Contemporary Research in Language and Linguistics
Volume: 1 Issue: 2

New Web-Based Corpus Tools for Language Teaching and Learning

Emin Idrizi[1]

Submitted:
02 November 2023

Accepted:
21 November 2023

Published:
30 November 2023

Abstract

Over the past decades, various corpora and corpus interfaces have been developed that have mainly served linguists and lexicographers for the purpose of language exploration and analyses. These interfaces have also been aimed at language teachers for language teaching purposes as well as language learners who wanted to engage in corpus consultation to gain knowledge on various aspects of the target language, such as grammar, words in context, word collocation, among others. The standard and traditional corpus interfaces, however, have not always proven to be easy tools for teachers and learners given that they require sufficient training before they can be used effectively. To overcome these challenges, a number of alternative web-based corpus tools have been introduced having language teachers and learners in mind. In this paper we examine and evaluate two state-of-the-art corpus tools: SKELL and Netspeak.

Keywords: Web-based corpus tools, corpus interface, SKELL, Netspeak, concordance lines, grammar, collocation

Cite as: Idrizi, E. (2023). New Web-Based Corpus Tools for Language Teaching and Learning. *Contemporary Research in Language and Linguistics, 1*(2). 71-79.

[1] Assist. Prof. Dr., International Balkan University e.idrizi@ibu.edu.mk
https://orcid.org/0000-0002-1924-2363

Introduction

Computer corpora and computer programs capable of analyzing corpus data were initially designed to serve linguists and lexicographers for the purpose of language investigation and language analyses. With the continuous development of technology, computers became affordable, and as a result, corpus interfaces started to become accessible for a wider range of profiles, including language teachers and learners. Tim Johns (1991) coined the term Data-driven learning or DDL referring to the type of language learning technique which had language learners, with some teacher help, discover the rules of language themselves using corpus data, while he called each learner a "Sherlock Holmes" (Johns, 1997, p. 101) indicating that every learner could be a language investigator given that the learner himself was able to access corpus data and investigate the rules of grammar and vocabulary.

Since then, corpora and corpus interfaces have been continuously evolving and advancing from being simple tools that do simple searches and generate simple concordance lines to advanced tools that can investigate and analyze language in more depth, such as provide reliable information on how words and phrases are correctly used in real context; provide accurate information on syntax and word patterning; find collocations for words, to name just a few. These have been made possible thanks to well-known corpus interfaces and concordancers, such as Sketch Engine, COCA, IntelliText, among others. Research, on the other hand, has been extensive in exploring the potential of using corpora for language teaching and learning as well as DDL as a technique. Research in general indicates that corpus consultation and DDL can be beneficial to learners. Some studies, for instance, show that corpus consultation plays a positive role in error correction in writing (Mull, 2013; Luo and Liao, 2015), while learners have been reported to have positive attitude towards the use of corpora for language learning (Boulton, 2010).

Corpus interfaces and concordancers, however, have not always been proven to be easy tools for learners to use, considering that they are required to undergo a significant amount of training before they are willing to use them autonomously (Gaskell and Cobb, 2004). In addition, standard corpus interfaces provide raw data, and it is the learner who is required to interpret it (Hunston, 2002), which may not always be an easy task. In other words, having language learners use the standard interfaces may not always be trouble-free, while for many teachers, engaging students in such investigations is a luxury that may not be easily achievable both time wise and training wise.

In response to these challenges, a growing number of corpus tools have emerged which can provide similar outcomes (i.e., provide language information based on reliable corpora), but which do not necessary include extremely long lists of concordance lines or require complex interpretation of raw data in order to identify language information about the target language. They are enormously simplified and user-friendly with the aim of serving language teachers and learners who have little or no training in corpus use. Another characteristic that distinguishes these tools from full-form and standard corpus interfaces is that they are more specific in searches. For instance, Linggle, an online web-corpus tool, is designed to help learners with word patterns. Finally, unlike standard interfaces which typically require the use of desktops or laptops, these new corpus tools are mobile-friendly which makes them more flexible and practical for language teachers and learners. Some of the most known tools are SKELL, Netspeak, Just-the-Word, and Linggle. In the coming section we discuss and evaluate two of them, namely, SKELL and Netspeak.

SKELL

Sketch Engine for language learning or SKELL (Baisa & Suchomel, 2014) is a free corpus-based tool that is specifically designed to assist language teachers and learners in exploring the English language as used by native

speakers. In fact, it is a part of Sketch Engine, a very well-known and full-size corpus interface. SKELL uses a very large corpus of English to provide teachers and learners with examples of words and phrases in context as well as relevant thesaurus entries and collocations for words. The corpus that the tool relies on is carefully selected text as well as quite balanced in terms of genres, and it includes text used in the web, such as articles, Wikipedia, blogs, and other reliable web content.

