Making a Learner's Dictionary of Academic English

Diana Lea
Oxford University Press
diana.lea@oup.com

Abstract

This paper gives an account of the development of the *Oxford Learner's Dictionary of Academic English*, a dictionary for non-native-English-speaking students who are studying academic subjects at tertiary level through the medium of English. First, a corpus of academic English was created, using high-quality texts from a broad range of disciplines, maintaining a balance between textbooks, containing typical student reading material, and journal articles, modelling expert academic writing. Drawing on the corpus, and on previous research in the field, a core headword list of "general academic vocabulary" was drawn up. This list was continually supplemented with necessary defining words, complements, collocates, synonyms and antonyms of these words as the work of compiling the dictionary entries progressed. In compiling the entries, particular challenges were encountered in reconciling the academic and pedagogic requirements of the dictionary. This can be seen, for example, in decisions about sense division and the wording of definitions and also in the selection of example sentences from the corpus. Editors had to find ways to represent academic language faithfully whilst making it accessible to learners. The result is a genuinely academic learner's dictionary that should offer real help to learners with their academic writing.

Keywords: learner's dictionary; academic English; EAP; corpus

1 Introduction

Academic vocabulary has received considerable research attention, in particular with the effort to identify a core academic vocabulary, as distinct from general English vocabulary on the one hand and discipline-specific technical vocabulary on the other. Coxhead (2000) proposed the Academic Word List (AWL), a list of 570 word families, divided into ten sublists, found to account for around 10% of the words in a corpus of academic English, as opposed to 1.4% of the words in a fiction corpus. The AWL was generally well received by teachers and has been quite widely exploited in published materials (Coxhead 2011). More recently, however, Paquot (2010) and Gardner and Davies (2013) have proposed alternative lists, addressing some of the perceived shortcomings of the AWL, notably its exclusion of the 2,000 word families of the General Service List (West 1953) as already "known" to students at this level, and its construction around whole word families, regardless of discrepancies in frequency (Paquot 2010: 17) and even core meaning (Gardner and Davies 2013: 3) between different word family

members. Hyland and Tse (2007), however, have questioned the whole validity of a single, cross-disciplinary, core academic vocabulary, partly on the basis that the same words may be used in widely different ways in different disciplines.

In contrast, there has been much less attention paid to the idea of a dictionary of academic English, as opposed to a word list. Kosem (2008) surveyed a number of dictionaries marketed for university students, mostly aimed at native speakers, and concluded that, apart from supplementary material on academic writing, these dictionaries differ little in content from the general-purpose dictionaries on which they are based. Learners' dictionaries, such as the *Longman Dictionary of Contemporary English* and the *Oxford Advanced Learner's Dictionary* have started to acknowledge the interest in academic vocabulary and writing, by marking words in the AWL and including their own academic writing supplements, but they remain essentially dictionaries of general English. In this paper, I shall give an account of the *Oxford Learner's Dictionary of Academic English (OLDAE)* (2014), which I believe to be the first widely available, genuinely academic, learner's dictionary.[1]

I shall begin by outlining the principles and parameters that were established at the start of the project. I shall then describe some of the challenges that were encountered in the course of the project, notably building the academic corpus on which the dictionary is based; determining the headword list; and above all, reconciling the academic and pedagogic requirements, especially with regard to writing the definitions and selecting the example sentences. I shall then conclude by evaluating the achievements and limitations of the dictionary and suggesting some possible future developments.

2   Principles and Parameters

The principles and parameters of *OLDAE*, as a learner's dictionary of academic English, will be found to differ quite widely from those proposed by Kosem (2010). Although I agree with the general thesis that no one is a native speaker of academic English and that both native and non-native-English-speaking students should be viewed as "apprentice writers" when it comes to academic writing (Kosem 2010: 49), I believe the particular needs of these two groups are sufficiently different that a single dictionary cannot serve both equally well. *OLDAE* is therefore designed to serve the needs of non-native-English-speaking students of English for Academic Purposes (EAP), at a range of levels from the B1 student on a foundation course, to students at C2 level writing their Masters' dissertations. The dictionary is also, however, partly for practical reasons, much smaller in scale than that proposed by Kosem. It largely excludes general and technical English, focusing essentially on a core academic vocabulary across the disciplines, but taking a broad view of what this might encompass,

---

1   I am aware of the Louvain EAP Dictionary (Granger and Paquot 2010), an innovative online dictionary-cum-writing-aid, which can claim to be the very first learner's dictionary to be based on analysis of academic corpora. However, it is currently only available to staff and students at the Université catholique de Louvain; moreover, with only around 900 headwords, it may be arguably more a writing-aid-with-dictionary-entries than a complete dictionary in itself.

not attempting to identify a single definitive list for all students, but exploring the nuances of usage of so-called "general academic" words across different disciplines. It is specifically intended to help EAP students with their academic writing across a range of genres. The most fundamental principle underpinning the dictionary was that it should be based on a thorough analysis of genuine academic writing; this meant constructing a new corpus, the 85-million-word Oxford Corpus of Academic English (OCAE).

