

CORPUS VS DICTIONARY IN EFL CLASSES

Ivana Cimermanová

Abstract: Dictionaries (bilingual, monolingual, visual, specific etc.) have become a natural aid for learners of English, teachers of English, translators, or simply speakers/users of language. Contemporary dictionaries are based on corpora. This fact can evoke a question – Can we use corpora in the classroom as a substitute of a dictionary. Can it be even more effective? The author tries to bring some possibilities of using electronic corpora in the language classroom to “substitute” dictionary.

Key words: corpus, dictionary, search, context, word meaning, word grammar

Introduction

Most of us have ever faced a question which dictionary to buy. For some people the basic criterion is a price, for others number of words or the way of the word explanation. There are many more criteria that should be considered when selecting a dictionary. It is interesting to observe how an increasing group of language users is using corpora as a first solution rather than consulting an unknown word with the “traditional” dictionary.

Dictionaries

The most important criterion for dictionary selection is a purpose of its use. Do we need it to learn the basic phrases used during holiday or do we need to translate the schemes of electronic circuits or probably to translate the results of speleological research. The next crucial factor is the user himself, especially his language level.

The third significant criterion is a number of headwords – generally the dictionaries are divided according to the number of headwords to the dictionaries for native speakers (200 000 - 400 000), for advanced learners of English (60 000- 100 000), for intermediate learners of English (30 000 – 40 000), for beginners (1 500 – 2 500) many times supported by various photographs or pictures).

What is equally important is defining vocabulary – number of words that is used to explain the headwords; idiomatic expressions and the sample sentences where user usually can learn a lot about the word grammar. The team of authors, as well as edition of a dictionary gives also certain information about a dictionary.

We can mention a number of other criteria that a buyer can/should consider, as e.g. does it provide pronunciation/spelling variants (American/British), year of publishing, user-friendly navigation, binding, paper quality, size, weight, etc..

Crystal (1995, p. 108) defines dictionary as “a reference book that lists the words of one or more languages, usually in alphabetical order, along with information about their spelling, pronunciation, grammatical status, meaning, history and use”. Different dictionaries serve for different purposes – there are terminological dictionaries, dictionaries of synonyms, antonyms, dialects, jargons etc.

Dictionaries are an important teaching/learning aid in language teaching. Students use it frequently and should master the dictionary skills that are not always trained, what more, teachers expect learners to use them effectively when they come to the language classes. Good dictionary is a source of a lot of information about a word, its meaning, grammar, collocations, idiomatic expressions, etc.

Corpora

The word corpus has several meanings in English. In linguistics we understand the term corpora as "... a collection of texts assumed to be representative of a given language put together so that it can be used for linguistic analysis. [L]anguage stored in a corpus is naturally-occurring, that it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology." Tognini – Bonelli (2001, p.2). Sinclair (1991, p. 171) defined corpus as "a collection of naturally-occurring language text, chosen to characterize a state or variety of a language". In 2005 he used more precise definition of a corpus as "A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research." This definition comprises all characteristics of electronic corpus - pieces of language text, electronic form, external criteria, to represent a language, linguistic research.

Many academics describe corpora as an electronic collection of the text, e.g. Baker et al define a corpus as "a collection of texts (a 'body' of language) stored in an electronic database. Corpora are usually large bodies of machine-readable texts containing thousands or millions of words" (Baker et al (2006, p.48). That would be true in these days, however in past the linguists worked with corpora manually.

History of corpora building dates back to the late 19th century. "In 1897, German linguist J. Kading used a large corpus consisting of about 11 million words to analyse distribution of the letters and their sequences in German language." (History of Corpus linguistics, 2012, ¶2) Before 1950s Franz Boas compiled small corpora to analyse the phonological aspects of the Inuit language, and there are also other examples of using corpora in the so-called pre-Chomsky period (L. Bloomfield, A. Hill, etc.). Some of them, "Fries and A. Aileen Traver also started to use corpus in pedagogical study of foreign language" (History of Corpus linguistics, 2012, ¶3). The first electronically readable corpus was compiled in 1960s and is known as Brown corpus (Kucera and Francis from Brown University started to work on it on 1961 – it compiles 1 million words and texts are divided into 15 categories).

