Monolingual Dictionaries as a Learning Tool: An Examination of the Collins Cobuild Advanced Learner's English Dictionary

<table>
<thead>
<tr>
<th>Language</th>
<th>総合政策研究</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>未定</td>
</tr>
<tr>
<td>Language</td>
<td>英語</td>
</tr>
<tr>
<td>Type</td>
<td>研究</td>
</tr>
<tr>
<td>Category</td>
<td>未定</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10236/13013">http://hdl.handle.net/10236/13013</a></td>
</tr>
</tbody>
</table>
Monolingual Dictionaries as a Learning Tool: An Examination of the Collins Cobuild Advanced Learner’s English Dictionary

Darby McGrath

This paper examines the Collins Cobuild Advanced Learner’s Dictionary and evaluates it usefulness as a language-learning tool. It begins by situating the Cobuild within the context of monolingual learners’ dictionaries in general, before taking a closer look at the philosophy that underpins the Cobuild, and at both the macrostructure and microstructure of the dictionary. Finally, it offers some suggestions as to what teachers can do to enable English learners to get more out of monolingual learners’ dictionaries in general, and the Collins Cobuild Advanced Learner’s Dictionary in particular.

Key Words: Monolingual Learners’ Dictionary, Cobuild

Introduction

There is an extent to which the evaluation of a learners’ dictionary independent of context is of limited value. It ignores the differences between both individual learners (Dalglish, 1995, p. 330), and sociocultural factors. This said, there has been a tendency in recent times to make possibly too much of learner variation. They are still human beings with human minds, and so we can draw a distinction between generally effective and ineffective pedagogical practices. In this sense, there are features of any given monolingual learners’ dictionary (MLD) which support language learning irrespective of context. It is these which will form the focus of this evaluation of the Collins Cobuild Advanced Learner’s English Dictionary (5th edition).

The approach below will also be shaped in part by the observation that there is a divergence between the way lexicographers traditionally represent language and how language ‘really works’ (Hudson, 1988; Rundell, 1999). The tendency of dictionaries to atomise language has been a contributing factor, Hudson claims, in the emergence of a ‘folk linguistics’ which views lexis and grammar as wholly discrete elements (1988). To avoid this, the intention here is to evaluate Cobuild in terms of how its different features each support language decoding and encoding, rather than focusing on how the dictionary supports lexis and grammar individually. This means that the bulk of the evaluation will look at the more important structural features of the Cobuild in isolation. The focus here will be on the more remarkable features of Cobuild, although preceding this is a brief analysis of the benefits and limitations which Cobuild shares with all MLDS.

Cobuild as a monolingual learners’ dictionary (MLD)

Frankenberg (2005) concludes that learners consider MLDS less useful than L1 dictionaries. Much of this probably has to do with the security offered by reference to L1. Beyond this, however, is an unavoidable consequence of the monolingual format. Simply put, if a user does not know a word, how will he or she know where to find it? (Bogaards, 1996; Rundell, 1999) This issue compromises the dictionary’s usefulness as a tool for encoding language. It is difficult to suggest a solution, beyond the possibility that learners use Cobuild in conjunction with a bilingual dictionary.

At the same time, what MLDS lack in usability here they compensate for in precision.
Anisomorphism presents a problem for users of bilingual dictionaries, inasmuch as genuine synonymy between languages is unusual. Monolingual dictionaries avoid the misconceptions this can create.

**Dictionary macrostructure**

Taking a ‘top-down’ approach, it is natural to begin with breadth of coverage. *Cobuild* purports to include “over 110,000 words, phrases and definitions, all simply explained in full sentences”. Bogaards on the other hand, estimates the number of “lexical units defined to be much lower, somewhere between seventy and seventy five thousand (1996). To be fair, the discrepancy may stem from the fact that this estimate seems not to include phrasal units. However, Bogaards also takes *Cobuild* to task for poor coverage of technical words, old words and Americanisms (1996). Some of the omissions he notes have since been included (‘grungy’, for example), but this aside there is the question of how broad the coverage of a MLD should be. Is it in the learners’ interests to include such words as ‘bulrush’ if they compromise the space given to more useful entries? The answer depends largely on the requirements of the individual learner.

The two remaining elements here have consequences for both the macrostructure and the microstructure of the dictionary. The first of these is the extent to which *Cobuild* ‘nests’ derivatives under a single headword. For the most part, derivatives are nested, with the exception of compounds. Landau makes the point that extensive nesting can have advantages to the user in terms of both decoding and encoding (2001, p. 365). The fact that the practice is “based on morphological similarity” allows learners to draw conclusions based on existing knowledge of the headword (assuming any exists). This approach has a number of analogues in the language classroom; it can be seen as the lexicographical equivalent of ‘scaffolding’. In terms of encoding, nesting forms is useful when the learner does not know the exact form of a derivative (Landau, 2001, p. 365).

