Advanced ESP and the Learner's Dictionary: Tools for the Non-language Specialist

Geoffrey Clive Williams
U.F.R Lettres et Sciences Humaines
Université de Bretagne Sud
4 rue Jean Zay
B.P. 92116
56321 LORIENT CEDEX
France
Geoffrey.Williams@univ-ubs.fr

Abstract
Advanced learner's dictionaries are a major asset for both the comprehension and production of texts. However, despite their usefulness they remain a tool that primarily addresses language specialists, whereas the ESP student tends to rely heavily on the bilingual dictionary. This raises the question as to whether general language advanced learner's dictionaries are adapted to ESP needs. The audience considered here are established scientists and doctoral students publishing in English. The study compares the entries for scientific words given in the major Advanced Learner's dictionaries with both general and special language corpora and discusses the possibility of adapting such entries to the needs of the Advanced ESP user.

1 Introduction

Learner's dictionaries are made for learners, but who are the learners in question? In the case of Advanced Learners Dictionaries, it seems that the key audience are students studying languages at school or university. The ESP student tends not to buy a monolingual dictionary but relies on the bilingual dictionary. Given that it is impossible to make across the board generalisations about users, this study concerns usage in French universities, and in particular the universities in which the author has worked.

In French universities, most students studying English as a language will buy one of the major advanced learners dictionaries, whilst those studying applied languages tend to buy only bilingual dictionaries, in both cases it is far from certain that dictionary skills are actually taught. Whereas virtually all students in French universities follow some form of English course as a compulsory module, they are generally under no obligation to buy any dictionary at all. The leitmotif seems to be that non-language specialists, which includes those in applied languages, need to build up a vocabulary from studying texts in class, for which the teacher will be the fount of knowledge as to word meanings. Bilingual dictionaries may also be used, but only for the building of glossaries based on simple equivalence. The problem is that the non-language specialist, especially in the sciences, will rapidly become a full-time

language user as the world of research requires the reading and production of highly specialised texts in English. The lack of dictionary skills and a reliance on simple equivalence often leads to some very odd formulations and a blind faith in automatic translation. One obvious solution would seem to be the introduction of greater language awareness through the use of advanced learners dictionaries, but this raises the question as to whether the content is adapted to the texts they are reading.

The needs of non-language specialist users are not ignored by the learners dictionaries, in the Oxford Advanced Learner's Dictionary (OALD), for example, we not only have a list of the 3000 keywords of general English, but also specialist lists of words from the world of art, science, business and finance. In this study, the Oxford Scientific 250 was taken as the point of departure. The aim of the study will be to compare usage of these words in three major learner's dictionaries, OALD and Collins COBUILD Advanced Learner's Dictionary (CCALD) and the MacMillan English Dictionary for Advanced Learners (MEDAL). The presence or absence of these words will then be investigated in a corpus of biological research articles to see to what extent the dictionary entries cover the senses found in the corpus. It is proposed that it should be possible to build a dictionary for advanced non-language specialist users by using more appropriate examples and, in some cases, expanding the definitional prototype to take in variations of use in specialised settings. In this paper, we shall restrict ourselves to a study of usage in the OALD with a comparison for a small number of words against the other two dictionaries and a special language corpus.

2 Oxford Scientific

The Oxford Scientific 250 consists of 265 word-forms, given that it includes spelling variations, for example *fibre* and *fiber*, and related forms, as *genetic*, *genetically* and *genetics*. According to the OALD, these are the "most common words (apart from the Oxford 3000) in writing that deals with science subjects (physics, chemistry, biology etc.). Knowing these will help you understand texts about these subjects more easily" (R115).

Insofar as these are going to help the learner, it is natural that their specialised nature be signalled in the entries. Consequently, we have gone through each of the entries in OALD to see whether subject labels are provided and whether the scientific nature of the word is explicit or implicit in the definition or example. This has been done for all 265 entries, plus one, as there is possibly an error in that it seems more likely that 'rate' and not 'rat' is the scientific word to be known.

The scientific aspect may be shown explicitly or implicitly in the definition. The latter usually occurs if the word has a very general scientific usage and the scientific nature can be surmised from the text. For labelled words there is generally no example, the user is deemed to need to understand it, not reproduce it. In many cases, the examples come only from business contexts, which seems to show that the expected audience will be stronger in the area of applied languages than in the sciences.

