

The ideal dictionary, lexicographer and user

DAVID CRYSTAL

University College of North Wales, Bangor

The title of this paper was not my idea, but, now that I have it, I am irresistibly reminded of another who was much concerned with idealisation. However, his account is relevant, in the present context, only with a great deal of modification. Lexicographical theory, in the terms of that account, would be concerned primarily with an ideal dictionary writer-user, in a completely homogeneous reading-community, who knows his language very imperfectly, is seriously affected by such lexically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance, and who has a great deal of motivation to learn, a great deal of time, and above all a great deal of money. The parallel is both silly and instructive. I suppose it is silliest in proposing the concept of 'writer-user'. We know that this hardly ever obtains: quite evidently, most dictionary users rarely attempt to write dictionaries; and I am quite sure that most dictionary writers do not have time to use them! But the parallel is instructive in that it does remind us of the theoretical problem facing lexicography as a branch of applied linguistics: namely, how to predict the performance limitations which constrain both parties to the enterprise — lexicographer and user — and to resolve them, so that we obtain an ideal end product, which satisfies everyone's criteria at a minimal cost in effort, time and money.

I speak from personal experience when I talk about the performance limitations of lexicographers. I have tried to write a dictionary on five occasions, in the last twenty years, and succeeded only once. My house is now filled with the remains of foetal dictionaries — thousands of cards representing sad attempts at dictionaries of colloquial English, British-American English, and speech pathology. Only one dictionary has ever come to term — the dictionary of linguistics and phonetics. But it is the first — and largest — of these projects which I was

particularly reminded of in preparing this paper, as in its conception, planning — and ultimate failure — it came closest to my ideal of a dictionary.

The 'Dictionary of English-speaking Peoples', as it was called, was commissioned as a pilot project by Cassell's in 1966, and was cancelled, as a result of the pilot project, in 1967. It was based on two main principles, which represent the primary dimensions of all lexicographical (as indeed all linguistic) work — linguistic variety and linguistic structure. I saw these dimensions as equal and interdependent, but commercially the variety dimension was felt to be primary, and it was this which gave the project its name. The aim was to provide in one work comprehensive coverage of all the standard regional varieties of English — Canada, Australia, South Africa, the Philippines, Ireland, and so on. The comparison of pages from existing dictionaries had shown that it was in relation to regional coverage that the gaps were greatest, both intranationally (the British English local dialect information of the *OED* contrasting with the American English dialect information of the Merriam-Webster *Third International*, for example) and internationally. The ideal of an unabridged, regionally comprehensive work seemed achievable, because in several parts of the world lexicographical projects were already on-going, and the project directors — people like Avis, Mackie and Murison — had agreed to be consultant editors. But as planning proceeded it emerged that the established projects were outnumbered by those where lexical work had hardly begun (the Philippines, Ireland and West Africa, for example) or where an enormous amount of updating was required (India and Pakistan, for example). The amount of funding which would have been required to ensure that all varieties were represented on an equally sound empirical footing turned out to be quite large. Cassell's went pale. Since then, several of the major regional surveys have been published, but the gap in even our largest dictionaries still remains wide. My ideal dictionary would fill this gap, as a priority.

The linguistic dimension to the DESP project also had its ideals. To handle pronunciation, for example, we aimed to develop an IPA-based 'meta-transcription' — a general transcription, interpretable via a specific set of rules (which would be explained in the introduction) into the different phonemic realisations for the different varieties. For semantics, we aimed to introduce information about semantic structure, using as a model an adaptation of Lyons's notion of sense relations. Information about synonymy, hyponymy, incompatibility and various kinds of oppositeness was to be systematically incorpo-

rated — not in the manner of uncontrolled ‘synonym essays’, where a series of items are brought together following an entry, and their meanings compared; but incorporated into the different senses of an entry. So, for example, different lexical relationships were introduced at various points in the entry for *babble*, as the following extract illustrates:

