

BOOK REVIEW

Collins Cobuild English Language Dictionary (Collins Birmingham University International Language Database). Collins, London-Glasgow, 1987. xxiv + 1703 pp. £ 7.95 (paperback).

PIETER A.M. SEUREN

This dictionary embodies an attempt at making (monolingual) dictionaries more systematic, more user-friendly, and more accurately descriptive of the living language (“helping learners with *real* English” it exclaims on the cover). There can be no doubt that the aims set by the makers of this dictionary, largely the English Language Department of Birmingham University, are timely and well-chosen. There certainly is a need for more serious *lexicology* to go into the *lexicography* of dictionary making. One might even say that a modern dictionary should also be organized in such a way that rapid computer-searches can be made on a sufficiently large number of practically useful parameters. Clearly, judicious use of up-to-date theoretical linguistic knowledge may well contribute to success in achieving the aims set. The question now is: to what extent has Collins Cobuild (henceforth CC) been successful in doing what it set out to do?

One clearly detects a desire to provide lively and readily usable meaning descriptions for the entries. The language used for the descriptions is simple, non-pompous, at times even colloquial. The dictionary clearly addresses itself also to users, English speaking or foreign, with limited experience in dictionary use. For this category of users the book is certainly more “user-friendly” than other more established dictionaries. Whether it also has greater “user-friendliness” for more advanced or even professional dictionary users is a different matter. These may detect a mixture of condescension and chumminess in this dictionary, which they would gladly exchange for greater efficiency and accuracy. The informal “you” is all over the pages: “If your thoughts are *fuzzy* or what you are thinking about is *fuzzy*, you are confused and cannot see an idea clearly or make a decision”. I’m not sure that all readers will appreciate being talked to in this manner. What is most helpful and enlightening, anyway, is a generous sprinkling of example sentences in the meaning description sections. For practical purposes of meaning description, few things are more helpful than well-chosen examples.

A great deal of thought and care has obviously gone into the actual meaning descriptions, which are generally clear and well laid out. Yet there is plenty of room for improvement on this score. To give just a few examples, for the verb *flog*, the first meaning given is: “If you *flog* something, you sell it; an informal word used in British English. EG *We thought we might be able to flog it to someone.*” As a second meaning we find: “If you *flog* someone, you hit them very hard with a whip or stick as a punishment for

something which they have done.” First, such descriptions are unnecessarily verbose: the part “if you *flog* someone” may be left out without any loss of information or user-friendliness. And although the element “punishment” does seem to be either prototypical or perhaps even constitutive of the meaning of this verb, the addition of “for something which they have done” is both tedious and misleading (it may also be for something they have not done, or for something that the flogger thinks, or pretends, or what not, that they did or failed to do). But secondly, the fact that the meaning “sell” is given first makes one suspect that the makers of CC have pandered to popular taste perhaps a bit more than they should have. It is the task of a dictionary maker not only to provide information but also to be an authority on what is considered to be accurate use of the words of the language in all sociolinguistic registers, not primarily, or preferably, the “informal” ones. This tendency to pander to popular taste is, unfortunately, present all through CC. It might be added that the description of the informal use of *flog* is misleading because of incompleteness: when one “flogs” something it is because one wants to get rid of it, the thing being considered of inferior quality and not worth keeping; one is then also prepared to accept a reduced price. A naive user of CC might think it appropriate to let a millionaire say, in an informal talk, that he will “flog” his Rembrandt painting to the British Museum.

Or, to take just one other example, the meaning description of the adjective *bald* shows a great deal of care. It is specified for persons (having little or no hair on the top of the head), for objects: “Something that is *bald* does not have the natural covering which you might expect it to have, for example fur or grass, EG . . . *a bald granite outcrop*”, and then in particular for tyres, and for statements, questions, accounts, etc. Yet, although this analysis shows that quite some thought has gone into it, it is also obvious that more is needed. For example, it is not made clear why a tree without its leaves cannot be called “bald”, but must be called “bare”, even when it is an evergreen, whose “natural covering” lasts the year round. The same goes for *a bare hill, a bare piece of rock*. Note that under *bare* the example “*a hilly patch of bare red rock*” is given. A user may well wonder why a granite outcrop is called “bald” but a patch of redrock should be called “bare”, or whether either would do in both cases. The point is that there is a certain division of labour between *bald, bare, nude, naked*, and perhaps a few others, which differs considerably from corresponding word families in other languages (e.g. German *kahl, bloss, nackt*). It would be useful to know to what extent there are unifying principles behind such divisions of labour, and to what extent the combinations are idiosyncratic. In the case of *bald*, for example, one might wonder whether there is any point in postulating that roundness of the object plays a part in the appropriateness conditions for this word (its etymology appears to be “balled”). We speak

