



From dictionary user to amateur lexicographer

Dr Costas Gabrielatos

25 March 2009

The topic for my lecture today will, on the one hand, look at how you move from a dictionary user to something of an amateur lexicographer, but also it will trace how dictionary-making moved from citation slips, meticulously collected by lexicographers, to language databases, some of them available online and, to an extent, for free, and which are now used to inform modern lexicography.

Perhaps I need to clarify, here at the beginning, that this is not going to be a lecture in the traditional sense. There might be some elements of lecturing, but what I aim to do is give you something of a guided tour of what resources are available on the internet, and I will concentrate only on those resources that are free, for obvious reasons. There are some websites that provide some information for free but require you to subscribe to get full access, but one of my aims is to show that, by combining the various different sources, that might not be absolutely necessary. But that is not my only aim. My other aim is to raise some issues about using online resources in particular, and using dictionaries or any reference source in general, to talk about their utility, but also their limitations. So in a sense, my aim could be seen as an irreverent approach to reference books, and I will be seen to encourage a less-than-trusting approach to dictionaries and reference books.

My approach comes from my experience as a linguist and, before that, as a language teacher, but also, and I think mainly, because of my being what you might call a dictionary enthusiast, and some people even say dictionary addict. I own maybe a bookcase full of dictionaries, different types of dictionaries - so I like dictionaries!

But I think, very fittingly for the occasion, my approach can be encapsulated in this statement by Samuel Johnson: 'Knowledge is of two kinds: we know a subject ourselves; or we know where we can find information upon it.' I think it is quite an empowering statement because, even though we do not know everything ourselves, it entails that we are not just passive recipients of somebody else's knowledge but that we can use different resources and actively seek knowledge ourselves.

So typically, when we are confronted with a word we do not know, or a word we are not sure about its meaning, the way we can use it, or the contexts we can use it in - what we do is we reach for the dictionary, and then we look the word up. In doing this, we make two assumptions, although perhaps not consciously. Assumption number one is what is in this book is the truth - we can absolutely and entirely trust the information; and the second assumption is whatever is in this dictionary is everything that there is to know about this word. I will try to, in a sense, shake things up a bit, and ask you not to make these two assumptions, particularly when you are using online sources.

Of course, we could use other print dictionaries, and I will be making this distinction between print dictionary or reference book and online dictionary or reference source, but the problem is that if we want to do the amount of cross referencing that we might need, we will need to invest quite a lot of money in dictionaries, and I would not ask people to do what I have done and spend quite a lot of financial resources on that! However, if we are willing to invest time, and we have access to the internet, then we can get as much information as is possible on a specific word or expression.

To give you a sneak preview of how my lecture/presentation/guided tour will end, we may choose not to only limit ourselves to juxtaposing the information in reference sources, but we might want to examine the use of a word or expression in the wild, so to speak. Again, that would be difficult before the internet. We would need to go out and meet as many people as possible and interact with as many people as possible

or read as many books as possible. Thankfully, we can now have access to resources where this experience is given in a condensed form, so to speak.

But we can begin with the question: 'How will the information be used?' One of the distinctions I would like to make here is what we are interested in the meaning of the word for. For instance, if we only want to understand a word in a passage, then we will need much less information than if we want to actually use this word in speech or in writing. This is because our choice of words and our choice of the surrounding language in which we will use this word will be subject to scrutiny by other people, and in many respects, who we are is judged by the language we choose to use.

So, as a springboard for my discussion here and the guided tour, I would like to use the deceptively simple question: 'Do you know this word?' Usually if people know the meaning of a word, or one of the meanings of the word, they might be tempted to be say, 'Oh yes, I know this word.' But what I would like to do is structure my presentation around the different aspects involved in knowing a word, and then, through that, show you how we can get information about this aspect using online sources. By no means am I implying that the example words I will be using are words that you may find difficult; it is only an arbitrary choice of words which I think can better demonstrate my point.