Use of electronic tools since 1980s dramatically and significantly changed the position of corpus linguistics (that was in fact rejected by Chomskyan linguistics) that is lately being introduced even to the language classroom and brings discovery techniques to language learning and teaching.

Nowadays, there are various corpora that can be used in a classroom; researchers (and teachers) can build their own corpora. Among the largest and well known English language corpora belong: American National Corpus (22 million words), British National Corpus (100-million-word text corpus), Corpus of Contemporary American English (425 million words), Bank of English (650 million running words).

There are different types of corpora (written vs. spoken, diachronic vs. synchronic, plain vs. annotated, monolingual vs. multilingual) and the texts are categorised (different corpora vary in categories but generally you can focus your searches and specify the subcorpora e.g. according to genre, register, style, etc.

Frequently we use web instead of a dictionary – e.g. if we are not sure about spelling or to make sure a certain word exists in target language. That is in fact using "corpora" as web is a free collection of expanding electronic texts (however without categories and it one cannot use corpus tools for analysis).

Corpora are used by lexicographers, computational linguists, for different literary studies, translation practice and it is slowly being introduced to language teaching and learning. Here we deal with a term data-driven learning (DDL), that is defined by Payne (2008) as the "application of tools (concordancers) and techniques from corpus linguistics in the service of language learning." The main benefits of DDL is that students work with

authentic materials to study new words, grammar structures and they are focussed “on real, exploratory tasks and activities rather than traditional «drill & kill» exercises”. (Rüschhoff, no date). A valuable source for data-driven learning (and teaching) can be found online – Compleat Lexical Tutor (www.lextutor.ca) that offers various tools that can be used.

Dictionary vs. Corpora

Different methods are applied in language teaching with the effort to make educational process effective. Humanistic approach, learner-centred teaching, building and developing leaning strategies, supporting individual and autonomous learning, etc. All these words are connected with key competences for lifelong learning – namely communication in foreign languages, mathematical competence and basic competences in science and technology, digital competence and learning to learn. DDL is one of the ways how to teach students to think creatively and linguistically (in our case), how to work individually, not to rely on teacher and rather to look for information on one’s own.

Thanasoulas (2000) describes positivism as one of the dominant philosophies of learning in the twentieth century, and compares it to ““traditional classroom,” where teachers are the purveyors of knowledge and wielders of power, and learners are seen as 'container[s] to be filled with the knowledge held by teachers’”. The traditional approach is opposed to positivism, where the “widespread notion that knowledge is attained by dint of the 'hypothesis-testing' model, and that it is more effectively acquired when 'it is discovered rather than taught’” (ibid.) He also stresses that it is related to self-direction and self-evaluation.

Constructivism “[i]n contrast to positivism, constructivism posits the view that, rather than internalising or discovering objective knowledge (whatever that might mean), individuals reorganise and restructure their experience“... “constructivism 'leads directly to the proposition that knowledge cannot be taught but only learned (that is, constructed)', because knowledge is something 'built up by the learner’”. Thanasoulas (2000) highlights that “constructivist approaches encourage and promote self-directed learning as a necessary condition for learner autonomy”.

I have already discussed dictionaries as a teaching/learning aid and we defined corpora as a source of data mainly for researchers but with possibility to use them in a classroom as well. Using corpora in the language classroom has been discussed lately by e.g. Reppen, R., Sinclair, J., Belles-Fortuno, B., Gea-Valor, M., O’Keefe, A., McCarther, M., and others, but as I have already mentioned CH.C.Fries and A.A. Traver.