- 2 talk in a particular way
 - a unintelligibly, incoherently, indistinctly, cf. *jabber*; esp. like some baby or animal, cf. *gabble*, *gibber*
 - b foolishly, idly, excessively *Disp[araging]* cf. *blab* 1, *chatter*, *gab*, *gabble*, *gibber*, *gossip*, *palaver*, *patter*, *prate*, *prattle*, *tattle*, *twaddle*
 - c familiarly, cf. *chat*, *gossip*
 - d in a low tone, cf. *murmur*, *mutter*
- 3 reveal secrets through 2b, cf. *blab*, *blurt*, *prate*, *tattle*

The corresponding entries for *blab*, *chatter*, etc. were to be done in the same way: several of the relationships would overlap with those for *babble*, but in no case studied in the project was there ever identity (nor would one expect such identity from a semantic network). But of course, by disregarding redundancy, the length of the proposed dictionary increased by a factor of four. Cassell’s went paler, and, shortly after, the project was quietly cancelled.

I don’t regret the enterprise. I gained a lot of experience, bought a lot of dictionaries, made a lot of lexicographical friends, and attracted several acerbic comments from those who had more dictionary experience in their big toe than I had in my entire left hemisphere! But the exercise wasn’t a total waste of time, for a few years later the semantic structure part of the proposal was reconditioned and submitted to the Longman lexicography panel as a consultative document. In due course it led to my supervising a project on the best way of incorporating semantic structural information into a dictionary, and after several years of hard work (not by me, but by Tom McArthur and others) a product emerged in the shape of the Longman *Lexicon*. Personally, I find that work one of the most exciting developments in lexicography in recent years, but it is very limited in scope, and the model it used could fruitfully be developed and extended in several respects. In my ideal dictionary the structure of the semantic fields which are the organising principle of that work would be maximally explicit, and there would be no restriction in the head words and defining vocabulary, which reduces the richness and power of a semantic description so much.

The two themes of variety and structure (or coverage and treatment) of course subsume far more than regional and semantic information. This can be seen from any comparison of sample pages from comparable dictionaries. Looking only at coverage, and restricting the issue to head words, the discrepancy factor (that is, the number of head words not shared divided by the number of head words shared) can be as much as 30 per cent. Nor are we dealing here only with regionalisms, local slang, recent coinages and the like — where one might expect weaknesses in coverage, given the written-language-based methods of data collection still in use — but with the standard technical terminology which makes up some 80 per cent of our vocabulary. To take just two examples: I had occasion recently to look up all the names used in the classification of dinosaurs, and found only 70 per cent of them in the largest dictionaries; similarly, only about 75 per cent of the terms defined in my linguistics dictionary were included. There is a strong and growing feeling amongst lexicographers that the widely quoted estimates for English vocabulary (usually around half a million) are well out: I am convinced that the unabridged coverage of my ideal dictionary would show this to be by a factor of two or even three.

So far, the requirements of my ideal dictionary have had to do with comprehensiveness of coverage and thoroughness of treatment. There is a sense in which biggest has to be best, with dictionaries as with any reference books. On the other hand, it is obvious that, to meet the needs of individual people and circumstances, information has to be selected and presented in usable form. This is where ideal lexicographers come in. These superbeings would know all there is to be known about the needs, motives, expectations and (in Chomsky’s sense) performance limitations of the dictionary user. What is so dispiriting, of course, is how little research of a relevant kind there has been, and how little we do know. From time to time linguists, publishers and others attempt to extract some information from those ‘native readers’ who claim some knowledge of Lexicographese. This information is always valuable, but it is invariably done using a questionnaire technique, and this approach has serious limitations. For example, here is the list of points raised by Quirk, in one of the earliest studies of its kind. He asked 220 British undergraduates about the following topics:

- When the subject last used a dictionary.
- Average frequency of use.
- Concern to consult a particular dictionary.
- The dictionary normally consulted.