What is apparent is a surprising degree of inconsistency with words that have a specialised usage being signalled in the individual entries, but not being on the list of scientific words. Conversely a number of words in the list have no subject label. Of the 266 word

forms listed, only 69 have a subject label. In other cases the scientific usage is given implicitly or explicitly in either the definition or the example, but 58 have no indication of a specialised status at all, which makes it difficult for the learner to see why it is useful to 'know' such words. The naming of sciences is also variable. It does seem odd that geology should be in the list, but not biology, whereas both are defined in the dictionary. In other cases, when a paradigm exists only part of the paradigm is indicated for specialised usage. For example, although the paradigm *gravity* includes *gravity, gravitation, gravitational,* only the adjective carries a usage note, and it is not in the Oxford list.

In order to map these inconsistencies it is necessary to map all the entries and also to compare with other dictionaries. This is the subject of research, but the aim here is not to present a full discussion of science subject labels in these dictionaries, but to look at a number of words that are in the list and a some that are not and to compare usage with that of a specialised context.

3 Specialised Usage

All the modern learner dictionaries are corpus-based, that is to say based on general language reference corpora. The aim of the reference corpus is to give a representative picture of the language at a given time. This means covering a wide variety of genres so as to give a balanced coverage. Consequently, sciences and technical fields are represented as genres rather than by field. It is obviously not possible to cover all areas of science equally and to attempt to do so would be to enter the area of special languages, which is beyond the brief of the reference corpus. Consequently only a very superficial coverage of scientific usage can be obtained and, until a broad based comparative corpus of special languages is built, it will not be possible to fully appreciate what is general scientific usage as opposed to specialised usage. However, it remains interesting to see how the usage reflected in a learner's dictionary differs from what might be found in a more specialised context.

The BIVEG corpus consists of texts from the field of plant biology. It was originally built as the source material for an experimental dictionary in the field of parasitic plant biology (Williams 2002). It is far from representative of science in general, or even plant biology, but, given its highly specialised nature, it is a useful point of comparison against which to check the Oxford scientific 250.

Of the 266 word forms from the list, 195 appeared more than 3 times in the corpus. Fifty eight of these were amongst those with a subject label in the OALD, 85 of the remainder had a scientific usage demonstrated in either the definition or the example. This can mean that either these are very general uses, or that the coverage of semi-technical language is rather good. Amongst the words that did not occur at all in the corpus, only 6 had specific subject labels, and none of these came from biology. If we look at the 58 with subject labels, 12 were labelled as coming from the field of biology, 5 from chemistry, 7 computing, 1 geology, 3 geometry, 4 mathematics, 1 medical, 1 phonetics, and 8 physics. A further 14 came under the general heading of 'technical'. Two others had a variety of attached subject labels; in the case of 'matrix' none of the labels or definitions covered what might be termed biological, whereas 'vector' carried a label for biology and therefore joins the biology grouping. This clearly

displays the multidisciplinary nature of science. The label 'phonetics' does not, however, show the influence of the language sciences as 'cluster' is used in the more general sense of a group of similar things. In order to go deeper it would be necessary to look closely at each different word in context, but this will be the subject of another article. In this study, I want to take just a small number of words amongst those present, and not, to see how the definitions compare across dictionaries and with the corpus.

The most frequent words with labels show clearly the nature of the corpus; these are 'gene', DNA, sequence and enzyme. DNA is labelled chemistry, whilst the others are attributed to biology. Instead of these, we shall look at the list alphabetically taking the letters *a* to *c* so as to compare the situation with two other dictionaries, the CCALD and MEDAL. This gives a list of 9 words, three of which are already labelled as biology and so will not be considered. The remaining six are: 'acute', 'array', 'axis', 'bond', 'clone', 'cluster' and 'compound'. Three other words not in the list will be dealt with in the next section, these are 'antibody', 'bank' and 'control'.

As table 1 shows the different dictionaries have different policies as to subject labels. The OALD is the most explicit for these words, in other cases the scientific meaning is displayed either explicitly or implicitly in the definition or example, in such cases the learner is expected to guess the meaning from the context. This should not be a problem for the advanced user. However, if we compare with the usage in BIVEG, the situation becomes more complex with the mapping from dictionary to contextual sense less clear.