3   The Corpus

Output from a corpus can only ever be as representative and appropriate as the corpus itself; the content of OCAE therefore needed to match, as nearly as possible, the materials that target users of the dictionary would be reading and writing themselves. In terms of "reading" content, this was provided by higher education textbooks, mostly aimed at undergraduate level, which at 42 million words constituted just under half the corpus. "Writing" content was more challenging: ideally, what was needed was some 40-50 million words of very high-quality student essays and dissertations, but creating such a corpus was beyond our resources. Instead, we substituted expert academic writing from journals, monographs and handbooks, adding up to a further 43 million words; this assured the quality, and meant the findings on the usage of academic vocabulary would be sound. However, it did mean text of a much higher level than our students would be attempting to write, which made the selection of authentic but user-friendly example sentences rather more challenging – but I shall return to this point later. In terms of the balance of disciplines in the corpus, we tried to match this approximately to the profile of disciplines being studied by international students at English-medium universities. Figure 1 shows the breakdown of the corpus into different subject areas.
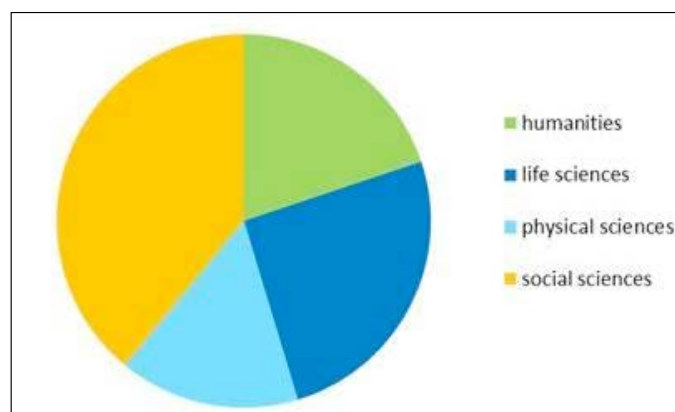

Figure 1: Breakdown of texts by subject area in the Oxford Corpus of Academic English.

Natural sciences (divided into life sciences and physical sciences) and social sciences each account for around 40% of the corpus, with the remaining 20% made up of humanities texts. The largest single disciplines were business and medicine, at around 8% each.

The Headword List

The headword list was built up organically as the work of compiling the dictionary progressed. We began with a core list, comprised of words from the AWL, after checking them against the corpus. We added to this four word lists of our own, extracted from the four subcorpora of OCAE, as compared with a fiction corpus. As work began on compiling entries for these words, the headword list was rapidly augmented with necessary defining words, complements, collocations, synonyms and opposites of the words in the initial list. Collocations were an especially rich source of additional headwords. If dictionary users were to be enabled to use the core words productively, the linguistic contexts in which they could be used would be all-important. This meant generous treatment of collocations in the dictionary entries: for nearly 700 of the most important, collocationally prolific words, a separate section of the entry lists collocations in the style of a collocations dictionary (see Figure 2). These collocation entries then fed back into the main headword list of the dictionary as it was obviously important that no word should be listed as a collocation without being defined and exemplified in its own separate entry. In this way, the headword list expanded beyond a core of 3-4,000 academic words, to encompass higher-level words that are more context-specific, though still mostly at the subtechnical level, as well as presenting in appropriate academic contexts the functional words that are actually basic to all forms of discourse. The process of enhancing the headword list continued throughout the compiling and editing process and was completed by a trawl through the corpus for items of a certain level of frequency that had not been picked up and that might also warrant inclusion. It would have gone way beyond the scope of the project to include every word that a particular student might wish to look up; the aim was to give thorough coverage to the core words that all students would need, plus a generous helping of supplementary words that would be useful to many.

Figure 2: Collocations of *factor* from the *Oxford Learner's Dictionary of Academic English* (2014).

5   Editorial Policy

There is insufficient space here to give a full account of the editorial policy of the dictionary. However, many of the challenges can be considered as different manifestations of the one central challenge: how to reconcile the academic and pedagogic requirements of the dictionary. We wished to make this both an academic dictionary and a learner's dictionary, but there were cases where a compromise was called for. I shall focus on two key aspects here: the definitions and the example sentences.

