**Frederick Jopling**
Locarno

# Electronic dictionaries, concordancers and lexical databases: tools for teachers and learners

*Negli scorsi anni un gran numero di dizionari nuovi è stato ideato. Le versioni su carta vengono completate da versioni elettroniche su CD. I nuovi dizionari sono praticamente tutti basati sui corpi linguistici. I corpi linguistici sono spesso disponibili anche su internet (per i link vedi l'articolo). Potrebbe essere interessante utilizzare delle concordanze per trovare informazioni in merito alla frequenza delle parole, l'uso e le collocazioni, ma anche per fare delle analisi letterarie. La maggior parte di queste concordanze sono per l'inglese, ma vengono sviluppate anche per altre lingue. (red.)*

Though most language teachers would probably agree that dictionary use comprises an essential part of language learning, it is likely that very few will have ever taken the time to introduce their learners to the vast array of dictionaries currently available. This is certainly due in no small part to time constraints, lack of teaching materials, to the specific demands of the curriculum, and many other factors. Yet it is probably also true that many teachers are themselves simply unaware that there are a number of interesting electronic alternatives to traditional paper-based dictionaries. Certainly in the ELT world, paper-based dictionaries are seen more and more as only one side of a product, the other side being the electronic dictionary that comes bundled on CD with the paper-based dictionary itself. These electronic dictionaries are, in a sense, the direct offspring of computer driven corpus linguistics, and most publishers of English learners dictionaries will tend to play up the fact that their products are 'corpus-based', and therefore reflect the 'natural', 'living' language. In fact, learners dictionaries have been drawing on corpora to determine dictionary content since 1978, so that nowadays all learners dictionaries – and indeed most dictionaries in general – are based on some corpus or other. It might, then, be worthwhile for ELT teachers to know just what a corpus is, in order to consider whether examining corpora might in any way be of use to their learners.

Simply put, a corpus is a body of texts in electronic form which have been collected for purposes of linguistic study.[1] Corpus content and size will vary according to the intentions of those who want to study it, but once created, a corpus can be accessed using a piece of software called a concordancer in order to find out things about the language contained in it (such as word frequency, usage patterns, collocations, etc.). These are the tools of modern-day linguists and lexicographers, tools which are becoming increasingly easy to use. The question then becomes: Might these tools also have a place in the language classroom? The answer is, of course, 'maybe', for everything depends on the learners involved. Ultimately, it is teachers who will have to determine whether concordancers are relevant for their learners, and this will naturally involve a degree of intuition and experimentation. Below are a few suggestions where this experimentation might begin.[2]

## 1. Concordancers – online and offline

There are several online concordancers which are freely available. For the most part these concordancers access corpora which are themselves relatively limited (a mere million running words or so), or limited versions of larger corpora, and as such they are best suited to queries about collocation and word frequency in a general sense. The more interesting of these include the Cobuild concordancer at http://www.collins.co.uk/Corpus/CorpusSearch.aspx; the page at– www.lextutor.ca, which allows the user to check concordances in a vari-

ety of corpora (the Brown corpus, or a corpus of spoken English, etc.) at http://www.lextutor.ca/concordancers/concord_e.html; and the British National Corpus (BNC) at http://sara.natcorp.ox.ac.uk/lookup.html. The results of a query are often displayed in so-called 'corpus strings', which allow the user to conveniently compare many examples of the word in use as it occurs in different contexts. This can be particularly useful for examining near synonyms. Comparing concordances of the words 'glisten' and 'glitter', for example, an observant learner might notice that many of the examples with 'glisten' contain some sort of liquid, whereas those sentences containing 'glitter' do not (at least not in *this* corpus). Unfortunately, such exercises tend to require a good deal of careful planning on the part of the teacher, and the free online corpora are often too small, too specific in content, or too outdated to be of much use. Nonetheless, these corpora are free and can be used to introduce such notions as word frequency and collocation, and, of course, as a way of demonstrating the idea of concordancing itself.

For deeper, more detailed inquiries into language one would have to subscribe to an online corpus or purchase a corpus on CD, such as the BNC. These *general-purpose* corpora on CD tend to be well balanced in terms of content, and are quite large (often more than 100 million running words). They can be accessed using free or professional software, just as the online corpora, but they draw upon a much broader spectrum of language and, often, contain more information about the source of the corpus strings (U.S. or British, man or woman, etc.). Having learners use concordancers to winkle out word meaning and usage patterns from actual examples of the word in context is called *data-driven learning*, and many teachers have reported success in engaging classes in linguistic research in this way. It is unsurprising, however, that many of these reports come from university lecturers or professors who are dealing with very advanced learners. Indeed, it is doubtful whether data-driven learning would be appropriate for any but the most advanced learners. For one thing, raw corpus data can be very messy, and it often takes near-native speaker competency to sort out the different nuances of meaning. As well, there is often a learning curve involved in using some of the more complicated concordancing software. Teachers must ask themselves: Is this the most valuable way to spend classroom time?

There are, of course, alternatives to using concordancers to examine corpora. Though the internet is not, strictly speaking, a corpus, it does certainly represent the most vast body of text ever collected. A search engine like Google, which scans the internet for key words, is *de facto* a kind of super concordancer, and can be used to do some of the things that concordancers do. It is especially useful to check up on the existence and use of new, faddish, regional or rare words which may not be contained in the latest dictionaries (such as 'concordancer', 'podcatcher', or 'vlog').

