Moving towards perfection: the learners’ (electronic) dictionary of the future

Leech, G. and Nesi, H.
Published version deposited in CURVE October 2015

Original citation & hyperlink:
http://dx.doi.org/10.1515/9783110947021.295

Note: This book was originally published by Max Niemeyer Verlag in 1999, and has been reprinted by De Gruyter in 2011. This book is no. 95 in the series Lexicographica. Series Maior.

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

CURVE is the Institutional Repository for Coventry University
http://curve.coventry.ac.uk/open
Geoffrey Leech and Hilary Nesi

Moving towards perfection:
The learners' (electronic) dictionary of the future

1 Introduction

The initial paper in this book, by Tony Cowie, has provided a historical account of the precursors of the present-day English learners' dictionary. It is appropriate, in this final contribution, to speculate on the learners' dictionaries of the future. Other preceding papers have made it clear that, in spite of the major advances in lexicographic practice that learners' dictionaries have made over the past two decades, such dictionaries fall well short of perfection. This is unavoidable, as far as the present generation of learners' dictionaries is concerned. Practical considerations determine that publishers have to produce a dictionary of a certain price, and of a certain size, which precludes the perfection of providing all the information which learners might ideally need. At the same time, economic opportunity has justified the considerable and continuing investment which publishers have put into these dictionaries, leading to the remarkable improvements we can already admire, particularly in tailoring the content and presentation of dictionaries closely to the needs of its users.

Among the sources of improvement, the development of electronic tools and resources, particularly corpora, has made the most groundbreaking contribution. In this respect, the learners' dictionary has set new standards: its example of responsiveness to the user's needs and accountability to linguistic evidence is increasingly being followed by older and more traditional branches of lexicography. Further, modern computational methods of information storage, text processing and book production mean that the existence of a learners' dictionary as a book of printed paper is almost secondary to its existence as (part of) an electronic databank. In fact, a number of different paper dictionaries - shaped to the purposes of this or that particular user-group - can be offshoots of the same lexicographical information bank. The developing market for electronic dictionaries now means that the paper dictionary is a secondary product. It is no longer essential even at the user's end: it can be replaced by a CD-ROM (such as those recently produced by Longman, COBUILD and Oxford University Press) or even by a tiny IC card, perhaps as an extension of an electronic notebook or personal organizer.

It is not surprising, then, if our vision of the future focuses on the fast-evolving opportunities of the new age of electronic communication and information technology. It would be unwise here to speculate about the demise of the printed book, an adaptable survivor from the Gutenberg technological breakthrough of the fifteenth century. The paper dictionary no doubt has a long and wordweight future ahead of it - although it is unclear how far forces of inertia will serve to prolong its vitality in the next millennium. What does
seem certain is that it will face growing competition from the electronic dictionary, which will provide increasing advantages as the price of hardware continues to drop and the power and convenience of the computer continues to increase. Hence in envisaging new horizons for learner lexicography in the next fifty years, we predict the electronic dictionary to be a better bet than the familiar 'hard-copy' dictionary, which we all know and love, but which appears to offer limited possibilities of further technological progress.¹

As always, in dealing with the pioneering and experimental stages of new technologies, there are strong negative points balancing the positive ones in favour of innovation. Like the motor cars of the early 1900s, electronic dictionaries, with the backup technology they require, are on the expensive side, and have their teething problems. It will be a long time, for example, before electronic dictionaries are available in the majority of primary schools in the world. Likewise, although search and retrieval software is getting cleverer, it does not do all of the things we would like it to do. But to judge the future potential by present achievement is to make the cardinal error of traditionalists through the centuries. So let us, at this point, summarise the main advantages of the electronic dictionary, not limiting our attention to what has been done so far, or worrying too much about some short-term limitations of current products. In passing, we will return to a number of points mentioned in previous contributions to this book.

2 Advantages in information access

The types of advantage noted in this section are those which apply to many areas of the electronic information revolution, not only dictionaries.