- Subject's ownership of a dictionary.
- Knowledge of both British and American dictionaries and basis of preference, if any.
- Ownership and use of dictionary in parental home.
- Subject's reason for most recent use of a dictionary.
- Subject's most usual reasons for use.
- Subject's failure to find what he wanted.
- Subject's suggestions for improving a dictionary.
- Should citations be from named¹ (and well established) authors?
- Comprehensibility of definitions.
- Adequacy of definitions in respect of subject's own knowledge.
- Use of a dictionary for pronunciation.
- Adequacy and comprehensibility of pronunciation symbols.
- Use of a dictionary for form-class information.
- Should dictionaries be complete, even with well known words?
- Should dictionaries have encyclopaedic entries?
- Use of a dictionary for etymology.
- Should dictionaries contain American English words?
- Should dictionaries contain slang words?
- Use of a dictionary for synonyms and antonyms.
- Adequacy of a dictionary for finding synonyms and antonyms.
- Should dictionaries contain regional dialect words?
- Should dictionaries contain phrases and idioms?
- Subject's further suggestions for improving a dictionary.

The first thing which strikes me about a list like this is its 'orthodoxy'. I am not here talking about the social-use aspect of the questionnaire, which contains no obvious theoretical pitfalls (although there must surely be problems over the accuracy of informants' intuitions about some of the questions asked. For example, could everyone here now confidently write down when they last used a dictionary, why they used it, and how often they consult one?). Rather, I am concerned about how we ask these questions, and the nature of our expectations and presuppositions when we ask them. Let us first consider the 'how'. We ask these questions using metalanguage which (questions of intelligibility aside) distances us from the informants' intuitions — we are dealing with secondary and tertiary responses, as Bloomfield put it, asking the informants to use or evaluate our labels, such as 'synonyms' and 'etymology'. What would happen if lexicographical data were presented 'straight', much as one would with acceptability testing in linguistics? This is in fact how publishers deal with advisory boards of linguists, when they want opinions about alternative layouts and styles: they provide a series of sample settings

and ask for reactions. It is an excellent way of proceeding; in the absence of any coherent lexicographical theory to which I could refer, I would not know where to begin, by way of critical comment, without the contrast in front of me.

Now, in these cases, the responses are not systematically evaluated; but there is no reason why this general procedure should not be used on a larger scale, as part of a psycholinguistic study of user evaluation. To take an example: one of the aims of using a restricted defining vocabulary, as in *LDOCE*, is to allow a higher level of lexical comprehension in the definitions; but a consequence of using the 'easier' vocabulary is that the length of the definitions often has to increase. Meaning which in an unrestricted dictionary would be 'packed' into a lexical item is now unpacked through an extended grammatical construction — often a relative clause. For example, the *LDEL* definition of *insurrection* is '(an act or instance of) revolt against civil authority or an established government', whereas the *LDOCE* definition is 'the act or occasion of rising against the people who have power, such as the government' — the relevant contrast being between 'civil authority' (where the constituent items are not part of the defining vocabulary) and 'the people who have power'. Now, the interesting research question is whether a user finds the packed or the unpacked versions easier to handle (where ease is operationalised in some standard way, as in psycholinguistic research paradigms)? The notion 'convenience of the user' is often cited, but rarely if ever tested. Systematic information on this point would be quite absorbing.

Any area of lexicographical practice could be investigated in the same way. Under which constituent item is it best to place an idiom (*kick* or *bucket*, etc)? Is a definition with a picture of more or less help than a definition without a picture? The answer is to be obtained not by asking the informant direct (along the lines of 'Which do you prefer?') but by giving him a task which requires that he use the information in the definition in some way, and then seeing whether his response is facilitated in the picture-present or the picture-absent conditions. The same procedure could be used for any aspect of lexicographical presentation or design — alternative pronunciation transcriptions, for example. Has anyone ever asked informants to read aloud the pronunciations of words represented by alternative transcriptions, to see whether speed, accuracy and other factors are influenced? Or checked on the interpretability of abbreviated labels (in grammar, etymology or whatever) in this way? Or tested the clarity of sets of usage labels (such as formal, informal, slang, etc)? To take this last

point, it is sometimes said that sociolinguistic research will help point the way towards a more consistent use of stylistic labels in dictionaries; but I personally doubt it, without the kind of experimental underpinning which only a psycholinguistic investigation can provide. In short, my ideal lexicographer is someone who can supplement his descriptive, naturalistic leanings with a rigorous experimental method. His traditional training in descriptive lexicography should be supplemented by an ability to practise what might now be called 'experimental lexicography'.