The results of using dictionaries and corpora can be presented on the following example - activity:

Choose the correct word

Dear Ms. Dawning

*I am writing to **apply** for/to the job of sales assistant advertised in last week’s Messenger. I think I would be **suitable** to/for this for a number of reasons. Firstly, my present job involves helping customers, so I am **good** in/at dealing with people. Although I do not have much experience of sales work, I learn quickly. I also have a good **knowledge** in/of the range of products that you sell. I would be available to attend an interview at any time.*

I look forward to hearing from you soon.

Yours sincerely,

J. Jones

(sample based on Longman General Dictionary Worksheets)

In case students do not know the answer they will probably consult a dictionary, we can presume they will use monolingual dictionary for learners of English. Depending on their language level they choose the most appropriate dictionary (as an example I can mention Oxford dictionaries – OBED – Oxford Basic English Dictionary, OSD - Oxford Student's Dictionary of English, OALD - Oxford Advanced Learner's Dictionary).

The Longman Dictionary of Contemporary English gives the following information about the verb apply:

- ap•ply** *past tense and past participle* applied, *present participle* applying, *third person singular* applies
- 1 **request** [intransitive] to make a formal request, usually written, for something such as a job, a place in a university, or permission to do something
 - apply for**
 - She applied for a job with the local newspaper.
 - We need to apply for planning permission to build a garage.
 - apply to**
 - I applied to four universities and was accepted by all of them.
 - 2 **affect** [intransitive, transitive not in progressive] to have an effect on or to concern a particular person, group, or situation
 - apply to**
 - Do the same rules apply to part-time workers?
 - The offer only applies to flights from London and Manchester.
 - 3 **use** [transitive] to use something such as a method, idea, or law in a particular situation, activity, or process
 - apply something to something**
 - New technology is being applied to almost every industrial process.
 - 4 **apply yourself to** work hard at something, especially with a lot of attention for a long time:
 - Stephen would do well if only he applied himself.
 - 5 **make something work** [transitive] to make something such as a piece of equipment operate, usually by pushing or pressing something:
 - apply the brakes
 - 6 **spread paint/liquid etc** [transitive] to put or spread something such as paint, liquid, or medicine onto a surface:
 - Apply the cream evenly over the skin.
 - 7 **apply force/pressure** to push on something
 - 8 **use a word** [transitive] to use a particular word or name to describe something or someone:
 - The term 'mat' can be applied to any small rug.

I mentioned that the results might differ and vary in different dictionaries. Below there can be found the result of online OALD:

Apply /verb/applies, applying, applied, applied for job/course

1 [intransitive, transitive] to make a formal request, usually in writing, for something such as a job, a place at college, university, etc

- You should **apply in person/by letter**.
- apply for something to apply for a job/passport/grant
- apply to somebody/something (for something) to apply to a company/university
- apply to do something He has applied to join the army.

use

2 [transitive] to use something or make something work in a particular situation

- apply something to apply economic sanctions/political pressure
- apply something to something The new technology was applied to farming.

paint/cream

3 [transitive] apply something (to something) to put or spread something such as paint, cream, etc. onto a surface

- Apply the cream sparingly to your face and neck.
- The glue should be applied to both surfaces.

be relevant

4 [intransitive, transitive] (not used in the progressive tenses) to concern or relate to somebody/something

- Special conditions apply if you are under 18.
- apply to somebody/something What I am saying applies only to some of you.
- apply something to somebody/something The word 'unexciting' could never be applied to her novels.

work hard

5 [transitive] to work at something or study something very hard

- apply yourself You would pass your exams if you applied yourself.
- apply yourself/something to something/to doing something We applied our minds to finding a solution to our problem.

press hard

6 [transitive] to press on something hard with your hand, foot, etc. to make something work or have an effect on something

- apply something to apply the brakes (of a vehicle)
- apply something to something Pressure applied to the wound will stop the bleeding.