So let us start with the form. In many respects, we can draw parallels between knowing a word and knowing a person, so one of the things we need to know is what this person looks like. So for my first example word, which is 'chagrin', what we need to know is that, if we are going to write it down, we use the sequence of the letters C H A G R I N. But then, we may want to use this word in speech, so we need to know how to pronounce it? This is never necessarily obvious, for instance, we can ask ourselves whether this 'ch' pronounced 'sh' or is it pronounced 'ch'? Notoriously, the English language has got a less than ideal correspondence between script and sound, and I think it baffles not only learners of the English language, like myself, but native speakers as well. The other question is where do we stress? We have two vowels here, so which one takes the primary stress? Is it the 'a' or is it the 'i'? Print dictionaries will give you the pronunciation using symbols. Unfortunately, not everybody uses the International Phonetics Association (IPA) symbols; different dictionaries will have their own type. So, as a dictionary user, particularly as a user of a multiple dictionaries, you need to familiarise yourself with one, two, three or four different phonetic spelling conventions, which might be irritating. But, on the internet, all you need to do is click and you will actually have this word very helpfully pronounced for you. So this is one convenience that online sources offer.

Another aspect of knowing the word is what you might call the grammar of the word. I know that the popular wisdom is that English does not have much grammar, but I know that there are lots of people doing, and have done, PhDs on tiny aspects of English grammar, so I think there is quite a lot of grammar.

One thing you need to know is the part of speech: is this word a noun, a verb, an adjective, or has this form got multiple grammatical functions? Is it countable or uncountable? Could you say that you have had one, two, three experiences, for example? Traditionally, particularly pedagogical grammars would say that experience is uncountable - you cannot count experience. The Longman Dictionary of Contemporary English online will tell you something even more helpful and important: experience is countable, or experience is uncountable - it depends on which meaning of experience you want to concentrate on, which meaning of experience you are using. So by quickly looking online I can easily find out within minutes that if experience refers to knowledge or skill, then it is uncountable; knowledge of life is uncountable; but if it is something that happens - 'This was my first experience of living with other people' - then we can count experience. So by using these sources, and sources that are so easily accessible all I needed to do was just click on the word, and I got what I was looking for.

Another aspect is whether we can add syllables to a word: something in the beginning, a prefix, or an ending, a suffix, and then turn it into another part of speech. This is also helpful because it allows us to enrich the way we speak and the way we write.

But let us move onto more grammar. The grammatical elements are ones that we can find in good dictionaries, but on the other hand, some dictionaries might not give this information. If we have online access, we can look up the word in different dictionaries and collate the information. But it is transitivity for verbs: so 'She gave a concert.' We have the, let us say, the action of the verb transferred onto an object or a situation or a person. Whereas, 'Prices have been rising,' there is no such transfer of the action.

I will be talking about synonyms and the pitfalls in the use of synonyms later, but here is a sneak preview. 'Help' and 'facilitate' overlap in meaning and so have a core semantic association, but we can talk about

policies that help or facilitate development, but we cannot facilitate a neighbour to move the sofa, we cannot facilitate a student with an assignment. So two words may mean more or less the same, but they may not do the same things in writing or speech. They might not have the same grammatical patterns or the same lexical patterns.

But the main reason why people use dictionaries is to look up the meaning of a word. But, again, meaning is not a single element; it is not as simple as we might think. We can distinguish between the central sense and that which is not central. For example, for something to be 'par for the course', it means that something is what you would normally expect to happen; but then you have the connotations of the word, let us say, the affective meaning, the emotional meaning of a word, and this comes next used to show disapproval.

This is a good point for me to also say, usually, native speakers of a language want to use dictionaries aimed at native speakers of a language, and I have seen people frown upon dictionaries for students. However, because learners of English are not expected to know connotations, this type of information is given much more often than in dictionaries for native speakers. So this piece of information is more probably found in dictionaries for learners than dictionaries for native speakers. So they are a good resource for information that might not be seen by the lexicographers as being salient.