An argument against nesting, that it can make units difficult to find when the learner does not associate the headword with its derived form, is largely circumvented in *Cobuild* by using both nesting and separation. As an example of this, we find that the suppletive forms ‘wept’ and ‘went’ are listed as derivatives under their canonical forms, but they are also given individual entries. A similar device is used in the treatment of compounds. If the first word of the compound corresponds to the headword, it is not nested: ‘horseplay’ is treated only in isolation because it follows so soon in the dictionary after horse. ‘Clotheshorse’, on the other hand, is cross-referenced because the entry is located elsewhere. However, this does not help with finding idiomatic phrases, which tend to consist of a number of words and so to the learner might be in any number of places. *Cobuild* nests idiomatic phrases, but Bogaards argues that the choice of headword is not consistent enough to make this useful (1996).

It is difficult to see a solution to this beyond the commercially unfeasible listing of phrasal elements under all the constituent words.

A second element that affects both macrostructure and microstructure is the treatment of polysemy and homonymy. To a degree *Cobuild* can be seen as continuing a trend among MLDs to dispense with the distinction between these relationships (Landau, 2001 p. 101). The entry for ‘ball’, for example, includes all senses from “a round object” to “a large formal social event”. One likely benefit of this is that it reflects the way language is ordered in our minds. Theoretical and psycholinguists have made the claim that to draw too clear a distinction between different uses of the same word is artificial, inasmuch as there is evidence that these words are often connected in a variety of ways in our ‘mental lexis’ (Hudson, 1988). However, in the case of longer entries this practice is not followed. Here the units are divided under “super headwords” (Bogaards, 1996), avoiding the need to “wade through” large amounts of information to locate the required sense (Scholfield, 1999). This is more significant than it might initially appear. A number of studies have indicated a tendency on the part of learners to look no further than the first sense when trying to decode language (Bejoint, 1981; Nesi & Haill, 2002; Rundell, 1999). With this in mind, any feature which encourages learners to look further is a positive step.

**Dictionary microstructure**

While an effective macrostructure determines the accessibility of information, the information itself is a feature of the microstructure. To this extent, the microstructure is more intimately connected with how well a dictionary supports language learning. *Cobuild’s* microstructure is shaped substantially by the fact that it, in Landau’s words, “[bet] the farm” on the use of the authentic language taken from the Bank of English (BOE) corpus (2001). This reliance is the source of many of the dictionary’s strengths
and weaknesses.

Using the BOE Cobuild has created a word frequency list, which in turn is used as the basis for selection, ordering senses within an entry, and it's system of 'frequency diamonds'. There is something instinctively appealing about this. For the purposes of encoding, common words are very important, as they comprise so much of the syntactic form of what learners aim to produce. However, there are a number of question marks over the use of frequency data for pedagogical dictionary writing.

One such question is of representativeness: whose language does the corpus use and from what modes of expression is it drawn? Kilgarriff makes the point that, for example, the vocabulary of journalists is overrepresented in Cobuild because of the large amount of "newspaper and newswire material" in the BOE (1997). This is also indicative of a bias towards the written mode.

What this means is that learners are asked to place priority on words that may not be as widely used as they believe. Moreover, even if we accept that the BOE is representative, are frequent words as useful as they believe. Moreover, even if we accept that the BOE is representative, are frequent words as useful as Cobuild implies? One remarkable feature of the literature is how few writers dispute the conflation of frequency and utility; in fact it is so widely accepted that many take it as a given (Bejoint, 1981; Kilgarriff, 1997; Landau, 2001). For the purposes of decoding at least, this seems to be based on rather specious reasoning. Landau (2001, p. 275) makes much of Fries’ study showing that 85% of written text is comprised of only 1,000 words. The implication is that 85% of texts are comprehensible if a learner is familiar with all the senses of these 1,000 words. What is as likely is that on average 8.5 words of a ten-word sentence will consist of these frequent words, but if the other 1-2 comprise the ‘content’ words (many of the most common words are ‘structure’ words), which are unknown to the learner, then the other 8.5 will be meaningless. In other words, it is conceivable that a learner will be familiar with 85% of a sentence, but understand nothing of its meaning. Consider the following, admittedly contrived but nonetheless plausible, example sentence:

There is some lobster in the fridge.

All the words bar the two nouns have three diamonds in Cobuild. The others have none, but clearly constitute the more communicative of the following ungrammatical expressions.