Students consulting the dictionaries can find immediately the answer along with grammar. In case of good dictionary skills they can find the answer quite quickly, both dictionaries introduce the item *apply for a job*. Grammar is included and information about the verb, the user can also find a sample sentence and to see how the verb functions in a sentence.

To find a correct answer students can also use different types of corpora as e.g. British National Corpus and after typing a query *apply* they can study the concordances (concordance lines). One can click a word to find more about the context and to get larger context. See the following figure with the BNC screen:

The screenshot shows the J-BNC interface with a search for the word 'apply'. The results are displayed in a table with columns for document ID, word, part of speech, and context. The word 'apply' is highlighted in red in the context lines.

ID	Word	POS	Context
4	G33	W_fict_prose	(6) and (6) of that section shall apply accordingly (2) This section shall apply to any premises
5	HXD	W_ac_polit_law_edu	employed by an agency . In this case the same precautions apply and it is even more important to deal with ownership of copyright
6	H83	W_misc) . The airlines usual liability , although limited , will apply and we will be able to obtain a recovery . It is
7	AMM	W_ac_nat_science	mysterious organisms live ? In this case it is difficult to apply argument from analogy , because the graptolites are really
8	BP5	W_commerce	behaviour . # National courts will have full powers to apply Article 88 under the Continental Can doctrine , to any merger
9	AYR	W_misc	than American Express . This is an especially good time to apply as we are offering you the chance to win two tickets to
10	B14	W_non_ac_medicine	. What has already been said about thermal hazards can also apply at work Where automatic stoking equipment is not in use ,
11	HXD	W_ac_polit_law_edu	. In practice , these provisions will be very difficult to apply but the preamble to the EC Directive may give some assistance as
12	J71	W_ac_polit_law_edu	he is normally resident in the UK and they can therefore apply even if he is physically outside the UK at the relevant time
13	K4W	W_newsp_other_report	wo n't get a free beer as with Wainwrights but can apply for a badge . The booklet can be obtained by sending 1.80
14	HW5	W_fict_prose	be able to devise a way round my official non-existence and apply for a gun myself , though even then , all things considered
15	FRN	W_commerce	, a sector which is being undermined by imports may successfully apply for exemption from restrictions on the formation of a cartel to
16	GXG	W_misc	send on your behalf . We understand that Newport BC will apply for the Order to build the barrage on the 28th of May
17	HRB	W_biography	' naturalists ' . When Gilbert decided to leave London and apply for the post of curator at the museum of the newly established
18	J7A	W_ac_polit_law_edu	second limb of s740(5) states that s65(6) (9) shall apply for the purposes of s740 as they would apply for the purposes
19	HA3	W_non_ac_soc_science	which type of work interests you before you decide where to apply for training A large number of firms offer general experience ,
20	BLW	W_religion	buy a motor motor bike , what college or university apply for what job to look for , a calm discussion
21	KP1	S_conv	(unclear) and do a good job ! (SP:PS50T) Royalty do n't apply for (unclear) (SP:PS50X) I du n no , they just (pause)
22	HL7	W_non_ac_polit_law_edu	be met only by using three-way catalytic converters and would apply from mid-1992 for new petrol-driven cars and from Jan. 1,1993 ,
23	HB4	W_misc	business remains very competitive but it has been necessary to apply further selective premium increases . A more selective
24	CB8	W_religion	for some time in mid-week housegroups is to follow up and apply further what has been taught on Sunday . This means that the
25	H8T	W_fict_prose	. Then he would be up and about , able to apply himself unfinished business . He tugged at the sheet and

Figure 1 BNC screen - (KWIC) search - apply

Using corpora let students to deduce the meaning and grammar rules based on the authentic texts, many times it might be more time consuming, however “aha-method” – discovery technique leads to understanding and more permanent association.

Students frequently have problems with collocations, and word choice, use of synonyms. Concerning collocations, they are not aware of the existence of dictionary of collocations and they mostly use monolingual or bilingual dictionaries. Corpora can very quickly bring collocates.

