Dictionaries on Computer: How Different Markets Have Created Different Products

Hilary Nesi

CENTRE FOR ENGLISH LANGUAGE TEACHER EDUCATION, UNIVERSITY OF WARWICK, CV4 7AL, UNITED KINGDOM

I. INTRODUCTION

Good paper-based dictionaries are too fat. That is the reason why students leave them at home, why teachers do not carry sets from classroom to classroom; and why all but the most enthusiastic users deal with only one dictionary at a time, rather than pooling the resources of several different volumes.

The fact that good paper-based dictionaries for advanced learners of English have gained weight is largely a consequence of changing attitudes to language teaching and learning. The earliest editions of the Oxford Advanced Learner's Dictionary broke away from the native speaker dictionary tradition by providing detailed information about the valency patterns of verbs. Since then there has been an increasing amount of grammatical information in all the learners' dictionaries, but the trend towards communicative language teaching, which stresses the importance of language appropriacy in addition to accuracy, has also led dictionary makers to add more and more information about word behaviour in context. Consequently all the major advanced learners' dictionaries now contain usage notes, comments on pragmatic function, warnings of register restrictions, and examples taken from corpora of authentic texts.

A further reason for the increased size of learners' dictionaries is their increased coverage. Publishers boast that each new edition contains a greater number of definitions, references, or word meanings (the distinction between these is important, but is often deliberately left unclear). For example, OALD has increased its coverage from 50,000 headwords and derivatives in the 1974 edition, to 57,100 words and phrases ("over 4000 NEW to this edition") in 1989, and to 65,000 definitions in OALD5 (1995). Likewise the COBUILD coverage has grown from 70,000 references in 1987 to 75,000 references in 1995, and LDOCE has shot from 56,000 words and phrases (1987) to 80,000 (1995). Expansion has doubtless been fuelled by competition between Oxford University Press, Longman and Collins, but also reflects the exponential growth in English terminology world-wide. According to Oxford University Press publicity material, the sixty readers employed to contribute to the "Oxford World Reading Programme" report on average 18,000 new words and phrases to the press each month. Obviously very few of these get included in any hard-copy dictionary, but reviewers and users often judge a new dictionary by its coverage of new words, and some new words are essential for the modern student, for example Information Technology terms relating to library use and word-processing.

No wonder that the learners' dictionaries are about to burst their covers. Yet despite the huge increase in content, reviewers, teachers and learners are still not satisfied with the amount of information they receive. Bolinger comments on the inconvenience of the fat learner's dictionary in his review of OALD4: "I suspect that hard-copy vademecum dictionaries of this type have about reached their capacity." (Bolinger 1990: 144)

Yet in the same review he calls for OALD to provide more illustrations, more examples of regional use, more idioms and
Kennedy (1992) suggests that more statistical information needs to be provided too. The corpora on which current learners’ dictionaries are based can provide an almost limitless quantity of data concerning word and structure frequency, likely learner error, and differences in language use as affected by age, region, gender, genre and mode of delivery, and Kennedy thinks that the information that informs the lexicographers should be made available to teachers too:

> it is not enough to tell teachers that curricula, reference works or teaching materials are based on corpus analysis. Increasingly, the most professional teachers expect evidence to justify positions taken, and teacher trainees should receive statistical information as part of the description of English or whatever language they are learning to teach.

(Kennedy 1992: 367)

Kennedy’s request for corpus evidence has been met in part by the latest edition of LDOCE, which places common words in spoken and written frequency bands, and provides charts to illustrate the relative frequencies of certain expressions and structures. COBUILD2 also indicates general word frequency. Much more such information could be included, however, if space permitted. We have reached the stage when the amount of information that lexicographers wish to convey, and that teachers and learners wish to acquire, far outweighs the capacity of any single-volume book.

Increased information content not only makes dictionaries fatter, but also poses organisational problems. Paper-based dictionaries organise information in a primarily linear way which is appropriate for the listing of a succession of separate entries, but inadequate as a means of grouping and regrouping words according to their semantic and pragmatic similarities, or their valency and collocational patterning. The A-Z sequence places headwords in an order that is virtually meaningless, shedding no light on the relationships between words that are alphabetically distant, and complicating searches for phrases and idioms. McArthur (1986) champions the thematic organisation of word books, on the grounds that: “Any reasonably well-constructed conceptual framework is far closer to ‘reality’ and how our minds work than anything that is alphabetically ordered.” (McArthur 1986: 151)

Yet although thematic organisation in a paper-based dictionary may improve look-up quality, it does not make searching any quicker or easier. As McArthur points out, “People can handle alphabetisation much more easily than thematisation, because they are used to it” (1986:153). Moreover, in a non-alphabetical thesaurus or lexicon two separate look-up processes are often entailed: the first to identify the semantic group(s) to which the search word might belong, and the second to find the search word within a given category. What learners really need is a fast and flexible access system which allows a variety of search routes, adjustable according to the user’s existing word knowledge, look-up preferences, and the information s/he specifically seeks.

