

EXTENDING VOCABULARY KNOWLEDGE WITH COMPUTERS

by **James Thomas**

Masaryk University

Brno, Czech Republic

thomas@informatics.muni.cz

When learners of a foreign language actively observe new vocabulary when reading, they might wonder if the new word, new phrase, interesting collocation etc, is worth learning. As experience tells us, "knowing a word" is much more than knowing its denotative meaning. In the effort to make a vocabulary item available not only for receptive purposes but for productive purposes as well, it is necessary to know the company it typically keeps. For example, what do you make of the third word in this headline from the Sydney Morning Herald?

Scientists all agog after discovery of fossilised menagerie

Sometimes a single context suggests the meaning, but this is not always the case. Dictionaries usually help, but not always: [Cobuild's E-dict](#) tells us: *If you are agog, you are excited about something, and eager to know more about it.* On the other hand, the fact that it is not in the [Cambridge Learner's Dictionary](#) (CD version), suggests that the lexicographers considered *agog* to be beyond learners' needs. Since it is no more practical for students to learn every new word they met, than it is for a dictionary to include every word in a language between its covers, it is necessary to prioritise. The following section discusses a variety of computer-aided approaches to prioritising.

One criterion we apply when deciding whether or not to learn a vocabulary item is the need it fulfils: does *agog* do anything that its synonyms don't? By putting the cursor on the word and hitting Shift F7, Microsoft Word 2000 offers the following synonyms for *agog*: eager (1351), excited (1803), impatient (677), keen (3710), avid (161), interested (8787), enthusiastic (1431), curious (2098). Notice that two of these synonyms appear in E-dict's definition above. And a list of synonyms certainly gives some idea of the field of meaning. An internet view on synonyms is the beautiful display at [Plumb Design Visual Thesaurus](#). Click this link and hit its "click to launch". Then at the bottom of the new screen, enter any word whose synonym you would like to follow. You can read a short review of this at [Yahoo Picks](#).

At the opposite end of the scale of "near-perfect combination of content and design" is another on-line source of synonyms called [Wordnet](#). After searching for a word, it provides you with that word's categories. For example, if you were writing an article describing an experiment and found that the word *experiment* was too frequent in your text, you might want to use a synonym. Wordnet offers three categories for this word, [\[click here\]](#) from which you can then choose to look at its synonyms [\[click here\]](#), and equally important for discourse, its hypernyms, hyponyms, holonyms, antonyms, meronyms, depending on what is available for your search word. The simplicity of design is not a liability.

A second criterion in deciding whether or not to learn a vocabulary item is an item's typicality. The numbers beside *agog*'s synonyms represent the absolute count of each word in the [British National Corpus](#) which has approximately one hundred million words: *agog* itself occurs only 25

times in the BNC which is a good clue as to why it was not included in the 35,000 meanings included on the CLD CD. If we accept that a non-native speaker's language production becomes more native-like as he or she employs more and more native-speaker like elements, the issue of typicality becomes important.

However, despite *agog* being a rare word relative to its synonyms, the Sydney Morning Herald chose it for that headline possibly because its rarity gives weight to the headline, possibly because "all agog" is an eye-catching cliché, and possibly because one would not expect scientists to be "agog". It is quite usual for a specific context to place constraints on the synonyms that can be substituted, so that of the eight listed above, the scientists could perhaps have been "excited" instead of "agog", but at the expense of a shade of meaning. Regardless, the newspaper was able to use this word because native speakers know it.

The scenario then is that students of a foreign language are meeting new vocabulary items while reading. Looking them up in a dictionary leads to a better understanding of the current text, but does not automatically lead to them being available for productive purposes. Meeting a word in a single context cannot exemplify the range of grammar patterns that a word has, nor the semantic fields it appears in, nor its various collocations. More focussed work is required if a student wants to learn a word. There are many ways to skin a cat, as the English idiom has it, and one way of elevating passive vocabulary to active vocabulary is by studying it in multiple contexts, and preferably authentic ones. One would have to read a lot of English text before coming across *agog* in such contexts. One could search a corpus, however. To be sure, a corpus is a database of multiple texts and a concordancer is a program which searches it.