	OALD	CCALD	MEDAL
Acute	Geometry	Implicit in definition	Implicit in definition
Array	Technical	Technical	General usage
Axis	Geometry	Implicit in definition	Explicit mathematics
Bond	Chemistry	Implicit in example	Science, chemistry
Cluster	Phonetics	General usage	Technical
Compound	Chemistry	Chemistry	Science, chemistry

Table 1. Comparison of subject labelling across three dictionaries.

Obviously the dictionary gives the more general sense; the reader must interpret following the context. In the case of 'acute' the geometrical definition refers to the angles of plant parts and is easily understandable as such. 'Array' takes the general rather than the technical sense given in the dictionaries as it refers to a collection of things. The use of the noun 'variety' in the OALD example helps with the BIVEG context. 'Axis' is used in the sense of a graph, again context words such as horizontal and vertical, which are also found in the scientific texts, assist in interpretation. 'Bond' and 'compound' do not present a difficulty as both are used in the chemistry sense labelled in all three dictionaries. However, in the case of 'cluster', even if we eliminate the phonetic sense, there remains the problem that the technical sense is not that used. The technical definition given in MEDAL refers to cluster as an event, whereas the sense used here is one of a group of things. The scientific use of this general sense is implied in both OALD and MEDAL through the definitions and examples, but not at all in CCALD.

What this small investigation shows is that subject labels must be used with care as in some scientific contexts it is the general sense, not the given technical meaning, which is used. In such cases, the user cannot really be said to 'know' the words as the scientific label is only partial. It is obvious that students in the sciences are going to find uses that go beyond the general language meanings found in the learner's dictionaries, quite apart from the specialised terminology that goes well beyond the brief of such dictionaries. However, it should be possible to build a more science-oriented learner's dictionary based on the successful models for general usage (Williams 2002), and also to broaden slightly the definitions for scientific usage in standard learner's dictionaries so as to better help students in this area. This is what I term 'tweaking the prototype'.

4 Tweaking the prototype

Prototypes have been in vogue for some time and their potential use in linguistics is well documented (Kleiber 1990, Taylor 1995). The model that seems to hold most relevance for the analysis of meaning in context is based on Wittgenstein's notion of family resemblance. It is this model that was used by Hanks (1994, 2000) to show how material extracted from corpora could be used to build definitional prototypes that could handle the problem of polysemy. As Hanks demonstrates with the word 'bank' (2000:209), both general and scientific usage can be catered for with a single prototype.

Given that many of the words used in scientific research contexts are not homonyms, but simply examples of specialised polysemy, they may not require an entirely new entry but simply a widening of either the definition, or the example. To overcome the problem of specialised polysemy the answer lies using Hank's prototype definition, in which case a specialised usage will only require an extension, or slight tweaking, of the prototype. This can be seen with the word 'bank'.

According to Hanks (2000: 209) the prototype for bank contains a number of characteristics, a bank:

- is an institution
- is a large building
- for storage
- for safekeeping
- of finance/money
- carries out transactions
- consists of a staff of people

This prototype, it is claimed, can easily handle scientific variants, such as *sperm banks*, where the notion of safeguarding something for future use can be activated. The definitions given in the learner's dictionaries reflect this in that the following definitions are found

- OALD – an amount of sth that is collected; a place where sth is stored ready for use: a bank of knowledge *a blood / sperm bank*
- MEDAL – a store of something that is available for use when it is needed: *a blood/ gene/sperm/organ bank*

However, concordance lines from the BIVEG corpus reveal two uses of bank; *clone bank* and *seed bank*. Concordances for 'clone bank' show that the prototype can be easily applied. The situation for 'seed bank' is different in that a *seed bank* is not something that is voluntarily stored, but a reserve of seeds that builds up naturally in the soil and which can be activated by natural conditions at a later date. In the case of weeds, this is far from being a good thing, hence the gardeners adage of "one year's weed, seven years seed". To overcome this an extra line indicating 'a reserve of something' can be added the prototype, which then becomes pertinent. The change is a small one, but essential to the correct understanding of 'bank' in this context. Such an extension would consolidate the prototype and be as valid in the general as in the technical environment.