The screenshot shows the BYU-BNC interface with a search for the word 'coffee'. The results are displayed in a table with columns for rank, context, and frequency statistics.

Rank	Context	ALL	ALL	%	MI
1	CUP	533	11774	4.53	7.45
2	TEA	388	7942	4.89	7.56
3	TABLE	145	19128	0.76	4.88
4	CUPS	117	1234	9.48	8.52
5	SHOP	117	10310	1.13	5.46
6	MORNING	110	19846	0.55	4.42

Figure 2 BNC screen - collocations - the word coffee (+2)

DISPLAY	SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT] [HELP...]	COMPARE	SIDE BY SIDE		
	CONTEXT	ALL	ALL	%	MI
1	HOT	428	58546	0.73	4.12
2	INSTANT	220	13147	1.67	5.31
3	FRESH	156	42895	0.36	3.11
4	STEAMING	108	2207	4.89	6.86
5	ICED	90	1516	5.94	7.14
6	ORGANIC	68	12778	0.53	3.66
7	TURKISH	56	5633	0.99	4.56
8	IRISH	50	12153	0.41	3.29
9	BITTER	48	9504	0.51	3.59
10	POURING	45	4850	0.93	4.46

Figure 3 BNC screen - collocations - the word coffee (adjectives +-2)

At this place I can also mention wordnetbrowser what is a system where lexicographer files organize nouns, verbs, adjectives and adverbs into groups of synonyms, and describe relations between synonym groups. These are organised lexically and semantically. E.g. concerning nouns the relations among words include hypernymy/hyponymy, synonymy/antonymy, entailment, and meronymy/holonymy. The system brings also derivative forms. The tool might be useful for students to realise relations among the words and (semantic) mapping is one of the cognitive strategies that is used in learning process and helps learners.