There is some in the lobster fridge

That frequent words are not necessarily more useful seems in part to be reflected in usage. There is evidence to suggest that learners have an overwhelming tendency towards looking up infrequent words. Nesi and Hall (2002) report that only 32% of ‘look-ups’ in their study involved words occurring more than 10 times per million in the British National Corpus, and Bejoint (1981) reports little or no interest on the part of his students for frequency information. This is not to say there is no value in frequency data, only that it is more useful when applied to content words.

Frequency also provides the framework for the ordering of senses within polysemous entries. This is an organisational feature rather than an informational one, but it has consequences for the learner. Firstly, it means that dictionary searches are likely to be less disruptive (Bejoint, 1982) (although again this assumes that learners will be looking for the more frequently used senses). This is valuable for decoding in particular, as excessive time spent looking up a word tends to detract from a learner’s overall textual comprehension (Hosenfield as cited in Scholfield, 1999). Cobuild’s method here also has its detractors because frequency makes no allowance for part of speech, leading to a scattered effect which Stanop calls “chaotic” (1988). Scholfield is more measured in pointing to research indicating that part of speech identification is a learner’s best first step in inferencing meaning, and it “seems perverse” to waste this tool by splitting parts of speech (1999).

In Cobuild’s case, however, it is tempting to agree with Bejoint. This is largely because part of speech is listed in the ‘extra column’, separated from the bulk of the text. Few demands are therefore made on the learner in identifying part of speech.

This extra column, unique to Cobuild, has broad approval in the literature (Bogaards, 1996; Landau, 2001; Rundell, 1998; Stanop, 1988). Cobuild itself divides the extra column into four different features. Of these, the ‘frequency diamonds’ have already been discussed. Another, ‘information on pragmatics’, consists of a set of seven labels which are designed to show the connotative meanings of the sense. The labels ‘approval’ and ‘disapproval’ in particular, will be useful in helping learners both decode such ostensibly neutral terms as ‘cosmopolitan’ (approval) and ‘cosmetic’ (disapproval).

The inclusion of superordinates, synonyms and antonyms in the extra column runs the risk of exaggerating what may be only a partial relationship. For example, while there is certainly is a relationship of sorts between ‘cosmopolitan’ and ‘parochial’, to list them as antonyms is misleading. In semantic terms, the former describes a state and the latter an attitude. However, this problem is possibly
outweighed by evidence that such devices reflect the way L1 vocabulary is stored in the mental lexis (Rundell, 1998).

Arguably the most significant feature of the extra column is the inclusion of grammatical information. In terms of presentation, opinion is divided over the quality of this information. Houseman and Gorbahn comment that the coding is “not immediately comprehensible” (as cited in Chan & Taylor, 2001), while Aards approves of the system because it includes “a minimum number of transparent symbols” (as cited in Bogaards, 2001). Houseman and Gorbahn’s criticism applies to the short term, so we can assume that if the learner takes the time to learn the coding, the information will prove useful. A problem is that few users seem to do this (Bogaards, 2001), perhaps because of the appearance of complexity.

Bogaards’ results might be explained by a preponderance of research participants applying a global rather than analytic learning style. A reality of the extra column treatment of grammar is that it does not serve field dependent learners well because it reduces syntax to decontextualised formulae. Cobuild goes some way to supplying a solution through its use of illustrative quotations (IQs) and its defining style. Each of these features provides what is essentially a mini-context, giving the learner comprehensible information for encoding a lexical item. The IQs do this by providing an authentic example of the language in use, while the ‘sentence’ defining style gives direct information about possible subjects, objects, transitivity and such like. For those daunted by the extra column, this provides a valuable alternative.

The usefulness of these features for decoding, and even encoding beyond the level of syntactic behaviour, is more questionable. Landau claims that authentic IQs can assist learners to decode variety of usage (such as humour and formality), connotation, collocation and meaning (2001, p. 208), but it is difficult to argue that this is true of many of many of Cobuild’s examples. The entry for ‘porosity’, for example, includes the IQ: “…the porosity of the coal”. It is hard to see how this explicates any part of Landau’s list. The context provided by the IQs is too limited to be of use in determining variety of usage or connotation. The IQ “London has always been a very cosmopolitan city” gives the learner nothing of the sense of approval that ‘cosmopolitan’ conveys. This example also shows the extent to which the usefulness of the IQs is limited by the learner’s level of sociocultural knowledge. The only thing that a learner unfamiliar with London is likely to gain from this IQ is part of speech.