Similar to some other web corpus tools, SKELL is aimed to be a quick and user-friendly reference tool for language teachers and learners who don't have any training in corpus use, enabling them to have easy and simplified access to corpus data. The idea stems from the fact that traditional corpus interfaces require training in, say, the interpretation of raw data, and that information about the target language is not always readily available to the user.

Figure 1 SKELL (Baisa & Suchomel, 2014): corpus interface for language learners

SKELL provides three categories of searches, and these include: examples or words in context, collocations and word co-occurrence, and thesaurus or similar word searches. When it comes to exploring examples for words, for each search of a word or phrase, forty examples are provided to the user containing the expression or phrase searched (see Figure 2). This gives students and teachers plenty of real language examples that can be used for various purposes in teaching and learning. According to the interface's website (SKELL, n.d.), the software does not list any sentence found on the web, but rather it selects the most appropriate examples for learners of English. In addition, unlike other corpus interfaces which include broader and extended concordance lines, SKELL provides sentence-long concordance lines which makes the example sentences look less confusing and more practical for language learners.

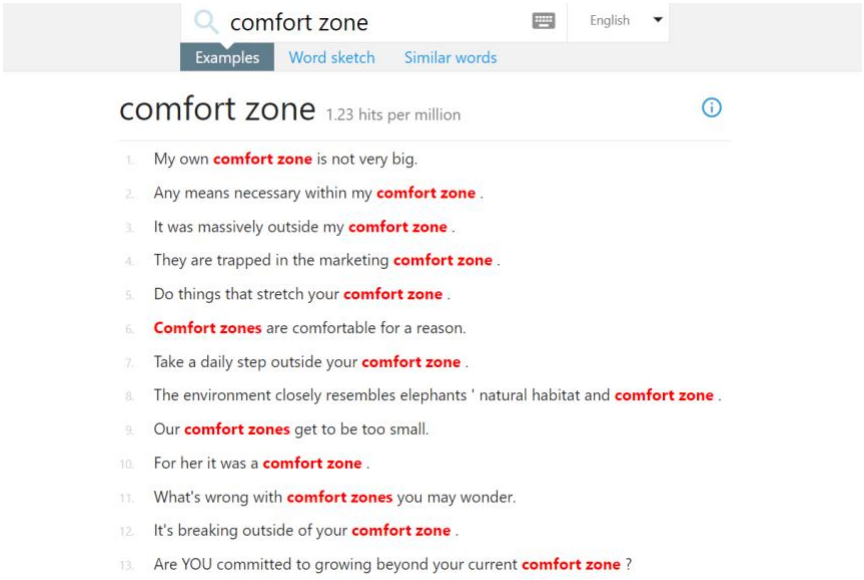

Figure 2 SKELL: examples or sentence-long concordance lines

Collocation search provides the user with a variety of words that co-occur with the word searched (see Figure 3). It is worth mentioning that this option provides more collocations for words than what is typically found in standard dictionaries or even collocation dictionaries. In addition, the collocates are grouped into grammatical categories, which include collocates that serve as a subject of the word searched (e.g., "fly" as seen in the figure below); object of the word; adjective collocates that come before the word, and so on. What makes this option more valuable is that for each collocate selected, forty concordance lines are provided, which the user can use to explore the co-occurrence in more depth and in more contexts. One more distinguished feature of SKELL in comparison to other corpus tools is that it can also provide collocations in phraseologies that include the conjunctions "and" and "or" in between. For instance, according to the corpus, "standards and guidelines"; "standards and requirements"; "standards and regulations" are some of the most frequent phraseologies of the term "standard" whenever the conjunction "and" is used.