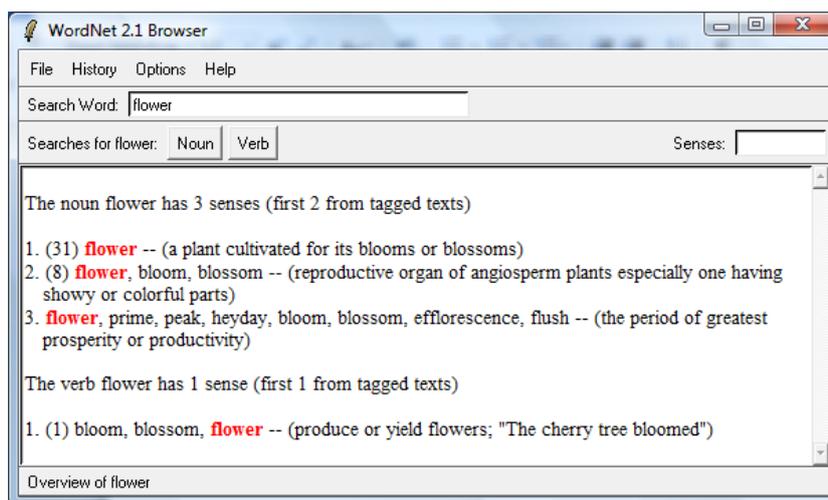


Figure 4 Wordnetbrowser output - search word flower

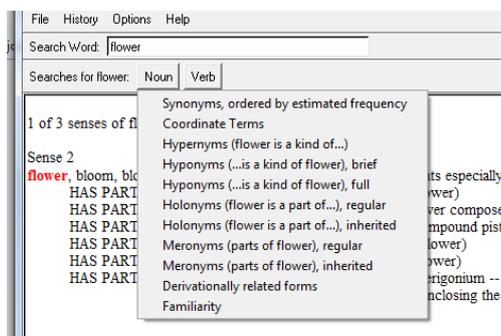


Figure 5 Wordnetbrowser output - Noun meronymy - word flower

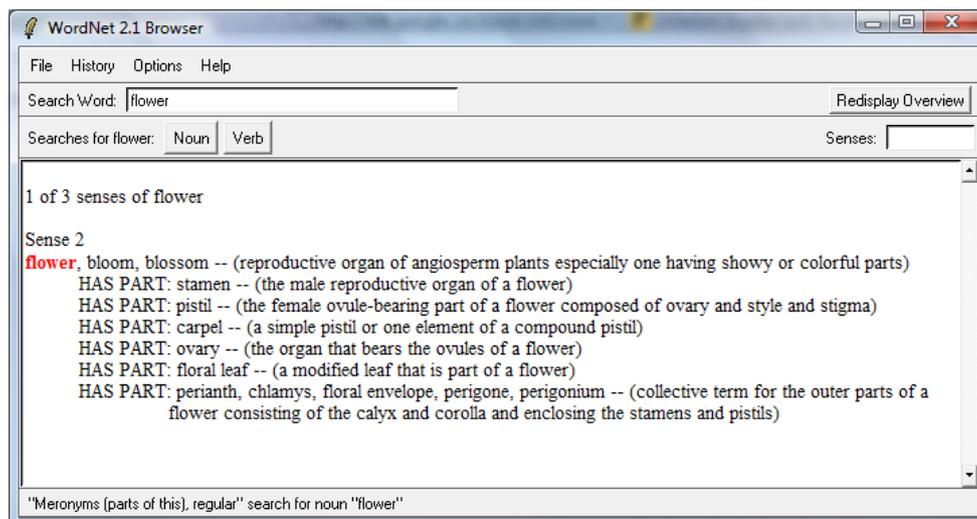


Figure 6 Wordnetbrowser output - meronymy word flower

Digitalisation has a significant influence on language – using emoticons, abbreviated/acronymed words (young generation uses the abbreviated words in speech – e.g. in Slovak language jj (instead of áno - yes); ee (instead of nie – no); phrases in English e.g. lol (for laughing out loud); JAM (for just a minute), L8R – (later) or words in English e.g. 4 – for, NE – any, M-am, etc.

The changes penetrate language rapidly and dictionaries and language even though they use corpora they reflect vocabulary of a certain period of time.

A former dot-commer working a McJob was listening to some headbangers while laying out the last of his dead presidents for longnecks and some less than heart-healthy Frankenfood.

This sentence was used to promote 11th edition of Merriam-Webster Collegiate Dictionary in June 2003. They built it (not only) on presenting newly created words and their presence in their latest edition of dictionary.

“A language has at a given time a finite inventory of words, the meanings of which are revealed in the course of general usage.” (Quirk in: Mouritsen, S. C., 2010)

We tend to use authentic materials in the classroom teaching learners to overcome linguistic barriers. Thus corpora are a valuable source of authentic material (in Lextutor there is a possibility to select the text K1000 or K2000 – according to our needs and students’ abilities and skills). Authentic material help students to be ready for a real language that is not adopted, language that is really contemporary.

The words from the advert above (e.g. dot-commers and longnecks) cannot be found in British National Corpus, however there are dozens of occurrences in Corpus of Contemporary American English.