For decoding meaning then, the learner is often almost entirely dependent on the definition. Cobuild claims that its sentence defining style is easy to understand and “natural” (Sinclair, 2004, p. vii). This is a claim which has not been universally accepted, for two main reasons. The first of these is that the defining style is not suited to all types of lexis, in particular realia (Landau, 2001, p. 180). What this criticism amounts to is that valuable space is wasted in some definitions. A more serious criticism is that many of the definitions, far from being easier, can become extremely complicated (Rundell, 1999; Scholfield, 1999). An example of this is the definition for ‘dwarfed’, which is likely to be difficult even for advanced learners:

*If one person or thing is dwarfed by another, the second is so much bigger than the first that it makes them look very small.*

The truth is however, that this criticism applies to only a minority of definitions. For the most part Cobuild’s defining style is clear, and particularly well adapted to defining metaphorical and idiomatic language (Landau, 2001, p. 179). What it does suggest though is that Cobuild’s “purity of focus” (Landau, 2001, p. 287) can be a disadvantage to the user. The editors should perhaps judge each entry individually and select a defining style accordingly.

A final element that is notable in its complete absence from Cobuild is the use of illustrations. Illustrations use a considerable amount of space, but it seems odd that a MLD eschews them completely, as they are an excellent means of unambiguously presenting concrete nouns and even many ‘action’ verbs.

**The role of the teacher**

The information contained in Cobuild is useful, but because a dictionary is a tool, *use* of a dictionary is a skill. The primary role of the teacher in this sense then is not necessarily to teach students using the Cobuild, but to enable learners to use it autonomously and appropriately, so that they can access the information it contains for themselves. Scholfield (1999) suggests six possible steps in learner use of a dictionary. Of these, the teacher arguably has a role in helping students with five:

*Making the decision to use a dictionary.*

Dictionary use is disruptive. As mentioned, it can detract from a learner’s comprehension of a text. This factor also applies to production, inasmuch as the listener’s patience is tried when a dictionary is used.
The teacher’s role here should be to offer alternatives so that the dictionary becomes a last resort. To this end teachers need to grade texts appropriately and encourage learners to skip words if they are not crucial to overall meaning. For production, learners need to develop a capacity and willingness to use periphrasis to describe an unknown lexical item. Simple describing and guessing games are an excellent way teachers can develop this.

**Finding the relevant entry**

Generally, this will not pose significant problem for users of Cobuild because of the policy grouping all senses under a single headword. Learners will need to know how alphabetising operates beyond the initial letter of each word.

**Finding the correct part of the entry**

Nesi and Hail (2002) report that the primary cause of learner error in their study was the failure to locate the correct sense. Students need to be encouraged to look beyond the first sense, but perhaps the key is for a more integrated use of dictionary and text. Drawing inferences will play a prominent role here. Users can establish much about a word before they even open the dictionary: part of speech, connotation, collocation. As all of these are clearly shown in Cobuild, learners will then be at a position to eliminate most of the inappropriate senses. It will be necessary to help students discover the location of these elements in a Cobuild entry.

**Exploiting the information**

Cobuild includes valuable information in every entry, which is for the most part wasted because learners often ignore the most useful elements (Nesi & Meara, 1994). The role of the teacher here should be not only to familiarise the students with the microstructure of the dictionary, but also to emphasise the need to use all of the features in conjunction to take advantage of Cobuild. One way of achieving both of these aims is activities whereby the students race to find a particular piece of information about a lexical unit and then have to encode it in some way. These can be used to take advantage of definitions and IQs, the IPA, the extra column, and collocation.

**Retaining the information**

This will not always be relevant to dictionary use, as often students will only be using a dictionary as a quick reference. When learners locate a word they consider useful, however, it is important that they have a genuine idea of what it is to ‘know’ a word. Beyond meaning and spelling, for production students will also need part of speech, collocation, pronunciation, connotation and usage, all of which are present in different parts of a Cobuild entry. This will mean recording a lot more than simply an L1 translation. If the class can agree on the words they consider important, then quizzes and production games focussing on all of these elements is an effective way of encouraging more detailed recording of vocabulary.

**Conclusion**

Cobuild 1 has been called “an interesting but not entirely successful experiment” (Stanop, 1988). In 1987-88, much of the criticism stemmed from the fact that the production team seemed so enamoured of corpus linguistics that the dictionary itself became unwieldy. The intervening editions have gone some way to rectifying past excesses by, for example, trimming the length of many of the illustrative quotations. There is no doubt however that problems remain, largely in terms of accessibility and a continued unwillingness to accept that discoveries about language (frequency data, for example) do not necessarily mean advances in language teaching (Cook, 1998). That said, if the learner can develop the skills to consult the information contained in Cobuild in an appropriate way, it certainly offers valuable support for language learning.
REFERENCES