of *bald tyres*, *bald heads*, and in some dialects (in America perhaps more than in Britain) of *a bald mountain*. In any case, combinations like *a bald statement* are conventionalized and idiosyncratic: they do not follow from any unifying principle of the lexical meaning of *bald*. Cp. the likewise idiosyncratic collocations *bare detail*, *bare fact*, *bare amount*, etc. Unfortunately, CC does not indicate, for these cases, that they are conventionalized rather than rule-governed.

It must be said in CC's defence that any commercially viable dictionary must strike a balance between investment in research and size on the one hand, and completeness and quality on the other. An ideal dictionary may just require too large an investment for a publishing firm to afford, even in conjunction with a university department. Such a comment is no doubt fair. Yet the question remains whether CC could perhaps not have profited more from already available insights in these matters, or whether the quality could not have been improved considerably with just a little more effort. It has, for example, already been pointed out that CC could easily have crammed more useful information into the same space, due to the repetitiveness and verbosity of the descriptions and the comments. It is clear, in any case, that much more attention should be paid in linguistics, as it is practiced in the universities, to lexicology, i.e. the theoretical study of lexical items, and in particular of lexical meanings. Lexicology should be recognised as a separate subdiscipline of linguistics, along with phonology, syntax, etc. The makers of commercial dictionaries could then draw on the insights and theories developed in theoretical lexicology.

Even so, CC could have done better, in particular as regards systematic lexical categorizations. There is no systematic distinction in CC between negative and non-negative words. Yet it is easily argued that of the following pairs the first member is positive while the second is negative: *assert/deny*, *increase/decrease*, *just/hardly*, *many/few*, *old/young*, *big/small*, *expensive/cheap*. This is not just of theoretical importance, because there is a further distinction in the positive category: some positive adjectives are used "neutrally" when accompanied by a measure phrase, such as "*how ___?*": a question like "How old is the baby?" does not imply that the baby in question is old; it is simply a neutral question about the baby's age. Not so, however, in: "How expensive is that coat?". Here there clearly is an implication that the coat is expensive. It must be noted that such phenomena are not predictable just on grounds of general semantics, because there are interlinguistic differences. The German "Wie teuer ist dieser Mantel?", for example, does not imply that the coat is expensive, even though the nearest translation of *expensive* is *teuer*. Likewise for English *old* and French *vieux*: the French *vieux* cannot be used to ask a neutral question about age, because it always implies advanced age, unlike English *old*. Now although CC does specify for *old* that it can be used neutrally to specify age ("If you

say that someone or something is a particular number of years, months, etc. *old*, you mean that they have lived or existed for that length of time”), whereas no such use is specified for, e.g. *expensive*, it would help the user if a category indicator were applied here, such as “positive neutralizing” for *old* etc., and “positive non-neutralizing” for cases such as *expensive*.

This is just an example, but many more could be provided. Thus, although there is an extensive literature on negative and positive polarity items, CC has in common with all other existing dictionaries that no mention is made of this lexically important distinction. For example, the verb *bat*, in conjunction with *an eyelid*, requires a negation or a (semi)negative adverb (such as *hardly*) for a main clause in which it occurs to be grammatical. CC lists this negatively polar verb as number 5 under the entry *bat* (which, unilluminatingly, has the meaning “cricket bat” as number 1, and our nightly flyer as number 3) in the following way: “If you say that someone did *not bat an eyelid*, you mean that they showed no sign of surprise or concern”. How can the user infer that if one does show a sign of surprise or concern, the expression *bat an eyelid* cannot be used? Conversely, for a positive polarity item like *bristle (with)* it is not specified that a so-called “echo”-effect occurs when this verb is used in a negative sentence: “The place did not bristle with policemen”.