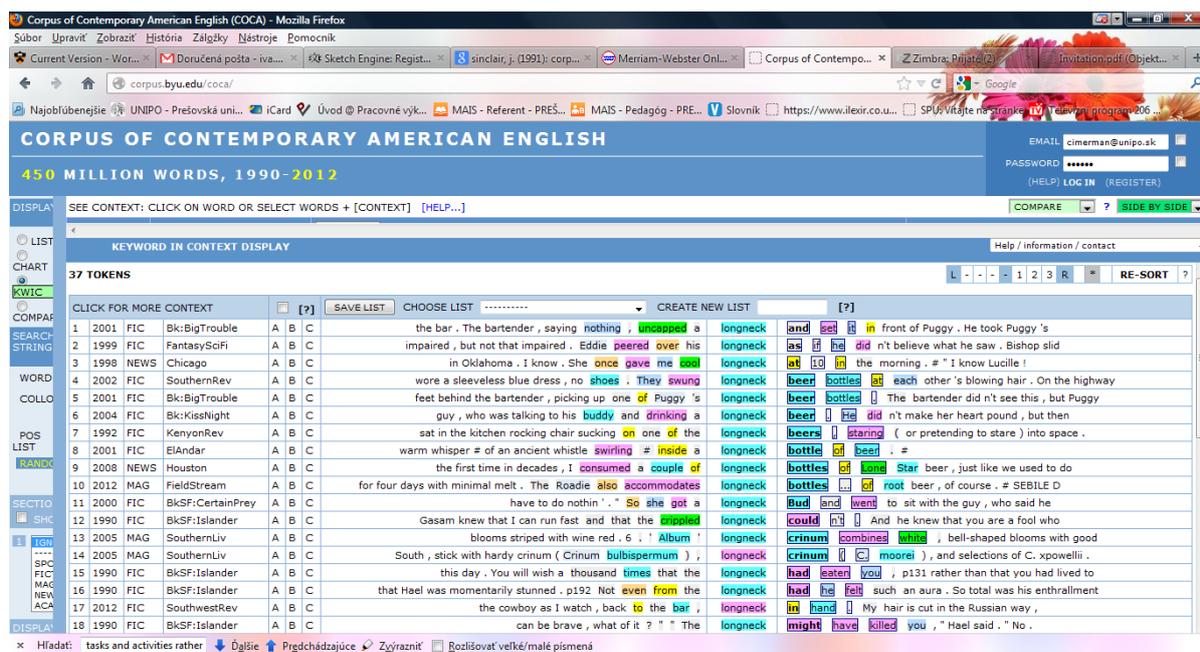


Figure 7 Screen of COCA - search longneck (37 tokens)

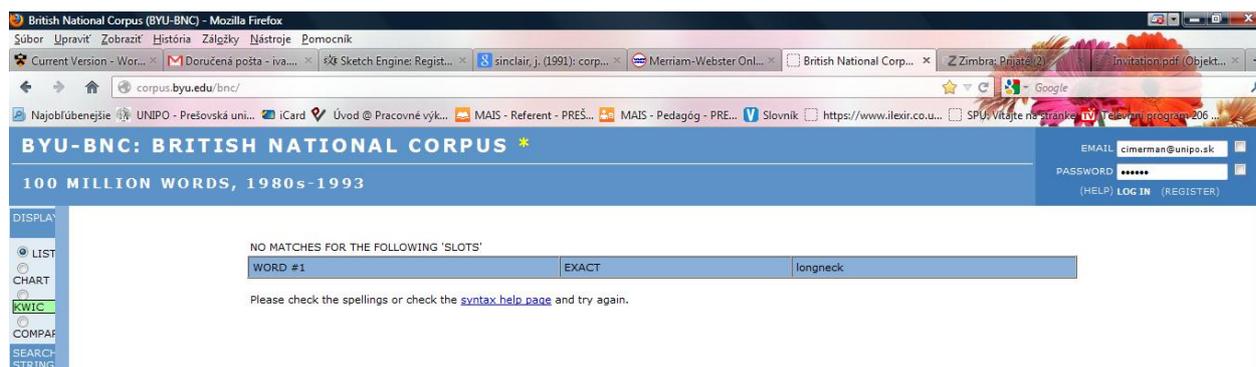


Figure 8 Screen of BNC - search longneck (no matches)

I mentioned Merriam Webster dictionary, however, that dictionary is for native speakers and the word longneck probably cannot be found in any of the dictionaries for learners of English. As we could see also in the BNC we could not find the samples of the use of this word, however COCA gave us 37 tokens and based on the concordances and the context we can be able to guess the meaning. The most frequent collocates are bottle and beer what might be very important information for the learner.