The Simple Concordance Program (a free download at http://www.textworld.com/scp/) is an interesting offline option. It is a piece of software which does not necessarily examine a corpus *per se*, but rather whatever text is fed into it. The texts can be examined in the same way the corpus is examined, to determine word (or even letter) frequency and collocation, but always within the confines of the text itself. This might provide a way of showing learners their own limited range when writing in English, perhaps with a view towards encouraging lexical variation. This can be done online, as well. The site at www.lextutor.ca provides an assortment of online text analysis tools which can do some very interesting things. One, for example, performs a frequency analysis of entire texts, breaking the text into four different colours, each representing the frequency group to which the word belongs (1000 most frequent words, etc.). This very visual way of demonstrating lexical complexity lends itself well to side by side comparisons of texts, or of examining the lexis of specific authors. Perhaps authentic texts could be compared with simplified easy-reader versions, to underscore the importance of reading at one's level (and, above all, of mastering the 2000 most frequent words in English as soon as possible![3]).

In any event, it is worth noting that even where concordancing tools might not be appropriate for learners, they might be appealing to the teachers themselves. They, too, will have questions about English which often cannot be answered by traditional dictionaries or native speaker colleagues, and concordancers might well provide the answer.

## 2. Electronic dictionaries

Electronic dictionaries[4] are beginning more and more to open up to dictionary users the corpora on which they are based. The latest Longman Dictionary of Contemporary English (LDOCE) includes an electronic dictionary, for example, which gives users access to a so-called 'example bank', providing a wealth of examples which do not appear in the electronic dictionary's main entry. These examples derive from a picked corpus which comes bundled with the electronic dictionary and can be displayed (and even printed) in 'corpus mode', i.e. as a search string. This has the enormous advantage of giving the learners many more examples of how the word is used than can be found in the paper-based dictionary. Knowledge gleaned from learners corpora (corpora comprised of error-ridden texts by non-native speakers) has also

provided the basis for a 'writing assistant' feature, which points out frequent errors made when using certain words or expressions.

The electronic learners dictionaries on CD from Cobuild-Collins, Macmillan, and Cambridge are also well worth examining. The Cambridge electronic dictionary is particularly interesting, as it seems to be the first electronic dictionary to attempt to develop a learners thesaurus.

Online dictionaries also abound, of course. Traditional, web-based dictionaries like those at www.m-w.com or the meta-search engine at www.dictionary.com might not necessarily be corpus based, but they are free and quite extensive. The open source dictionary at www.wiktionary.com is especially interesting, with a sort of bottom-up approach to lexicography that gives it very broad and up-to-date content. The Visual Thesaurus at www.visualthesaurus.com draws from the WordNet lexical database, and the site designers have developed a fun and innovative way of displaying word relations: Type in a word, and related words sprout out, as if from a seed. These, in turn, sprout new related words, so that the end effect is of an organic mass of sense-related words hovering around the original search term as if it were the nucleus of a cell. It is so visually appealing that it is bound to be popular with learners (which does not, alas, guarantee its usefulness in terms of language acquisition), but unfortunately it is geared to English native speakers, for whom it effectively combines a thesaurus and a dictionary (with a beta-version which cross-references to Spanish, French, German, Italian and Dutch!). The website does, however, provide free downloadable lesson plans for teachers complete with exercises for learners (http://www.visualthesaurus.com/educators.jsp) from which some exercises might easily be cannibalized for use with learners of English.

There are so many online dictionaries available that it is difficult to recommend which are best for learners. The entire content of the LDOCE is available for free at www.ldoceonline.com , and many Cambridge dictionaries for learners can be found at http://dictionary.cambridge.org/. Appropriate dictionary type will of course be determined by the task, and online tasks will naturally be best suited to online dictionaries. Once learners begin to deal with authentic texts, they will probably be more sensitive to the advantages of these newer, more flexible electronic dictionaries.

### 3. Concordancers as reading aids

There are a number of interesting online projects where concordancers have been integrated with specific texts (or collections of texts) in a way which is meant to assist reading. One example can be found at http://www.dundee.ac.uk/english/wics/wics.htm, which concordances the poems of the English romantic poets, essentially giving all the instances within the body of the poet's work where a given word has been used. At www.lextutor.ca there is an online text tool called the 'hypertext reader' which uses concordancing as a reading aid. Here users can upload any text (in '.txt' format), whereupon they are given a link to a web-site which is divided into three frames: one for the uploaded text, one which provides a concordance of clicked words from the text itself, and the third which provides a link to an online dictionary. The hope is that this combination of dictionary and concordancer use will lead to a richer processing of words, a deeper comprehension of the text and to more effective vocabulary acquisition.

When testing these electronic resources it should be borne in mind that linguists, lexicographers and software developers are not necessarily language teachers, and thus are not likely to be aware first hand of the usefulness (or uselessness) of their own linguistic tools for language learners. It is the teachers themselves who will have to start examining these tools to find applications (or combinations of applications) which might be useful for them and their learners. Ultimately, using concordancers and electronic dictionaries with learners will only make sense if it is introduced as part of a broader discussion about language and dictionary skills, so that learners' expectations might be tempered by their knowledge of what they can reasonably expect to find using such tools. In any event, one thing is certain: learners will never use language tools they do not know exist.

### Notes

[1] Meyer's overview of the subject is both readable and comprehensive (Meyer, C. 2002. *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.).

[2] See also: Conrad, S.M. 1999. 'The importance of corpus-based research for language teachers'. *System* 27 (1) 1-18.

[3] See: Hirsch, D. and P. Nation 1992 "What vocabulary size is needed to read unsimplified texts for pleasure?" *Reading in a foreign language* 8 (2) 689-96.

[4] Some background for teachers can be found at Nesi, H. 2000 "Electronic dictionaries in second language vocabulary comprehension and acquisition: the state of the art" in: Heid, U., Evert, S., Lehmann, E. and Rohrer, C. (eds.) *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000* Volume II. Stuttgart: Euralex, 839-847.

**Frederick Jopling**
teaches English at the Liceo cantonale di Locarno.