Now, when the word we want to look up is an object, of course we may find a brief description in a dictionary, but this brief description might not be helpful or it might not be sufficient. We could go to an encyclopaedic dictionary and read a much lengthier description of that object, but sometimes, it is much simpler to go to an online image search, write in the word, and see, not one or two, but hundreds of images of that thing. So by looking not at one photograph or reading not just one description, we can gain a better understanding of what an object is than any a one-page description might give us, and we might also realise that the object might come in different shapes and sizes and still be described by the same word.

There are some pictorial dictionaries and there are some encyclopaedic dictionaries which provide a mini-entry and a photograph, and of course these are much more helpful than a brief description. But what the main difference that online resources make is that we do not need to be happy with just one image; we can have as many as we need, and compare and compile a composite picture of that object in our minds.

Sometimes, it is not a word that we are concerned about; it is an expression; it is a combination of words. We might know what collateral is, and we might know what debt is, and what an obligation is, but in view of the recent/current credit crunch, we read in newspapers about 'collateralised debt obligations', and we might want to know what these are.

If we were to type this phrase into a search engine, we would soon arrive at InvestorWords, which is a website that specialises in financial terms. Here, not only would we find two definitions - one quite lengthy - but each definition has got what we call clickable words. These include such words as: security, debt, bond, loan, mortgage, deal, CBO, CMO, risk, investor, exposure, purchase etc. So this is not the end of your looking up; this is the beginning of your looking up. In fact, by clicking on these words, you may end up not only knowing what Collateralised Debt Obligation is, but actually deriving some knowledge on financial instruments along the way. Of course, you can stop here if that is all the information you need, but at least further relevant information is right there at only one click away.

But many people will not know to look in InvestorWords for this particular phrase, nor in any other specific type of dictionary which is aimed at particular areas. However, the way to find these different specialist dictionaries is to type your search into the OneLook dictionary search, which is something like a junction. It will then give you links to different types of dictionaries, which will give you different amounts and types of information. Some of these dictionaries will be good for some words, but poor for others, and vice versa, so the important point here is that you should not be happy with the first webpage you land on.

Another thing we might need when we want to use a word is what has been described as the lexical company a word keeps. The technical term, if you come across it is, collocation, being in the same place or very near.

Why is that important? Well, we have here three instances of something having gone bad: stale bread, rancid butter, sour milk. All these are particular instances of something going bad, but the individual words for the different types of food going bad do not transfer between them; bread does not go sour or rancid and nor do milk or butter go stale. This shows that words are particular about the company they keep. Because two words might have the same meaning, it does not mean we can use the same adjectives. Two nouns might be synonymous to a large extent, but they may not share the same lexical patterns, the

same collocations.

This is not a rare case, but it is not a very common either: one word might not even have a life on its own. Par is not a word that forms relationships freely. Par seems to be committed to relationships with other words, and is found only in specific expressions: so 'on par' with something, 'below' or 'under par', 'par for the course', as we said before, and this expression 'par value' and 'par excellence'. So knowing the patterns of a word is very helpful when we use this word in speech or writing. Not every dictionary will provide this information, so the more information we get on a word, the better equipped we are to use it accurately and appropriately.

Then we have the situations that a word can be used in. To use a word appropriately, we need to know whether it is general or specialised. So, 'heart attack' is the everyday term, but 'acute myocardial infarction' is the medical term. So knowing this distinction, and knowing that these two words are synonyms - they describe exactly the same condition - but they are not at ease in the same contexts of situation. A word might be archaic, formal, informal, colloquial, or slang. For examples of these consider the following words that mean the same thing, but which would be used in different contexts: offspring, minor, child, kid, brat. These words make our listeners and readers feel different when they read them. They may mean the same in the general sense, but they do not do the same things.