II. THREE TYPES OF ELECTRONIC DICTIONARY

We conclude, then, that paper-based learners’ dictionaries can neither hold all the information they need to provide, nor store existing information in a sufficiently accessible way. Computer-based dictionaries can do both. An entire fat dictionary can be stored on a credit-card-sized IC card with space to spare, and as with any other electronic database the various pieces of information that constitute a dictionary entry can be broken up and stored separately in electronic form, enabling users to call up entries in terms of such categories as word class, grammar, meaning and/or examples, rather than simply according to the position of the headword within an alphabetical sequence.

The principles of electronic storage are fundamentally the same whether just one dictionary is held in a pocket-sized device, combinations of reference works are stored on computer hard disk or CD-ROM, or hundreds of dictionaries are available for comprehensive searching on the Internet. The difference between an electronic notebook, a PC or the World Wide Web is really just one of size, so we might expect the same quality of information and the same range of search routes in all three storage systems, with the promise of increased multimedia capacity at all levels as technology advances.

In practice, however, the three storage systems offer very different dictionary products, because they are subject to very different market forces.

A. Hand-Held Electronic Dictionaries
Least widely-known of all the different types of electronic dictionaries are the hand-held or pocket variety. They are largely ignored by lexicographers and reviewers, although Taylor and Chan (1994) and Nesi (forthcoming) attest to their popularity with users, and Sharpe (1995) discusses their pedagogical potential. These dictionaries appeal to users who may find paper-based learners' dictionaries inaccessible, but unfortunately it is difficult for teachers and academics to check on their accuracy, coverage and treatment of words. For one thing they are sold in electronic goods stores rather than bookshops, and are advertised in terms of their technological rather than their lexicographical features; for this reason they are not accompanied by editorial notes or the “front matter” found in hard-copy learners' dictionaries. Hand-held dictionaries are particularly popular in South-East Asia, and most are bilingual or multilingual, making a thorough assessment even more difficult for the reviewer who does not know Chinese, Japanese or Korean. Perhaps the biggest obstacle to comprehensive reviewing, however, is that there are hundreds of hand-held devices on the market; older models are continually being replaced, and each costs many times the price of a hard-copy dictionary, but unlike hard-copy dictionaries hand-held devices are not available for consultation in public libraries.

Sharpe (1995) examined a number of Japanese-English electronic dictionaries, including dedicated hand-held devices and notebooks with optional dictionary extensions, in order to identify ways of adapting such devices to help English learners of Japanese. However, because it is difficult and expensive to obtain a wide selection of currently available hand-held dictionaries, other researchers have asked users to act as informants to piece together a comprehensive picture of the role of such dictionaries in language learning. Taylor and Chan (1994) surveyed 494 student informants in Hong Kong, most of whom preferred hand-held dictionaries to dictionaries in book form because of the ease and speed of electronic look-up, even though they also believed that paper-based dictionaries were more detailed and accurate. Taylor and Chan's study suggests that many hand-held devices combine up-to-date access software with out-of-date text, containing all the defects associated with the smaller hard-copy bilingual dictionaries. Of the twelve Hong Kong English language teachers interviewed by Taylor and Chan only four had used hand-held dictionaries themselves, and although they had all taught students who used hand-held dictionaries only one teacher claimed to actively encourage their use. All the teachers said that they would prefer their students to use printed dictionaries.

Although there is not yet much use of hand-held dictionaries in Britain, data from ten overseas students who used hand-held dictionaries at Warwick University is reported in Nesi (forthcoming). According to these users, the prime advantage of hand held devices is that they are easy to carry around and use, but the informants also appreciated the variety of access routes that their dictionaries provided. Rather than typing in a headword, they noted that they could search for words via their synonyms, antonyms, a "sound-like" spelling, or a first language equivalent. Some informants had hand-held devices with an audio feature, so that they could hear the correct pronunciation of the search word, and most commented on the fact that their hand-held device could be linked to a computer and printer, and/or expanded by the addition of cards or mini disks.

By linking and expanding the hand-held device it can be converted for use in a variety of ways, many of which may be unrelated to language learning. The Warwick University users spoke of electronic games, an alarm clock function, calendars and calculators, and most of the devices used by Taylor and Chan's informants doubled as personal organisers. Taylor and Chan report on the availability of IC card extensions in Hong Kong which provide practice exercises for English language examinations, but Sharpe complains that

most EBDs (Electronic Bilingual Dictionaries) seem to use the content of printed dictionaries as their database without making any additions or alterations to take full advantage of the EBD's greater capacity for holding information.

(Taylor and Chan 1995: 48)

As yet there do not seem to be any extension cards available which function as a means of increasing dictionary coverage, or providing more examples of use, regional variation, idioms, collocations or syntactic patterns. Presumably this is not something that consumers demand. Educationalists and lexicographers, who support the development of collocational and pragmatic information in conventional learners' dictionaries, seem to have virtually no influence over the design, marketing and purchase of hand-held devices.