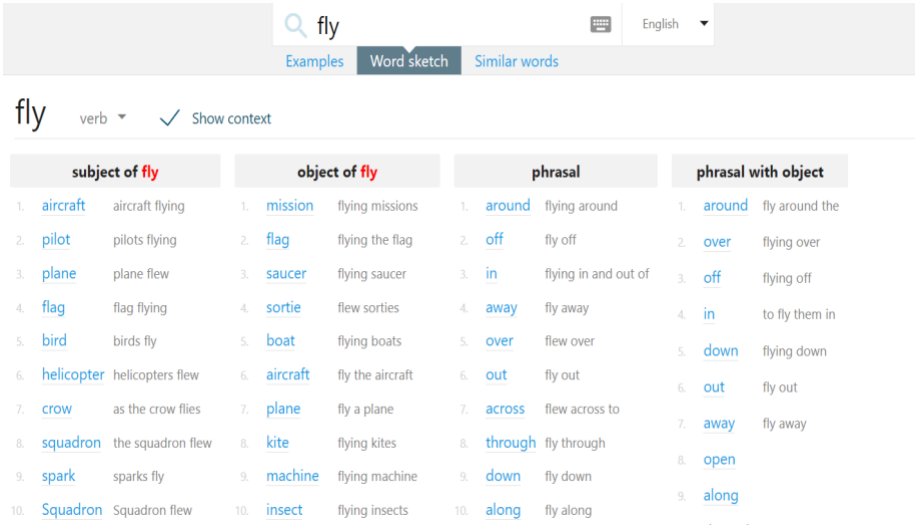

Figure 3 SKELL: word combination or word collocation search

Lastly, SKELL also provides a thesaurus for the words searched. For instance, for the same word, "standard", the interface generates the following similar words: *requirement*, *rule*, *policy*, *regulation*, *practice*, etc., which somehow appear in similar contexts in the corpus. Something to point out, however, is that it is not always clear how some words are similar to the word searched. For instance, it would be difficult for a learner to know how the word "issue", which appears in the results, is similar to "standard". The feature, nevertheless, still proves to be a valuable option for language analysis.

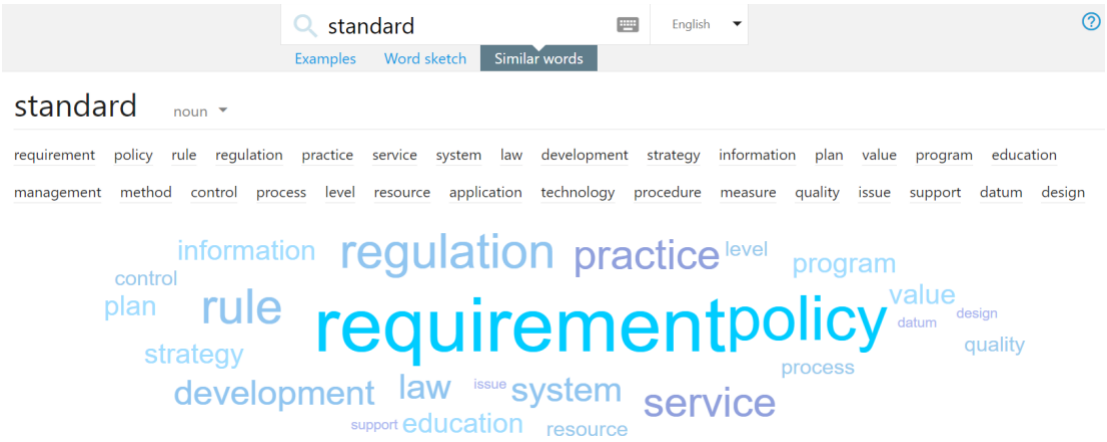

Figure 4 SKELL: similar word search

SKELL is a valuable tool for teaching and learning. It can be used as a simple reference tool, as we do with dictionaries, to search for quick language information, or it can be used as a tool for learning in classroom settings. Teachers can use SKELL to find more authentic examples for words in context, be that for making the meaning of words clearer or for exercises. Learners, on the other hand, can be given a task of using the tool to explore collocations for particular words. Thus, by having learners use the tool effectively they would have a sense of accomplishment considering that they are able to explore the target language autonomously, finding useful collocations for words and common phraseologies.

SKELL is to some extent flexible when it comes to who can use the tool. The interface can be used with teenagers and adult learners, while it may not be an appropriate tool for young learners. This is due to the complexity of the interface as well as due to the level of language used and different genres to which its language belongs. In addition, the tool is mostly appropriate for students of higher levels of proficiency. Corpora contain authentic language, and it does not provide sufficiently simplified language appropriate for low-level learners.

NETSPEAK

Netspeak is another web-based corpus tool available online that mostly relies on Google Books as corpus and it is aimed to assist foreign language teachers and learners with correct phraseologies, word patterning and collocation in the target language. It resembles a search engine (see Figure 4) both in terms of layout and some functions. However, the results don't exactly match normal search engine results. The page has a search box, requires no sign up, and it is free for its users.