2.1 Storage capacity

In considering the perfect learners' dictionary in relation to current learners' dictionaries, the image comes to mind of a giant octopus being confined inside a wooden box. The perfect learners' dictionary, one which answered all users' needs, would have to be a vast open-ended information source. It would need to extend its tentacles in many directions from the central core of a dictionary to other types of lexical information which might be required by this or that learner: e.g. lexicogrammatical, multilingual, visual, encyclopaedic, etymological. But the paper dictionary confines this open-ended requirement to a fairly strict limit. What Bolinger wrote in reviewing the OALD4 is just as apt today: "I suspect that hard-copy vademecum dictionaries of this type have about reached their capacity. Any really dramatic advance would burst their covers." (1990:144) Many of the comparative

¹ On the range of electronic learners' dictionaries already available, and a review of their strengths and weaknesses, see Nesi 1996 and 1999.
arguments about whether this or that dictionary is preferable ultimately come down to issues of priority: how much room do we have for this kind of information rather than that? On the other hand, electronic storage media have massive capacity compared with storage on paper. A single CD-ROM can contain the most gigantic dictionary of English ever written: the twenty-volume Oxford English Dictionary. Currently, paper learners' dictionaries, although weighing over ten times as much as a CD-ROM, have less than a twentieth part of the capacity of the OED. In consequence, by transferring a learners' dictionary to CD-ROM, one could cram in at least 200 times as much material. This can allow the expansion of an existing learners' dictionary to a much larger size, or the addition of further reference information sources (such as are currently provided by the Longman Interactive American Dictionary and the Collins COBUILD on CD-ROM): for example, a visual library, a grammar, a usage dictionary, a dictionary of common errors, a large corpus of texts.

2.2 Multimedia

Apart from text and illustrations, the electronic dictionary is already beginning to use channels of information not available to the book dictionary: sound, flexible vision (including moving images), as well as text. OALD on CD-ROM, The Longman Interactive English Dictionary (LIED) and the Longman Interactive American Dictionary (LIAD) for example, all provide audio recordings, and LIED and LIAD also include a selection of video clips.

Contributors to this volume have at various points noted the limitations of making meanings clear on the printed page. Learners' dictionary lexicographers have done their best, using definitions with a limited defining vocabulary, usage boxes and illustrations showing special visualizable fields of vocabulary. But many expressions (the example of make the bed was used earlier) are difficult to explain in simple language, and the restrictions of the defining vocabulary may result in some verbose and clumsy definitions. However, the availability of animation can change all that: in principle, the depiction of action in real time will enable any physical action, process, or change of state to be directly portrayed to the learner. This potential has not yet been fully exploited, but some of the most recent electronic dictionaries such as LIAD are now beginning to use animation in a limited way to illustrate verb meanings. Similarly, the great advantage of the direct auditory representation of a word over its phonetic transcription scarcely needs to be laboured. Hearing any vocabulary item pronounced "at the click of a mouse" is a far better alternative to phonetic transcription, a bugbear to many learners. LIED, LIAD and OALD on CD-ROM provide audio recordings of every dictionary headword, and LIAD also offers a record and playback facility, enabling learners to compare their own pronunciation with the dictionary recording. With a "sound picture" of each word or phrase in a dictionary, phonetic transcriptions are hardly necessary. However, rather than banishing phonetic transcriptions from the electronic dictionary, one would like to keep them as an optional
extra - as a kind of information which is made available only when it is needed and for the benefit of those who need it - see 3.3 below.

2.3 Easier and faster access

It is far easier to type a few characters on a keyboard - sufficient to find the headword you are interested in - than to ruffle through more than a thousand pages in order to home in on that same word. Access to an entry in an electronic dictionary is (normally) relatively instantaneous, whereas looking up in a printed dictionary is an inefficient, hit-or-miss, often frustrating process. The printed book is essentially a linear storage device, whereas conceptually, a lexical information resource is multi-dimensional. The alphabet, which has tyrannised over dictionary arrangement and access for centuries, ceases to be a tyrant when your dictionary is electronic.

Ironically, it has been argued (see Sharpe 1995: 50) that the very speed and convenience of access is a drawback of electronic learners' dictionaries, because information so easily extracted may be just as easily forgotten by the user! On the other hand, research by Guillot and Kenning (1994) has found that students using an electronic dictionary show "an increased capacity for sustained effort," with much exploratory browsing of the dictionary and more multiple searches than would otherwise take place.