Conclusion

The aim of the presented article was not to present the idea either...or, but rather to present few different possibilities that dictionaries and corpora can offer. Corpora are not used in our schools, similarly as dictionary skills are not developed systematically at our schools. Knowing the meaning of the words is not enough, one must be able to use it in conversation and writing and to use in appropriate context and grammatically correctly. Knowing different types of dictionaries may make learners to be more effective especially in writing can later positively influence their communicative skills (both fluency and accuracy).

Using corpora allows learners to see the word in different contexts, study collocations at the same place and it does not give a direct answer but rather force students to think

linguistically and deduce the meaning of the word what may help them to overcome linguistic barriers without panicking using different learning strategies, especially compensation strategies – as deducing meaning from context, guessing based on partial knowledge of the target language, etc.

Both, dictionaries and corpora are valuable source for language teaching and combining them in the classroom can help learners to develop their communication skills, to build their self-confidence, builds their language ego, teaches them to take risks and this can have a positive influence on their language (and social) competence.

References

- Baker, P., Hardie, A., McEnery, T. (2006): *A Glossary of Corpus Linguistics*. Edinburgh University Press. 2006. 187pp., ISBN -10 0 7486 2018 4
- Crystal, D. (2003): *The Cambridge Encyclopedia of the English Language*. CUP, ISBN 9780521530330
- History of Corpus linguistics* (2012): available at: <<http://www.cl2011.org.uk/history-of-corpus-linguistics.html>>, retrieved: November, 1st, 2012.
- Longman General Dictionary Worksheets*, available at: <<http://www.pearsonlongman.com/loginNowRedirect.asp?redirectTo=/dictionaries/members/pdfs/gen01.pdf>>, retrieved: October, 27th, 2012
- Mouritsen, Stephen C. (2011): *The Dictionary Is Not a Fortress: Definitional Fallacies and a Corpus-Based Approach to Plain Meaning* (February 1, 2011). Brigham Young University Law Review, p. 1915, 2010. Available at SSRN: <<http://ssrn.com/abstract=1753333>>, retrieved: November, 1st, 2012.
- Payne, J. S. (2008): *Data-Driven South Asian Language Learning*. Available at: <<http://salrc.uchicago.edu/workshops/sponsored/061005/DDL.ppt>>, retrieved: October, 27th, 2012
- Rüschhoff, B. (no date): *Data-Driven Learning (DDL): the idea by Bernd Rüschhoff*. Available at: <http://archive.ecml.at/projects/voll/rationale_and_help/booklets/resources/menu_booklet_ddl.htm>, retrieved: October, 27th, 2012
- Sinclair, J. (1991): *Corpus, Concordance, Collocation*. London, England: Oxford University Press. 1991, ISBN 0 19 437144 1
- Sinclair, J. (2004): *Trust the text: Language, corpus and discourse*. London, England: Routledge. ISBN-10: 041531768
- Sinclair, J. (2004): *How to Use Corpora in Language Teaching*. Amsterdam; Philadelphia: John Benjamins B.V., 2004, ISBN 90 272 2282 7
- Straková, Z. (2003): *Vedíme žiakov k samostatnosti*. Prešov : MPC v Prešove, 2003. ISBN 80-8045-321-7.
- Thanasoulas, D. (2000): What is Learner Autonomy and How Can It Be Fostered? In: *The Internet TESL Journal*, Vol. VI, No. 11, November 2000, available at: <<http://iteslj.org/Articles/Thanasoulas-Autonomy.html>>, retrieved: October, 27th, 2012
- Tognini-Bonelli, E. (2001): *Corpus linguistics at work*, Amsterdam: John Benjamins, 2001, ISBN 1 58811 061 3

Author

PaedDr. Ivana Cimermanová, PhD., The Institute of British and American Studies, Faculty of Arts, The University of Prešov, Prešov, Slovakia, e-mail: ivana.cimermanova@unipo.sk