The situation is different with a more specialised term as 'antibody'. As can be seen below (table 2), the three dictionaries agree on what these are, but do not feel the need to signal them as scientific. This does seem odd as some very general language words are signalled for a scientific use, whereas a blatantly technical term is not. The problem for the biology student in this case is that the reader may well come across plant antibodies, something which goes against the established prototype. However, in this case only a specialised learner's dictionary would need the extended definition in which mention of artificially introduced antibodies are part of biological research techniques.

OALD	a substance that the body produces in the blood to fight disease, or as a reaction when certain substances are put into the body.
CCALD	**Antibodies** are substances which a person's or an animal's body produces in their blood in order to destroy substances which carry disease.
MEDAL	a substance that your body produces in your blood to fight illnesses and infections. Antibodies are an important part of the **immune system** that protects your body against disease.

Table 2. Definitions for 'antibody'

'Control' is another word with a distinct scientific usage that is not in the Oxford 250. It is, however, clearly signalled in both OALD and MEDAL as having a particular scientific usage. Only CCALD does not signal this. In OALD, the term gets the mention 'in experiment' and with a 'technical' subject label, in MEDAL, it is signalled as 'in scientific tests'. Both dictionaries give a clear definition with an example. This is fine for general usage, but the definition would need enlarging in the specialised learner's dictionary so as to take in to account other uses of 'control' in science. These are variations on the general noun and verb uses, but to be usable in scientific writing a wider variety of specialised contexts of use would be needed.

What is true for this word must be true of numerous other forms of scientific usage, which means that rather than reworking dictionaries and terminologies to present so-called scientific usage it is possible to simply adapt resources by resorting to specialised corpora, which in themselves will enrich the general language prototypes.

5 Conclusion

What role then do these word lists play? Are they designed to help non-scientists understand popular science or to help scientists gain a greater awareness of language? The answer

is most likely the first in that those texts with usage notes will potentially help the reader decide as to which field the particular meaning is used. But what about the unassigned words? Some are simply rhetorical, but, as scientists are assumed to argue logically, logical connectors of discourse are included Other word in the lists simply belong to the large perceived register of science and therefore the reader is expected to guess the field of use from the definition or example. In all cases the aim seems to only be comprehension as examples of usage, so useful in production, are often missing. This leads me to a second question. The fact that a great majority of the words in the list are to be found in a highly specialised corpus of published scientific research seems to point to the fact that many must be either very general words or a genuine scientific core vocabulary. If we take the second premise then the question arises as to whether an advanced learners dictionary could be adapted to the needs of the advanced learner of scientific English. To do this a minor tweaking of the definitional prototype and the inclusion of more scientific examples may be all that is needed. Maybe it is time for a fruitful collaboration between special language corpus linguists and dictionary makers, the market for language specialist learners is big, and that of the non-specialist learner must be vast indeed.

References

A. Dictionaries

OALD – Oxford Advanced Learner's Dictionary. 7th Edition 2005.
CCALD – Collins COBUILD Advanced Learner's English Dictionary.
MEDAL – MacMillan English Dictionary for Advanced Learners. 2002.

B. Other Literature

Hanks, P. (1994), 'Linguistic Norms and Pragmatic Exploitations or, Why Lexicographers Need Prototype Theory, and Vice Verse.' *Papers in Computational Lexicography: Complex 94*, pp. 89-113.
Hanks P. (2000), 'Do word meanings exist?', in Kilgarriff & Palmer (eds) 2000. Sensival: Evaluating Word Sense Disambiguation Programmes. *Computers and the Humanities*. 34/1-2, pp. 205-215.
Kleiber, G. (1990), *La sémantique du prototype*. Paris, Presses Universitaires de France.
Taylor, J.R. (1995) (2nd edition), Linguistic *categorisation: Prototypes in linguistic theory*. Oxford, Clarendon Press.
Williams, G. (2002), 'Corpus-Driven Lexicography and the Specialised Dictionary: Headword Extraction for the Parasitic Plant Research Dictionary.', in Braasch & Povlsen (eds) 2002, *Proceedings of the Tenth EURALEX International Congress*, pp. 859-864.