We return to accessing issues - specifically to variable paths of access - in 3.3 below.

3 Advantages in terms of needs and resources

Turning now to advantages of the electronic learners' dictionary which are related more to educational value, first, let us consider a simple model of dictionary use in terms of needs and resources. A learner (here we are thinking of an EFL learner who is not a native speaker) uses a dictionary especially when trying to solve a communicative problem - one either of comprehension or of production. This is the immediate need for the use of the dictionary, but there is also a more long-term need or objective: the dictionary is a tool to improve the learner's competence in the target language. This is a goal that may be gradually attained over many years of learning.

A dictionary is a resource that may be consulted or accessed, in order to help satisfy the needs mentioned above. Those who compile the dictionary, however, may rely on more fundamental resources - such as a corpus of written texts and spoken transcriptions, or a citation bank. These are what may be called primary resources - direct evidence of the use of the language. Compared with them, the dictionary itself is a secondary resource: a compilation of lexical information about the language, systematically organised so as to be readily available to the user. The general relation between the dictionary and the dictionary
Moving towards perfection

user is represented in Figure 1, which shows means of access as the channel linking the user to the dictionary.

In the case of a printed dictionary, the three components of Figure 1 are relatively fixed, whereas in the case of an electronic dictionary, possibilities of variability are potentially very large. It is this multi-access potential, with its adaptability to need, that is a major advantage of the electronic dictionary.

3.1 Variability of access points

Let us consider first the learner, the dictionary user. The English learners' dictionaries chiefly considered in this volume (OALDS, LDOCE3, COBUILD2, and CIDE) are all roughly comparable in the level of learner that they cater for. Their lower bound of user competence is determined by the fact that these are monolingual learners' dictionaries: all the explanatory materials, as well as all the words and phraseologies to be explained, are in English, the target language. Their upper bound, although fuzzier, is imposed by the limited list of headwords they contain (entries for rare, archaic, or highly technical words being generally omitted), and perhaps also by the use of a limited defining vocabulary. In spite of the user-friendly aspects of these dictionaries, which make them rather enjoyable to use, a user with near-native competence would probably want to "upgrade" to more comprehensive or specialised native-speaker dictionaries. The user population, for these learners' dictionaries, is therefore confined to a broad band of intermediate-to-advanced learners, as shown in Figure 2.
Learners excluded from this band include both those who are "too elementary" and those which are "too advanced". Extending the band of users downwards, to include elementary learners, would necessitate invading the territory of the bilingual dictionary, and giving access to the information in the dictionary via the learner's native language (or a second language the learner knows). This would not only be an open sesame for less advanced learners, but would also overcome the problem, for dictionaries employing a limited defining vocabulary, of unhelpful over-long definitions for something that could be more easily and economically explained in the learner's own language (see 2.2 above).

There is no obvious reason why future electronic learners' dictionaries should not provide such variable access points. For example, for speakers of French, a core list of headwords in French (with main sense and phraseological subdivisions within each headword) would provide a more accessible "front end" to the English dictionary vocabulary for productive use. For receptive use, the English defining vocabulary could be cross-referenced to equivalents for the major world languages. Of course, these NL-specific add-ons would take up space, but we have already noted that this would hardly be a problem for electronic dictionaries of the future.

Learners "too advanced" or specialized for current learners' dictionaries would have dictionary needs similar to those of native speakers. They would want to be able to extend their vocabulary into more peripheral areas of the lexicon: for example, into technical terminologies, archaic terms, slang expressions, dialectal forms and so on. Consider just one area where it has been noted that current learners' dictionaries leave much to be desired: the lexical characteristics of different national varieties of English. At present (see Heath this volume) learners' dictionaries make only limited efforts to cover national varieties apart from American and British English. Regrettably, there can be little room to spare for detail on Australian, New Zealand, Caribbean, Singaporean and other regional

---

**Diagram**

- Expanding competence
- Zone where monolingual learners' dictionaries are currently most useful
- Learner's progress
variants. But again, if we "think electronic," there is little danger of dictionaries running out of space. Various national vocabularies can be add-on resources, supplementing the lexical databank with information optionally available to the user.