Concordancing is a computer-based activity undertaken by people studying language for both academic and practical purposes. After searching a database of natural language, a corpus, for a particular item, usually a word, the program typically returns a page of text which has the target item vertically aligned down the centre of the screen. In this way, concordancing offers two simultaneous views on the target item: the horizontal view in which multiple contexts of the item can be observed, and the vertical in which the item's co-text is displayed. This data can then be interpreted and used to address the question that initiated the search. The diversity of a language can never be comprehensively presented in a dictionary and few dictionaries provide much information about a word's grammar or its collocations. For simple searches, it takes a concordancer seconds only to search a corpus, and the human interpreter of the data can often find the answer almost as quickly if the question is well-framed and the search query properly formulated. These skills are as learnable and as useful as the multitude of computer skills that people now take for granted. For language students, teachers, translators and people writing in a foreign language, access to data for checking one's intuitions on the fly is invaluable.

The corpus I use with my students is the Collins Cobuild Corpus Concordance Sampler (henceforth CCS), available on-line most of the time. They occasionally have server problems, e.g. a hacker brought it down for ten days in November 2002. Being a sampler means that the CCS does not offer a full range of functions - for this, one must become a subscriber. For example, the above BNC word counts for *agog*'s synonyms cannot be derived from the CCS. Nevertheless, as an introduction to how a corpus can help a learner address such issues as a vocabulary item's "company", its typicality and usefulness, it is ideal.

One way of learning to use something is to use it. So without further ado, click this link [Collins Cobuild Corpus Concordance Sampler](#) and Type in your query, agog. Then click Show Concs. Eight concordances appear - an unusually and gratifyingly small number to deal with.

What do you notice about the layout?

- The search word is centred.
- The concordances (the lines of text) are "cut off" which means that complete sentences almost never appear.
- The text on the screen cannot be read as a single text because the concordances come from different sources.

What do you notice about *agog*?

- Two concordances are preceded by *all*, as is the newspaper headline above.
- *Agog In The Ether* looks like the name of something. To resolve this, copy and paste the phrase into an internet search engine: Google found the phrase 52 times (when enclosed in inverted commas - "Agog In The Ether"), almost always capitalised as here, and clearly the name of a piece of music, so this phrase is probably not important in "knowing the word".
- Four concordances(50%) are preceded by some form of BE.
- And you can often see the reason for someone being *agog*: the news, the performance, over the arrival, to see the world, to hear the news. Thus, it would seem to be a reaction although there is no indication of it being a positive or negative reaction.

What has been learnt about the word in the process of making these observations?

- The company the word keeps is easily spotted: grammatical (be), lexical (all), semantic (the reaction).
- It is a very infrequent word if it only occurs eight times in "56 million words of contemporary written and spoken text", as is written at the top of CCS's main page.
- You are now in a position to consider if the word is worth learning, and if so, what there is to learn about it.

The CCS allows a great many types of searches which can answer many simple questions that crop up when reading and writing: this includes translating and marking students' written work. My website, [A Ten-step Introduction to Concordancing through the Collins Cobuild Corpus Concordance Sampler](#), aims to instruct people in forming search queries and utilising the results. As its title indicates, it is an introduction to concordancing, not specifically a manual for the CCS. Relevant linguistic concepts are illustrated and a wide range of search activities is provided. There is also a plethora of links from this site to related topics.

DIY concordancing

Once students are familiar with reading concordances and understand what a unique and valuable perspective can be gained by observing vocabulary items in multiple and authentic

contexts, there is another resource available on the Web which can be utilised, if the reading text is available in electronic form.

[The Compleat Lexical Tutor](#) offers a wide range of vocabulary learning facilities. We will look here at only one of them. Click the link and go to the Hypertext Builder in the "Teachers" column. Once you have a text pasted into the text box, the program creates a hyperlinked version of it. This means that while you are reading and studying it, you can click on any word to see it concordanced in a lower part of the screen. You can also click to see the word in a WordNet window with the facilities described above. It literally takes seconds to copy and paste a story from the Internet, e.g. from a news service such as the Sydney Morning Herald or [The Guardian](#), into the Hypertext Builder and have the page available for reading and studying in these ways.

In conclusion, it has to be admitted that it takes some investment in time to learn how to use concordancers and derive the benefits that are available. Part of this investment is coming to understand what vocabulary per se is, and my [Introduction to Concordancing](#) website is written with that in mind. In an era which so highly values learner autonomy and discovery learning, concordancing has much to offer. In an era when more and more teachers and students have easy access to on-line resources, dictionaries, thesauri and concordances open up paths to linguistic independence undreamt of even a decade ago. But these are little more than toys in the hands of novices: language learners need learner training to understand what they need to know.