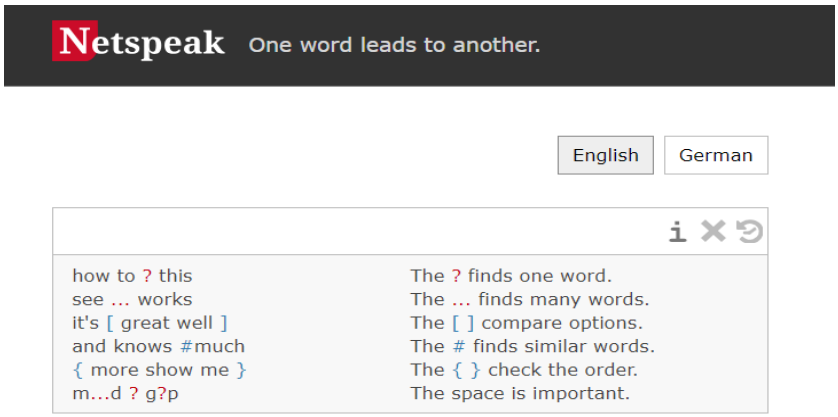

Figure 4 Netspeak: search box

Netspeak provides various and unique searches for its users that can hardly be done via other platforms. One of its very useful features is searching for a missing word in a phrase. This is done using some particular symbols in queries. For instance, if we search *"I am ? interested in"*, the results will show all possible words that commonly occur in between ***am*** and ***interested*** in this particular phrase. Note that the question mark asks the software to identify possible words that can appear in that position in the phrase. Netspeak gives a list of words that can occur in this position, and as figure 5 shows, the results are ranked based on the number of occurrences and frequency or percentage, meaning from the most frequent at the top to the less frequent down the list. Users can click any option provided which then opens plenty of concordance lines with the phrase selected. In addition to concordance lines, if the user wants to expand or see a concordance line in a broader context, Netspeak provides a link that leads to the electronic Google book online where the phrase is originally used.

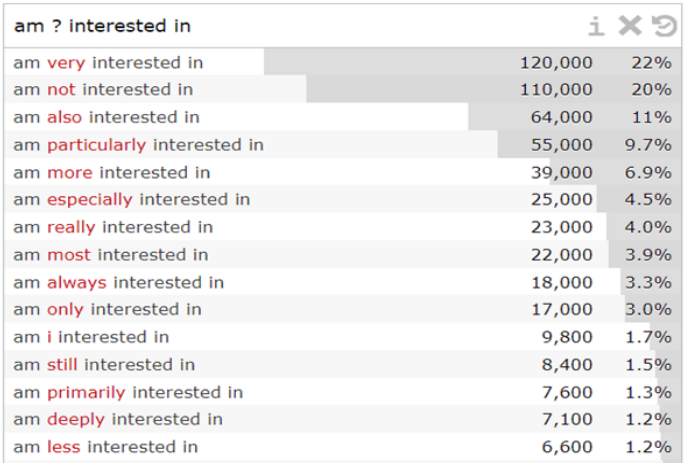

Figure 5 Netspeak: "missing word" search

Another exceptional feature is finding the best option. Netspeak can quickly and reliably tell which combination of words is more common in English when provided with two or more combinations. For example, if we search "[strong powerful] engine", we are asking Netspeak to identify which combination is more standard, ***strong engine*** or ***powerful engine***. According to the results, ***powerful engine*** is the correct and the most common co-occurrence. Note that in order to get accurate results for such query, the possible collocates of ***engine*** are put in square brackets with space around them.

Netspeak can also provide queries about English syntax, which is not something one can do with other available corpus interfaces. In order to ask for a correct order of some words in English, the words should be typed within curly brackets. For instance, if one searches for a correct order of these three words " { brick beautiful house }", Netspeak will provide the following order: beautiful brick house (Figure 6), which is the correct order in English.

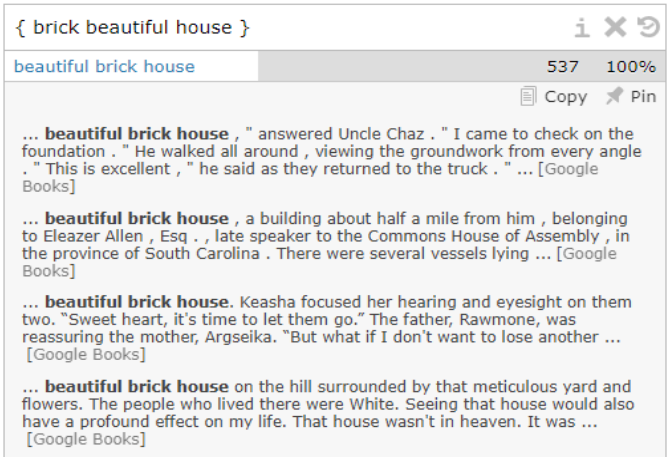

Figure 6 Netspeak: finding a correct order of words

There are also more specific queries available on Netspeak. One is finding a synonym for a word in a phrase. For example, if one searchers "and knows #much", the interface provides a list of synonyms for the word ***much*** (see Figure 7). Note that the symbol **#** put in the front of a word (i.e. much) asks the interface to find a range of synonyms for it. As the figure below shows, the most frequent phrases include *and knows a lot*; *and knows a great deal*; *and knows a good deal* etc. Referring back to finding a missing word feature discussed earlier, users can also use the question mark in all positions, be that in the front of a word or a phrase or after, for more word combinations or collocation; can use more questions marks to identify more than word in a position etc. In addition, users can use other symbols for additional special queries. For instance, the symbol * will generate none, one or more missing words at the same time in a phrase.