Another example of variable access points is the provision of a "sound like" facility, whereby the user can access a word via a word similar in pronunciation. LIED, COBUILD on CD-ROM and even some of the smallest hand-held electronic dictionaries provide this facility. Given the vagaries of English spelling, this is bound to be something of a hit-or-miss affair. However, developments in automatic speech recognition open up the possibility of the purely auditory look-up of a word. If this seems futuristic, there is already the device (for example, in LIED, LIAD and OALD on CD-ROM) of providing a word's pronunciation automatically on look-up, as an immediate auditory feedback to the user. Other approximate look-up procedures, such as the partial spelling of a word using wildcard characters, are also helpful. Such procedures can be an escape from the limitations of traditional orthographic look-up, where paradoxically a student needs to know how to spell a word in order to look it up (and find out how to spell it!).

3.2 Variability of destinations

The preceding paragraph has already introduced the idea of variability of destinations: different users, with different needs, will need to track down information from different lexical resources. Already, electronic dictionary producers are thinking pluralistically, combining a number of different resources in the same CD-ROM package: a dictionary accompanied by a grammar, a usage dictionary, an error dictionary, a corpus and so on. At present, however, there is a temptation to use the space on the CD-ROM rather like shelf space for books in a library: simply combining coexisting reference resources, without integrating them. The facility to access more than one resource from the same user interface is only a first step towards integration. There is much more to the integration of lexical resources than this. There is need for "multi-referencing": for simultaneous signalling to the user that the same query item is to be found in a number of different resources. At present, many of the teething troubles of electronic dictionaries, promising as they are, are to be found in the imperfect integration of a number of resources. For example, there is clearly a pedagogical difficulty of integration between a dictionary and a corpus, where the corpus (consisting of authentic, unpreselected text or discourse data) inevitably contains many items which are not represented in the dictionary, and vice versa. No doubt it would be better to provide, not a "raw corpus," but an extensive citation bank, in the form of a concordance in which each item in the dictionary is illustrated by a set of examples in context. There is also an absurdity, for example, in providing a single route of

---

2 But there are regionally specific dictionaries produced in some of these English-speaking countries. An example, for Singaporean usage, is the Times-Chambers Essential English Dictionary, Singapore and Edinburgh: Chambers Harrop & Federal Publications, 1997.

3 See Nesl 1996 on the difficulties of cross-referencing.
access to both a learners' dictionary and a pronouncing dictionary, where the latter has many more headwords than the former, but supplies only pronunciation. Such problems of integration can lead to much frustrated searching for information which turns out to be unavailable. But they can be overcome if the multiple-database provides signals to the user of the available sources of information for a given item at the point where the query is made. There should also be an easy way of "switching off" a given information source (say, a regional dictionary of British English, or a pronouncing dictionary) where these are not needed by a particular person or for a particular application.

Existing printed dictionaries suffer from problems of information overload: too many different types of information in the same dictionary can clutter the page and distract the reader from the information quest in hand. Especially grammar codes and phonetically transcribed pronunciations can be disproportionately distracting to users for whom these more technical features have no relevance and are difficult to handle. There is a big difference, here, between linguistically sophisticated users (typically teachers?) and linguistically naive users (typically students?) who may nevertheless have a good practical knowledge of the target language. For the former, grammar codes and phonetic notation may be meat and drink; for the latter, meaning, spelling and phraseology tend to be the major concerns, and grammar and phonetics can be baffling and offputting. Of course, one can argue that students ought to be interested and enthused by grammar and phonetics, and so should be continually exposed to these types of information. But this argument loses its force if the result is to put the students off using their dictionary entirely.