5.1   Definitions

The starting point for most of the definitions in the dictionary was the definitions in the 8th edition of the *Oxford Advanced Learner's Dictionary (OALD)* (2010). These definitions have the advantage of accessibility for learners, especially as they are written within a carefully controlled defining vocabulary of 3,000 words (in fact reduced to 2,300 for *OLDAE*). In many cases, *OALD* definitions were retained unchanged. However, there were important decisions to be made, often not only over the wording of definitions but over the content. One example is that of *variable* as both noun and adjective. The *OALD* entry (Figure 3) distinguishes two separate senses for the adjective and just one, coverall sense for the noun. The derivative *variably* is nested in the adjective, undefined.

OLDAE, however, recognizes that, for academic purposes, *variable* is a very important word, and EAP students need to know a lot more about it. The noun entry (Figure 4) separates out two further, much more specific

meanings that are important in an academic context: one that is relevant to all experimental sciences, and a basic meaning in mathematics that students from a wide range of disciplines will need to know. As well as conveying much more information than the *OALD* entry, this splitting of senses enables the definitions to be much more precisely worded: students have more to read in this entry, but the burden of interpreting and applying what they have read is actually lighter, because these definitions spell things out much more clearly. The adjective entry (Figure 5) distinguishes four separate meanings, each with different synonyms and antonyms, whilst *variably* (Figure 6) is recognized as an important academic word in its own right, not just a derivative of *variable*, and is given its own entry with two distinct meanings.

Figure 3: Entry for *variable* from the *Oxford Advanced Learner's Dictionary,* 8th edition (2010).

Figure 4: Entry for *variable, noun* from the *Oxford Learner's Dictionary of Academic English* (2014).

Figure 5: From the entry for *variable, adjective* from the *Oxford Learner's Dictionary of Academic English* (2014).
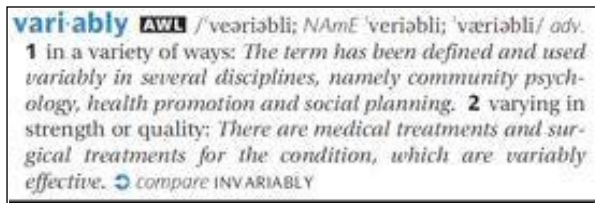
Figure 6: Entry for *variably* from the *Oxford Learner's Dictionary of Academic English* (2014).

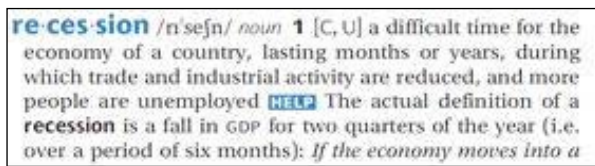A slightly different challenge is posed by a word like *recession* (Figure 7):

Figure 7: From the entry for *recession* from the *Oxford Learner's Dictionary of Academic English* (2014).

The main definition that comes first is closely based on the definition of this word offered in the *OALD*. However, as our economics adviser pointed out, it is not, strictly speaking, a definition at all, but a description. On closer inspection, this will be found to be true of many "definitions" offered in general learners' dictionaries, and the dictionaries in general are all the better for it. They offer learners the degree of understanding they need in a form that is accessible to them. For the EAP student, however, the case is different. The student of economics (or history or geography or a number of related subjects) is not well served by a mere description of a recession, when it is in fact a very precisely defined economic term. Our solution was to offer the description first, followed by the "actual definition", clearly signalled as such. (The economics adviser, it must be confessed, was not happy with this solution, and felt that only the exact definition should be offered; however, this was a view offered from an entirely academic perspective, with no concession to the particular needs of foreign learners, and so the editor's view – that both general description and precise definition should be offered – prevailed.)

5.2 Example sentences

Selecting the example sentences was probably the most challenging aspect of compiling most entries and policy on this evolved over the course of the project, in some cases necessitating late revision of earlier compiled entries. Consultation with academics and EAP tutors at the planning stage impressed on the editors the need for extreme caution when lifting and editing examples from the corpus. Some were uncomfortable with the idea of editing corpus text at all. However, when faced with the reality of raw corpus text, set against the practical needs of the intended users of the dictionary, it became clear that many of the selected corpus examples would need some degree of editing to render them useful and appropriate for learners. Potential difficulties with unedited corpus text were numerous: very high-level vocabulary; difficult constructions; extremely long sentences; obscure and dist-

racting detail; general oddness. Editors also had to take into account the fact that the academic genres in the corpus – textbooks and journal articles – were not the genres that students themselves would be writing. Textbook examples were often tempting, as they were clear and accessible, but many textbooks employ a tone of "expert speaking to student" that would not be appropriate in a student essay.[2]

We initially approached the task of selecting example sentences from an academic corpus with the feeling that it was in some way a different task from selecting examples for a general learner's dictionary from a general English corpus. Experience persuaded me, however, that this task was not in fact different in kind, though perhaps it was different in level of difficulty. The most useful examples are the most typical, which often means the most general:

> Taylor makes the following argument:...
>
> This approach yields dramatically lower estimates.
>
> Several other factors played a role in the decision-making.
>
> The most persuasive argument against this idea comes from Foster (2009).