Another question we can ask is how frequently a word is used? If we use it, will we blend in within that speech community, or will we stand out as the person who used this rare word in this context? And if a word is frequent, in which contexts is it frequent? Where is it safe to use a word, and when is it safe to use a word? Unless we actually want to stand out, which is fine, we would not otherwise want to find ourselves standing out in a situation where we would rather we did not.

Another aspect is the meaning of the word in relation to the meaning of other words. A very good example to use here is the word, 'expand'. Its meaning could be: amplify, swell, distend, inflate, and dilate. All these meanings share the common core sense of making something bigger, but this making bigger can take different forms. Is it bigger in this way or is it bigger in some other way?

To finish with knowing a word, in terms of the information we get about this word, it is good to know the history of that word. To come back to the analogy I used earlier, the reason that it is good to know the history of a word is because it is akin to our knowing a person better through knowing their history or their past. For example, for the word, 'sequacious', we can find a very concise piece of history: mid-17th Century, comes from the Latin, which meant 'inclined to follow'. For some people or for some purposes this might be enough, but for other purposes it may not. If, for your purposes, you need more information than this, you will need to go to a more specialist dictionary. For instance, WorldWideWords is a website which offers lots more wonderful information about the history and the historical forms of use of the word. If we look in it for 'sequacious', we will find an encyclopaedic entry on the history and uses of this word, current uses, older uses, with examples, and information about its frequency now, whether it is a word that is used, whether it has got a strong currency, or whether it is a word that is used rarely, for effect. I thoroughly encourage you to visit the site as it is clearly maintained by someone who is infatuated by the English language and they have spent an inordinate amount of time working on it - it is a wonderful resource!

Lexicographers have used examples to reach decisions about the definition they would actually write for the dictionary. Of course, and this is part of my irreverent approach, there have been cases where some entries were written according to what the lexicographer knew about that word. Why? Because he or she was a native speaker of that language, and of course they spoke the language, of course they knew what this word mean, and the only problem is to write down a clear and helpful definition of that word. But contemporary linguistics and in particular my strand of linguistics, corpus linguistics, has found that one conclusion that can be drawn is that intuitions are wonderful and really helpful, but they are not always correct. What is more, if you ask ten people to give you their intuitive response about the meaning and use of a word, you may get ten different and not entirely overlapping views.

So we have then the examples, on one hand, and the intuitions, what people may know about the word through their experience with it; but then, we need to ask ourselves, 'What kind of examples and whose examples are they?' I think that we should here listen to the authority on word meaning, Humpty Dumpty: 'The question is, said Humpty Dumpty, 'which is to be master - that's all.'" So who is going to be master? Who knows what words mean? Who knows how words behave? Who knows what friendships they have, grammatical friendships and lexical friendships?

Well, one answer is, of course, the speakers of the language know - they know quite a lot. Unfortunately, they do not know the same things, and also, unfortunately, they do not know everything there is to know about a word. We saw that there are quite a lot of aspects of knowing a word. So what do we do if we are lexicographers, and what do we do if we are amateur lexicographers, and this is where we go into the realm of our all becoming amateur lexicographers.

There are two sources of examples when we use the internet. One is corpora, which comes from the Latin for 'body'. These are bodies of language, language collections, electronic language databases, but they are not corpora in the traditional sense, which was the whole body of work of a novelist or a playwright, like 'the corpus of Shakespeare plays'. The reason for this is that you cannot put all of language in a database - it is simply impossible. So what corpora do is they try to recreate a microcosm of language, with samples from different sources of language use and different situations of language use.

Let us take, for example, the British National Corpus, which was compiled partly in my own university, Lancaster University. It consists of 100 million words. It sounds impressive but in fact it is not that impressive - it is only about 2,000 books. But in fact the BNC contains much more than 20,000 books can contain because it does not contain entire books; it contains samples of books, so a few pages or maybe a chapter from each of the books in the sample. It contains formal and informal language, spoken and written language. It has demographic information about the age of speakers, their socio-economic status, whether they were men or women. It contains newspaper text, it contains letters, it contains just about any type of text that people produce. So it is, let us say, a snapshot of English language, as actually used by real people - and I say 'real people' because there is some sort of abstract unreality to examples thought of by the lexicographer.