One way to hide information of no importance to the reader's current concern is, as already said, to switch off or mask the information types not needed. Another is to use a hypertext structure for the database, so that a user who wants to explore more detail can optionally explore links (say) to a concordance of corpus examples, or to a list of collocations associated with a given item. Hypertext referencing also helps to solve another problem of information overload. One of the difficulties of conventionally printed dictionaries is the great size and complexity of entries for some very common words, such as in or get. But for these, a hypertext information structure can provide not merely cross-referencing, but layering of information at different levels of detail, so that a top-level panoramic single-screen view of an entry (like the summary entries in the Longman Language Activator) can supply the reader with links to further details in terms of subsenses, multi-word expressions, and so on. It is true that current printed learners' dictionaries have become sophisticated in the simultaneous presentation of many kinds of information. But the various devices learners' dictionaries use to reduce information overload include not only features such as typography and layout, but also simplifications which verge on oversimplification. Simplifications which reduce the amount of information available (and most simplifications do) are unnecessary if an electronic dictionary makes use of hypertext methods of managing the presentation of information and other methods available with a modern windows-based user interface.

---

4 See Nesi 1996 on the LiED problem in this respect.
3.3 Variability of search or access paths

Again, we have anticipated this type of variability in the above paragraphs. The multi-access feature of electronic dictionaries not only includes multi-referencing to different resources, but also extends to cross-referencing (e.g. via hypertext links) within and between resources. This means, for example, that on looking up the word brave in the dictionary, we can immediately click on one of a list of synonyms (courageous, valiant, bold etc.), and be able to compare the two synonym entries in separate windows on screen.

Or it would be feasible to have a reception-oriented learners' dictionary cross-referring to a production-oriented dictionary like the Longman Language Activator, which provides detailed information on the choice between synonyms or quasi-synonymous words. The production-oriented dictionary, on the other hand, because it tends to cover a smaller vocabulary in greater detail, would benefit from cross-references to less common synonymous expressions in a reception-oriented dictionary, where meanings and examples for these expressions could be sought. In fact, in the longer term one could envisage a merging of the dictionary (essentially reception-oriented) and the thesaurus (a production-oriented lexical reference resource), providing access links via realisation (spelling or pronunciation) or by similarity of meaning as required. The difference between a dictionary and a thesaurus is basically a question of whether form or meaning is the access route to the lexical information which both types of publication ideally provide.

Possibilities of cross-referencing are numerous. At present, the dictionary shelf of the library contains many lexicographical tomes with overlapping information, but also with different areas of specialization. The need is to combine these so that cross-reference between the different special sources of information is easy. For example, a dictionary of faux amis between native language and target language is useful for learners, as is also a dictionary of common errors. Such dictionaries already exist as separate paper volumes. But to make the most of the information they contain requires cross-referencing, so that clicking on the flagged word in one dictionary will automatically bring up the relevant information in the other. Pushing this way of thinking a little further, we reach a stage where, for the user, different dictionaries do not exist as separate entities; rather, there is a single lexical database with links between the different categories of information associated with the same headword entries.

A final point about variable routes of access is that there should be links between primary and secondary resources. An obvious implementation, in a multi-access electronic dictionary, is a link from the dictionary item to corpus material - providing, for example, frequency data (including collocations) or KWIC concordance listings. The opposite link - going from corpus word to dictionary entry - could also be a possible and valuable type of connection, enabling the learner to use the bank of texts stored with the dictionary as a set of materials for reading comprehension, for example.
4 The interactive dictionary

Having elaborated on the advantages of variability in the electronic dictionary, we move on to a final advantage, closely associated with variability, viz the truly interactive potential of the electronic dictionary.

The printed book is not interactive: it is an inert resource, passively waiting to be consulted by the reader. On the other hand, an electronic dictionary offers a wide range of choices to its user - in terms of the points of departure, destinations, and search routes. Perhaps there is too much choice for some users; but for those who want to take a more active stance, there is a great opportunity for developing dictionary-using skills. The electronic dictionary can provide a stimulating environment for exploring the lexical resources of the language using their own initiative, and developing their own word power. Exercises and tasks comparable to those provided by CALL [computer-aided language learning] courseware can be integrated with the dictionary, and with the concordance materials which may accompany it. The Electronic Wordpower Dictionary and OALD on CD-ROM already include vocabulary games such as crosswords; more elaborate electronic dictionaries of the future might provide more focussed feedback on vocabulary learning activities, with suggestions for further exploration. In other words, the dictionary is adaptable for the needs of the learner, and to the learner's changing behaviour.