The second concern I have about the 'orthodox' appearance of lexicographical questionnaires relates to our preconceived ideas about what dictionary users like and know about. Because we know what 'should' be in a dictionary, as good linguists and lexicographers, we ask questions relating only to these notions — questions to do with lexical relationships, form class, etymology, and so on. But an ideal lexicographer should always be striving to go beyond this — to discover whether there are other parameters of relevance to the user. Left to themselves, what do people want to have in their dictionary? According to the Reader's Digest preliminary enquiry which led to the *Great Illustrated Dictionary*, a large number of people expect their dictionary to contain specific facts — dates, people, and other basic encyclopaedic information. In Quirk's survey only 44 per cent of his student informants wanted this kind of information, but the figure is much larger when the 'person in the street' is asked. Also, a large number of people apparently appreciate the value of a dictionary in colour: now *there's* something which no linguistically inspired lexicographical theory could have predicted. When I asked some people about this, following the publication of the *GID*, one said, 'It's a pleasure to look something up.' One of my own children likes to use it because it has lovely pictures. So, if the truth be known, do I. Or again, we can consider the reasons people look something up: in the Quirk survey, etymology, pronunciation and usage loomed small; by contrast, 57 per cent of the uses to which a dictionary was put in a parental home were in relation to word games (Scrabble, crosswords, etc) — a figure which can err only in being too low. Quirk makes the wry comment that 'some of the dictionary features which seem of particular centrality to lexicographers are decidedly peripheral to the ordinary user'.

So who should bend? The lexicographer, or the user? Should lexicographers try to change their methods in order to meet the demands of the potential consumer? Or should they attempt to educate

the consumer into new ways of behaving — at the very least, a national campaign to persuade dictionary users to read their prefaces? There are merits in both approaches, with publishers on the whole leaning towards the first, and academics on the whole leaning towards the second. But both are required in order to create my ideal users in their ideal lexicographical world. Such users have been taught to understand dictionary conventions as a routine part of early education, starting in junior school, where they were given a nice-looking dictionary, written at their level, and not (as is more usual) a book which looks boring, is falling apart, and which is written in a language well above what would be expected of them in other aspects of the curriculum. During school they have taken part in several of the national dictionary-using competitions which have been sponsored by national academic bodies and publishers interested in lexicography. As adults they have continued to go in for such competitions, and have probably subscribed to *English Today* (or the equivalent magazine in their own language). As a consequence of all this, they have bought two dictionaries, or preferably one dictionary at two levels, organised along the lines of the *Encyclopaedia Britannica's* micropaedia and macropaedia — or, to adopt a more homely analogy, *A University Grammar of English* and *A Grammar of Contemporary English*. They have signed on to receive the annual lexical supplements. They use the data collection forms which are issued routinely by the publisher, on which they note and send in (for a small fee) any changes in usage which they have observed. They know their transcription symbols, because they have listened to the recording the publisher has thoughtfully provided on a small disc inside the back cover. Playing with dictionaries for them is a leisure activity, made especially appealing because the data base (to move forward in time a little) is now available in electronic form, which their terminal allows them to access, and to which they can plug in one of several lexicographical computer games. If they wish to look something up, they have the option of referring to their lexicopaedias, or addressing the data base direct through their voice-activated terminal. They know their access code words. 'Dictionary', if they have a word in mind which they want information about; 'Thesaurus' if they have a meaning in mind which they want words for. Then 'American', 'British', or whatever, to access the appropriate sub-file in the data base. Then 'Meaning', 'Pronunciation', 'Usage', 'History', 'Picture', 'Spelling', 'Idioms', or whatever, as required — the information to be made available in sound, on screen or in print, depending on which mode selection they make.