Let me say here that examples that are produced by the lexicographer for the dictionary can be very useful because they encapsulate some central aspects, but they cannot encapsulate everything there is to know about a word. To get that picture, we need to read much more than just one example. So, when we are in doubt, we might want to look for that information ourselves.

As an example, let us have a look at the word 'egregious'. These are five definitions from five different sources:

- Conspicuously bad or offensive.
- Often of mistakes, extremely and noticeably bad.
- An egregious mistake, failure, problem etc is extremely bad and noticeable.
- Extraordinary in some bad way; glaring; flagrant: an egregious mistake; an egregious liar.
- Conspicuous ; especially : conspicuously bad : flagrant <egregious errors> <egregious padding of the evidence - Christopher Hitchens>

If you do not know what it means and do not know how it is used, you will look it up, and if you are not satisfied with just one source, you may very well come across these five definitions. These definitions tell us the connotations of this word are negative - they all use the word 'bad' - so if something is egregious, this is not a good thing. We can be fairly confident, because it is in all the definitions and so this badness is conspicuous. We can continue to build up what we know about this word by looking at these definitions by noticing that it tends to be used with mistakes - we have 'mistakes' or 'errors' in nearly all of the definitions - but we also see that it can refer to a person, an action, or the result of an action. So far, we have been doing quite intense looking up, but we might not have completely accurate information. If we stop here, we might legitimately ask what the context to use 'egregious' in is? It seems that these definitions as we have them tell us nothing more than that it is when we talk about mistakes, blunders or errors, but that is perhaps not completely helpful.

If we look up 'egregious' in the British National Corpus, there are only 36 instances. This already tells us that it is not a word in general use and so occurs quite infrequently, which is a piece of information that most dictionaries will not give you. But, by looking into the types of use in which the word appears in the BNC, we can learn even more about the word. We can split up the instances of 'egregious' in the BNC in the following ways:

- Action/behaviour/event/result etc. (15)
- Person/organisation etc. (12)
- Mistake/error etc. (6)
- Object (2)
- Other (1)

From this, we can see that the most frequent use was when attached to action, behaviour, event and result, closely followed by a person or an organisation, and mistake/error was the third most frequent use, with an object having two instances, and one use that I could not categorise. Therefore, if we have a look at these instances of the word in the BNC, even by just reading through 36 examples, which does not take that long, you can perhaps gain information that reading through five or six or ten dictionary definitions might not.

I am not saying that you should throw out your dictionaries and take out a subscription for full access to the BNC. In fact, these examples come from a free interface, which gives you a maximum of fifty examples. So if a word has got 1,000 examples, it will tell you you could have had access to those 1,000 examples, but that we can only give you fifty of them. With 'egregious', we were lucky - we got all there was in the BNC. So I am saying that this is something that you could do to supplement or enrich your online looking up.

Before I finish, I would just like to show you the different types of print dictionaries that exist, or at least these are the ones I have:

- General - alphabetical
- General - thematic (concepts)
- Thesaurus
- Terminological
- Pronunciation
- Collocations (lexical patterns)
- Pictorial
- Encyclopaedic