Factivity of verbs or adjectives is not mentioned in any way, even though this notion has been around in semantics for about twenty years now. A verb or an adjective is *factive* when it takes a *that*-clause in subject or object position and presupposes the truth of that clause. Thus the verb *realize* is factive: a sentence like “Henry realized that he had been fined” carries the logical consequence that Henry had been fined. Likewise for predicates such as *know*, *have forgotten*, *regret*, *be surprised*, *be a pity*, *be regrettable*, *be advantageous*, etc. Factivity distinguishes between verbs such as the factive *know that . . .* and the non-factive *be convinced that . . .*

All sorts of lexical processes, which are regular and productive, yet hardly ever fully predictable (if they were fully predictable they would be part of the grammar), are, though often duly reflected in the descriptions, never explicitly identified. Adjectives, for example, are often derived from nouns in such a way that their meaning amounts to a preposition phrase (usually but not always with the preposition *of*) over the noun in question. Thus one has, in English, *monumental sculptor*, which means, apart from “sculptor of monumental proportions”, rather “sculptor of monuments”, Likewise *nocturnal prowler* (“prowl during the night”), *attitudinal change* (“change of attitudes”), etc. Such derived adjectives are invariably limited to attributive use. CC does mention the exclusively attributive use, but not the noun-derived character.

Many adjectives can be used causatively. Thus, *a sad story* is not a story that is in a state of despondency, but rather a story that *makes* one sad.

Likewise for, e.g., *happy ending*, *proud victory*, etc. Not, however, for **glad story*, **nervous event* (but e.g. the Modern Greek for “nervous” does allow for causative use). Again, CC does, on the whole, mention the possible uses (and does so rather better than most other dictionaries), but it does not, unfortunately, mention the categories.

It is a pity that CC has not paid more attention to specific lexical categories of meaning description (and instead of the unnecessary verbose elements in the descriptions). CC has taken a first step in this direction by putting all information of a more technical nature in a separate column to the right of the meaning descriptions, which are given in ordinary prose. This “technical” column could, and in my opinion should, have been a great deal more precise and more complete.

This column contains, for one thing, the information on what is usually called *subcategorization frames* for predicates. These frames specify the number and kind of terms that can go with a verb or adjective. The subcategorization frame specifications given in CC are, however, very often incomplete and sometimes even sloppy. In this respect CC compares badly with, e.g., the Oxford Learners Dictionary.

The main advantage of good and systematic categorial information in a dictionary is that such information makes for a vast increase in possibilities for computer application. A computer-stored lexical database with such information allows for rapid and efficient search procedures, which can be of great value, for example, in computer-assisted translation. It will enable a translator who must turn, say, the Modern Greek equivalent for “nervous light”, i.e. in the causative sense of “nervous”, into English: a simple query will tell him instantly that English *nervous* does not allow for causative use. Adequate lexical categorization will enhance the speed and accuracy of computer-assisted translation. It will enable a translator to know instantly that, e.g., the Dutch equivalent for “a glad face” must be rendered in English as *a happy face*, or that the English phrase *a funny man* is to be rendered in Italian as *un uomo spiritoso* only if it means “a man with a good sense of humour”, but that this same phrase in the sense of “a slightly strange man” will properly correspond to *un uomo un po’ strano*, or *un uomo ridicolo*.

Clearly, lexicology has not developed far enough to provide sufficiently adequate categorization systems for all such cases, but some progress has been made, and one has the feeling that CC could have made better use of this progress. It is anyway important to mention the practical advantages of adequate lexicological categories for writers and translators. A translator who is fully competent in the two languages involved will not or hardly need such help. But such translators are rare and expensive. Given the dramatic increase in the amount of translation work to be done, especially in the context of the European Community, research into ways of facilitating

translating procedures should have a high priority. We know now that fully automatic translation will remain a pipedream for quite some time. But that does not mean that translation has to remain the way it is, entirely done "by hand". The appearance of CC is a welcome opportunity for stressing that it is realistic to work towards the development of computational techniques that will assist translators and generally writers in a foreign language.

As with all dictionaries, one can spot omissions in CC. Not so much in the area of "obscene" terms, notoriously neglected in English-language dictionaries. Such terms are, on the whole, given, though with shamefully deficient meaning descriptions. There still are, however, glaring omissions, such as the expressions *Dutch treat*, *Dutch uncle*, *Dutch wife*. One also misses, under the entry *real*, the meaning "just like but not in reality", as in "John is a *real* diplomat/actor/ . . .", implying that John is *not* really a diplomat, or an actor, or what have you.

On the whole, CC is an interesting dictionary, which may well be quite successful. It is an attempt at opening windows in lexicography, windows that have remained shut for too long. In doing so it opens new and enticing perspectives for dictionary making. If it falls short of what a devoted lexicologist would hope for it is easily forgiven, mainly because it underscores the importance and the feasibility of significant advances in lexicography, based on advances in lexicology.

*Nijmegen University
Philosophy Institute
P.O. Box 9108
6500 HK Nijmegen
The Netherlands*