers, and course books, to create generate exercises and activities, analyze and compare the appropriate context and usage e.g. *obtain vs. get*.

However, corpus-based method has several disadvantages. Firstly, as the data is stored in the web, there is a dependence on the Internet connection. Secondly, it requires some training for teachers and students as it is not a ready-made resources. Thirdly, despite the fact that corpora provide specific contexts (Electrical Engineering, Academic), there is still the need to develop narrower collections of words for ESP contexts.

In conclusion, the given outline of the possible applications of corpora in EFL and ESP represents the idea of their application in the classroom. The future studies in this field may include the deeper research of corpora and its application in ESP, teacher training and applied linguistics.

RESOURCES:

1. Barlow, Michael (2002) Corpora, concordancing, and language teaching. Proceedings of the 2002 KAMALL International Conference. Daejon, Korea
2. Biber, D., Conrad, S., Reppen R. (1998) Corpus Linguistics, Investigating Language Structure and Use, Cambridge: CUP.
3. The Corpus of Contemporary American English (COCA) [Электронный ресурс]. – Режим доступа: <http://corpus.byu.edu/coca/x.asp?r1=&w=1093&h=614>. – Назва з екрану.
4. Leech, G. (1997) 'Teaching and language corpora: a convergence' in A. Wichmann, S. Fligelstone, T. McEnery and G. Knowles (eds.) Teaching and Language Corpora, pp. 1-23. London: Longman.
5. Willis, Jane (1998) Concordances in the classroom without a computer, In Brian Tomlinson (Ed.) Materials development in language teaching, Cambridge