You have the general alphabetical dictionary, which is the typical dictionary that we have in mind when we generally talk about dictionaries. There are also general thematic dictionaries, where things are organised by concept, the Longman Lexicon is a very good example and a very good dictionary of this sort. A thesaurus, when you will get synonyms and antonyms of a word, but usually with no other information as to the overlap in meaning and use. Terminological dictionaries are things such as dictionaries of finance, economics, science, technology, computers, etc. In pronunciation dictionaries, you get the different pronunciations a word might have - good ones include regional dialectal pronunciations as well. Dictionaries of collocations or dictionaries of lexical patterns are not that common - there is one general one, and one for learners of English - but increasingly, general dictionaries for learners include collocations. So if you cannot find information in a dictionary for native speakers, it is not a bad idea to look it up in a dictionary for learners. Pictorial dictionaries generally, unfortunately, only provide one picture, or sometimes a black and white drawing, which is not the same with the hundreds of images that you can get through an online search engine. And then there are encyclopaedic dictionaries, where you get a more wordy entry, and sometimes photographs.

So, having come this far, it is time to ask what the possibilities of online searches are? Is it the ultimate reference source? It would be tempting to say that it is. But we can also see the route down which online searches have begun and which further technology could perhaps take us further. For instance there might one day come a time when we have virtual reality dictionaries, where we could smell smells, and touch objects. If you could smell an odour or feel the texture of a fabric, you would not need a definition in the traditional sense. So perhaps it is best to say that online searches for the meanings of words are the ultimate so far, but as to what will come in the future, we will see.

Online searches are free, which is very good. You can compare different sources quickly and easily, so instead of depending on one single authority, you can pit authorities against one another and then see whether there is any core to their definitions, such as we found for 'egregious'.

Another great advantage of online searches are what we might call 'lexical journeys'. These begin when you write a word or expression into a search engine and click enter. Every click you take from then on will send you to another site, with more information and more links to other sites. If you have got a job, perhaps you should not carry out these lexical journeys during the weekday, because you just do not know where you are going to end. I have read somewhere the expression 'Zen navigation', which I take to be the case where you follow someone or something, it might not take you where you want to go, but it certainly will take you somewhere interesting! So that online searches are a great way to have your Zen navigation lexical journeys.

Finally, I should note that you yourself can do primary research along the lines of what current modern lexicographers do with corpora and what, in older days, lexicographers did by collecting meticulously citations and then putting them in shoeboxes or envelopes, and then having to sort them out, spending hours.

I think that these few main features that I have mentioned are what recommend online searches and what can turn the avid dictionary user into an amateur lexicographer.

© Dr Costas Gabrielatos, 2009

A sample of free online lexical sources

Dictionaries

- The Longman Dictionary of Contemporary English Online: <http://www.ldoceonline.com>
- Cambridge Dictionaries online: <http://dictionary.cambridge.org/>
- Dictionary.com (Dictionary + Thesaurus), based on the Random House Dictionary: www.dictionary.com
- Compact Oxford English Dictionary: <http://www.askoxford.com/dictionaries>
- Merriam-Webster Dictionary (+ Thesaurus): <http://www.merriam-webster.com>
- The Free Dictionary (+ Thesaurus + Pronunciation-audio + Encyclopaedia), based on The American Heritage Dictionary of the English Language (4th edn.), 2000: <http://www.thefreedictionary.com>
- PONS (+ Thesaurus, collocations): <http://pons.eu/dict/search>
- One Look (links to 30+ online general and specialised dictionaries): <http://www.onelook.com>
- Wiktionary: <http://en.wiktionary.org>
- Word Net (+ Thesaurus, collocations): <http://wordnet.wordmind.com>

Encyclopaedic entries (and much more)

World Wide Words: <http://www.worldwidewords.org>

Corpora

British National Corpus (max. 50 hits): <http://sara.natcorp.ox.ac.uk/lookup.html>

Free-access online corpora (+ full access to the BNC), maintained by Mark Davies / Brigham Young University: <http://corpus.byu.edu/>

Collins Wordbanks Online: <http://www.collins.co.uk/corpus/CorpusSearch.aspx>

Michigan Corpus of Academic Spoken English: <http://quod.lib.umich.edu/cgi/c/corpus/corpus?c=micase;page=simple>

WebCorp: <http://www.webcorp.org.uk>