Examples like this may not be taken directly from any one text; often they are a distillation of a number of different concordance lines, all of them very similar. Other examples – the majority – do contain context derived from a particular source text; and, where appropriate, may be taken from that text unedited. This helps them to feel more authentic; nonetheless, it is important that the context does not get in the way of understanding the linguistic point being presented in the example. The examples are intended to "feel authentic" but they cannot actually be authentic – even if completely unedited, they are inauthenticated the moment they are taken from their context and set in italic type in a learner's dictionary. Ultimately, though, the needs of the learner trump other considerations. Learners using this dictionary are not expected to immediately start writing fluent expert academic texts; what they need to acquire is a style that approaches more closely an appropriate academic style, whilst still being accessible from the level they are currently at.

6   Conclusion

The learner's dictionary is not an academic genre but a pedagogic one. A learner's dictionary of academic English needs to pay close attention to the rules and conventions of academic writing, and represent them as faithfully as it can, but the learner's needs still take precedence. *OLDAE* is designed

---

2   The use of "we" is a case in point. Textbook writers use it frequently with the meaning of "you and I", to include the (student) reader in a comment such as, "In this chapter, we shall see ..." This style is not employed in research articles, where experts are writing for other experts, and it is not to be recommended for students writing for a tutor or examiner. A usage note at the entry for we in the dictionary explains this point.

to meet the specific needs of tertiary level students writing assignments in English in a wide range of disciplines. It covers a generous "core" academic vocabulary, showing not only the meanings of words, but how to use them in context. We hope it will be a valuable new resource for students. Its limitations are largely those imposed by the relatively limited size and scope of the project. Coverage does not go much beyond the "subtechnical" level of vocabulary, but it is assumed that the technical vocabulary of the student's own discipline will be explained as part of their subject course. The traditional format of print dictionary plus CD-ROM may also limit its appeal for some of today's students. However, there is a lot of scope for presenting, combining and expanding the content in different ways to make it even more useful and accessible to a wider range of users. For example, a customizable online subscription model could allow users to combine *OLDAE* content with both more general content and more technical content from other dictionaries, according to the subject they are studying. To make this a reality would require rather more work, both editorial and technical, but it seems worth aspiring to.

7   References

Coxhead, A. (2000). A New Academic Word List. In *TESOL Quarterly.* 34(2), pp. 213-238.

Coxhead, A. (2011). The Academic Word List 10 years on: research and teaching implications. In *TESOL Quarterly*, 45(2), pp. 355-362.

Gardner, D., Davies, M. (2013). A New Academic Vocabulary List. In *Applied Linguistics.* First published online August 2, 2013. Accessed at: applij.oxfordjournals.org [25/03/14].

Granger, S., Paquot, M. (2010). The Louvain EAP Dictionary (LEAD). In *Proceedings of the XIV EURALEX International Congress, 6-10 July 2010.* Leeuwarden, The Netherlands, pp. 321-326.

Hyland, K., Tse, P. (2007). Is There an "Academic Vocabulary"?. In *TESOL Quarterly.* 41(2), pp. 235-253.

Kosem, I. (2008). Dictionaries for University Students: A Real Deal or Merely a Marketing Ploy? In *Proceedings of the XIII EURALEX International Congress, 15-19 July 2008.* Barcelona, Spain, pp. 1575-1584.

Kosem, I. (2010). Designing a model for a corpus-driven dictionary of Academic English. PhD thesis. Aston University, Birmingham, UK.

Longman Dictionary of Contemporary English. 5th edition (2009). Harlow: Pearson Education Limited.

Oxford Advanced Learner's Dictionary. 8th edition (2010). Oxford: Oxford University Press

Oxford Learner's Dictionary of Academic English (2014). Oxford: Oxford University Press

Paquot, M. (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis.* London, New York: Continuum International Publishing Group, pp. 11-17, 55-61.

West, M. (1953) *A General Service List of English Words.* London: Longman, Green and Co.