But enough of these fantasies, which are more appropriate after cocktails than before. In indulging in them I have not forgotten the realities which face us all in attempting to make sense of this difficult domain. My proposal for an experimental dimension to lexicography, within the tradition of applied psycholinguistic studies of performance, is by no means a fantasy, though it is an earnest hope. It would help to provide a foundation of empirical fact about lexicographic practice which would supplement the strong descriptive traditions of the subject. In this way one might even, in due course, move towards a characterisation of what counts as a well formed dictionary entry. Perhaps Chomsky's definition of competence was not so irrelevant after all.

ALLUSIONS AND REFERENCES

- Bloomfield, L. (1944), 'Secondary and tertiary responses to language', *Language*, 20, 45-55.
- Chomsky, N. (1965), *Aspects of the Theory of Syntax*, p. 3. Cambridge, Mass.: M. I. T. Press.
- English Today* (1985), quarterly magazine, Cambridge: Cambridge University Press.
- GID* (1984), Reader's Digest *Great Illustrated Dictionary*. London: Reader's Digest.
- LDEL* (1984), *Longman Dictionary of the English Language*. London: Longman.
- LDOCE* (1978), *Longman Dictionary of Contemporary English*. London: Longman.
- Lyons, J. (1977), *Semantics*, vol. 1. Cambridge: Cambridge University Press.
- Quirk, R. (1973), 'The social impact of dictionaries in the U. K.', in R. I. McDavid and A. R. Duckert (eds.), *Lexicography in English* (Annals of the New York Academy of Sciences, 211), 76-88.
- Quirk, R., and Greenbaum, S. (1973), *A University Grammar of English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1972), *A Grammar of Contemporary English*. London: Longman.
- Webster's Third New International Dictionary* (1961). Springfield: Merriam.

DISCUSSION SUMMARY

- (1) In addition to questionnaires and task-based experimental techniques, there are other ways of investigating dictionary use. These include protocols based on users' own accounts of their procedures, and recorded observations of what they actually do. Furthermore, users of computerised dictionaries can have their procedures logged on the computer itself.
- (2) New trends in British educational psychology (perhaps influenced by the 1975 Bullock report, *A Language for Life*) seem to encourage children to appreciate language in both structural and functional terms, and to become acquainted with dictionary conventions at an earlier age. At the same time, studies of the

explanatory techniques used by children suggest a five-stage development (as for *broom*): pointing to a broom, use of deictic forms ('That's a broom'), person-oriented explanation ('My mummy's got a broom'), functional explanation ('A broom is for sweeping'), and (not usually until eight to ten) consideration of form (the physical appearance of a broom) leading to the classical Aristotelian formulas ('A broom is a certain kind of object with a certain function'). This may have implications for explanatory techniques in dictionaries.

- (3) Users' expectations of dictionaries may vary from country to country and from period to period. The encyclopaedic dictionary is well known in France, Germany and America, and in Britain (especially Scotland) was once more widespread than it has been until recently. Johnson's dictionary, and *OED* later, provided an alternative non-encyclopaedic model of the dictionary which proved particularly influential in Britain. Some market research suggests that users welcome encyclopaedic features even here.
- Dictionary features developed in one country might profitably be adapted elsewhere; thus Professor Crystal's model entry for *babble* owed something to the *dictionnaire analogique* features pioneered by Robert in France.
- (4) The disciplines of terminology (dealing with concepts and their names) and information science (dealing with the structural organisation of knowledge) may provide a bridge between dictionaries (dealing with 'words') and encyclopaedias (dealing with 'things'). Workers in the field of artificial intelligence feel increasingly a need to draw on the lexical and indeed encyclopaedic information to be found in real dictionaries.
- (5) The inclusion of a pronunciation recording in the back of an American dictionary was a commercial failure some years ago; on the other hand, there is now great demand in the Far East for hand-held computer disks with pronunciation.