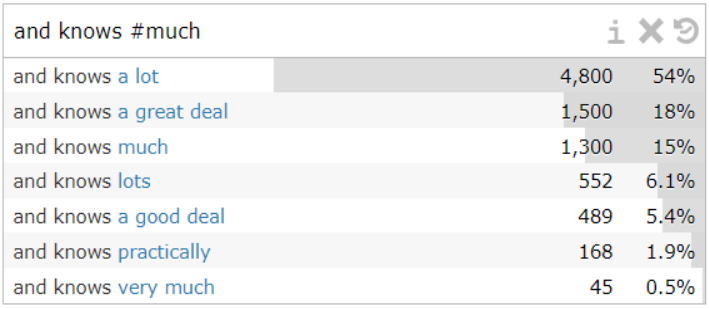

Figure 7 Netpeak: finding a synonym for a word in a phrase

There are some considerations one should have in mind when using the interface. The results are typically very clear and easily interpreted; however, there may be room for misinterpretation. For instance, as "look forward" appears to be more frequent than "looking forward" in the corpus, this does not mean that the later is wrong or less English; both are acceptable and common forms in English. In addition, if one doesn't find a phrase in

the corpus, this cannot be always interpreted as not being English. The phrase may simply not be part of the corpus Netspeak uses.

Netspeak can be a very useful tool for both teachers and learners of English. Teachers can use the interface for their teaching needs, such as find answers for more complex questions they may have about English or use the tool while preparing teaching materials and lesson planning. Learners, on the other hand, similar to using dictionaries, can use Netspeak as a reference tool whenever they have doubts about language use or can utilize it as a reference tool when writing or improving writing drafts. The interface may also be very useful in university settings. English language programs in countries where English is a foreign language can introduce and integrate Netspeak in more professional language courses for language analysis purposes, such as grammar and linguistic courses. Advanced learners of English can make use of the tool to do tasks that require them to engage in more inductive type of learning, such as Data-driven learning (DDL) or Inquiry-based learning.

CONCLUSION

The state-of-the-art corpus tools discussed in this paper can be very useful tools for both language teachers and learners. They can serve as a good alternative to full-size corpus interfaces, although this does not suggest that the later are not appropriate and ineffective for language teachers and learners. Standard corpus interfaces will continue to be appropriate alternatives for more in-depth and extended language analyses and DDL. The corpus tools presented in this paper can also be good complementary tools to Learner's Dictionaries. Taking into consideration that available dictionaries online still lack important information about words, it is these web-based corpus tools that can fill in the gaps. Teachers and learners can use these tools along with online dictionaries to make language learning an easier journey.

REFERENCES

Baisa, V., & Suchomel, V. (2014). SkELL: Web Interface for English Language Learning. In A. Horák & P. Rychlý (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing*, (pp. 63–70). Raslan.

Boulton, A. (2010). Learning outcomes from corpus consultation. In M. Moreno Jaén, F. Serrano Valverde & M. Calzada Pérez (Eds.), *Exploring new paths in language pedagogy: Lexis and corpus-based language teaching* (pp. 129–144). Equinox.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? System, *32*(3), 301–319.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Johns, T. (1991). "Should you be persuaded": Two samples of data-driven learning materials. In *Classroom Concordancing ELR Journal*, *4*, 1–16.

Johns, T. (1997). Contexts: the background, development and trialling of a concordance-based CALL program. In *Teaching and language corpora*, 100-115.

Luo Q., Liao Y. (2015). Using Corpora for Error Correction in EFL Learners' Writing. Journal of Language Teaching and Research, *6*(6), 1333-1342.

Mull, J. (2013). The learner as researcher: Student concordancing and error correction. Studies in Self-Access Learning Journal, *4*(1), 43-55.

SKELL. (n.d.). Examples, collocations and thesaurus for learners of English.
https://www.sketchengine.eu/skell/

Websites used:

https://skell.sketchengine.eu/#home?lang=en
https://netspeak.org/