In another sense of "interactive", the electronic dictionary can interact with other computer tools and applications. These can vary from CALL materials and concordancing software (two already mentioned) to printers and word processors. From a word processor it is already possible, in existing packages, to access a dictionary or a thesaurus on-line. In this way, the dictionary becomes a writing aid comparable to a spelling or grammar checker, only more educationally useful. Similarly, it is already possible to link up a dictionary and a printer, allowing selective print-outs. Where students do not have individual access to computers, these provide, for example, multiple copies for classroom activities and study tasks.

5 Conclusion

As much of this paper has painted a rosy picture of future technological progress, it is time to end on a more sober note of re-appraisal. We are now in the period of first generation electronic learners' dictionaries, which are heavily influenced by the printed dictionaries from which they originate. It will not be surprising if many who use them have reservations about their superiority to printed dictionaries. In part, this is because the potential of the electronic dictionary is only beginning to be exploited. In part, it is also due to technological conservatism, especially among the older generations of dictionary
Moving towards perfection

users, which include many teachers. In part, too, it may be because we appreciate the familiar advantages of the printed dictionary and do not entirely appreciate how an electronic dictionary could ever match them.

Paper dictionaries are relatively cheap and accessible. We do not have to acquire special equipment or know-how in order to use them. We can casually dip into them, or scan them in a "browsing" mode, without using electronic implements such as a mouse. A paper dictionary is felt to be more ornamental, in one's living room, than a PC or a CD-ROM. It can be carried around the world (e.g. by air) without fear of serious damage or loss of functionality. Moreover, oddly enough, although a printed dictionary is an example of yesterday's technology, we can feel relatively confident that it will not become obsolete so quickly as the latest up-to-the-minute hardware and software. Finally, at least for some people, books can have some of the fetish value of a cuddly toy: we can treat them lovingly and take them to bed with us. It will be a long time before a computer with a CD-ROM drive will compete on all these fronts.

However, it is worth thinking here about hardware in the practical delivery of lexical information. Our assumption, up to now, has been that the computer plus CD-ROM will be the main mode of delivery. But we cannot look decades into the future, any more than a computer pioneer in the 1950s could have foreseen the age of terminals, keyboards, personal computers, diskettes, CD-ROMs, and all the other things that have become familiar to the average child today. Bolinger (1990: 145) was probably right, for example, to say that "the dictionary-consultor of the future will tap out inquiries on a hand-held computer." Already there is a growing market for electronic dictionaries as hand-held devices. Although these tend to be neglected by the educational and academic sectors as being lexicographically beneath serious notice, they are highly popular in certain student markets, and in the near future could well improve substantially in coverage, quality and price. If so, the problem of expense and portability associated with the first generation of learners' electronic dictionaries could well be on the way to solution. Another likely mode of delivery in the future is on-line, via the Internet, where the problem of information capacity can be ignored. It would not be too difficult, in view of current technology, to predict the advent of hand-held electronic dictionaries which can also interact by remote communication with large data resources on the World Wide Web. We can only speculate where technology will take the electronic dictionary in the future, but it seems unlikely that most of their current limitations, which may loom large in people's experience of electronic dictionaries today, will persist in the future.

There remains, therefore, a large gap between the promising (but limited) pioneering developments of the present, and the likely achievements of the future. Further, this gap cannot be bridged merely by technological advances. It is clear that first-generation electronic learners' dictionaries have some shortcomings which cannot be remedied without considerable expenditure of effort and money. Much of this progress - e.g. towards the optimal cross-referencing of resources - may turn out to be labour-intensive. The human and technical resources which have been expended on printed learners' dictionaries in the past twenty or thirty years will need to be applied, with equal dedication, to the development of the electronic dictionaries of the future. If that is done, it
is likely that the overwhelming advantages in principle of electronic learners' dictionaries will become overwhelming advantages in practice. Although the perfect learners' dictionary will always remain imaginary, moving towards perfection will have become a reality.

Bibliography