**B. Dictionaries on CD-ROM**

The situation is entirely different regarding dictionaries on 12cm CD-ROM. These products are strongly identified with their publishing house and the hard-copy dictionaries from which they were derived, and are intended as much for institutional use as for private use. Product development is thus led by current lexicographical and pedagogical trends
rather than by consumer demand, and the covers of learners' dictionaries on CD-ROM stress their educational value:

- "The First English Language Teaching Multimedia CD-ROM" (Longman Interactive English Dictionary)
- "Helping learners with real English" (COBUILD on CD-ROM)
- "The easy way to improve your English" (Longman Interactive American Dictionary)
- "The dictionary that really teaches English" (Oxford Advanced Learner's Dictionary on CD-ROM)

The storage capacity of a CD-ROM is about 600 MegaBytes, large enough for the half million headwords defined in the Oxford English Dictionary on CD-ROM, or the 44 million word database of the Britannica CDÖ. Most dictionary packages on CD-ROM use only a fraction of the storage space available, but still combine sources that exist as separate volumes in hard copy.


In each of these three packages the component volumes are cross-referenced to each other, so that when consulting one component the user may be directed to additional information about meaning, pronunciation, grammar or use to be found in the companion sources.

Not all dictionary packages on CD-ROM have taken this approach, however. The two Oxford learners' dictionaries, the Oxford Interactive Wordpower Dictionary and the Oxford Advanced Learner's Dictionary, contain electronic versions of just one published volume, and only the smallest collection in the Oxford Reference Shelf series, the Oxford Study Shelf on CD-ROM, permits two volumes to be searched as one (the Oxford School Dictionary and the Oxford Study Thesaurus). Other collections in the Oxford Reference Shelf series offer on a single disk several dictionaries designed for native speaker adults which can be installed separately or together, but which must be searched independently of each other. Thus the four dictionaries in the Oxford Compendium on CD-ROM and the sixteen dictionaries in the Oxford Reference Shelf on CD-ROM do not allow the user to conduct a search across several volumes at a time, although each volume is presented in the same format with the same search facilities.

The Oxford Interactive Wordpower Dictionary on CD-ROM and the Oxford Advanced Learner's Dictionary on CD-ROM are the two most recent electronic learner's dictionaries (a smaller version of Wordpower was originally produced on floppy disks). The decision by Oxford University Press not to include multiple sources, although doubtless partly influenced by the lack of compatible volumes from the Oxford stable, may also have been made in the light of the problems already encountered by its competitors. Print books created independently of one another have different numbering systems, different cross-referencing systems, and different levels of coverage of the same words. This means that mistakes occur when searches are made across multiple volumes (see Nesi 1996).

Sometimes cross-referencing also results in the over-emphasis of relatively unimportant data, or words which the dictionary does not adequately cover. This is the case with the Longman Interactive English Dictionary (LIED), which automatically directs users to its extensive Pronunciation Dictionary wordlist, rather than to the shorter wordlist for the Dictionary of Language and Culture. The Pronunciation Dictionary wordlist contains hundreds of rare and interesting words for which no information is provided, other than a pronunciation guide. For example, a search for the word sheep opens up a window listing not only sheep but also sheepdip, sheepdog, sheepfold, sheepish, sheepmeat, sheep's eyes, sheepshank and sheepskin. Only five of these words are defined; LIED provides just a transcription and an audio pronunciation for sheepfold, and transcriptions alone for sheepshank, sheepmeat, and sheepbit. Longman Interactive American Dictionary (LIAD), published more recently, avoids this problem by not including a pronunciation dictionary at all.

Slightly less frustrating, but potentially irritating, is the interface between the Language Dictionary and the Grammar in Collins COBUILD on CD-ROM (COBUILD). When the user types in a word; COBUILD looks for "hits" in each of the four reference areas: the Language Dictionary, the Grammar, the Usage and the Word Bank. Searches for common words frequently result in unexpected hits in the Grammar, not because the word is in any way grammatically unusual, but
rather because it is listed in the Grammar to exemplify what is typical of a word class or subclass. The word hat, for example, is flagged because it appears on one of the Grammar's many lists as an example of a common count noun.

Cross-referencing to the Word Bank is much more successful, perhaps because the corpus has not been edited for independent hard-copy publication. Most searches will result in Word Bank hits, although common words obviously produce a greater number of concordance lines. The problems of using Word Bank information are those inevitably associated with raw corpus data: there are numerous scanning errors, some Word Bank words are not defined in the Language Dictionary, and only a limited context is provided so the meaning of some expressions is difficult to establish. Nevertheless the Word Bank is a useful innovation, indicating the potential of electronic dictionaries to provide far more corpus data than their printed counterparts. There is no reason why electronic hand-held dictionaries of the future should not provide concordance lines on IC cards, or why the next generation of dictionaries on CD-ROM should not provide selections from more specialised corpora, distinguishing, for example, between spoken and written use, and/or British, American and Australian varieties of English.

Despite some minor difficulties with cross-referencing, the multiple sources of LIED, LIAD and COBUILD provide users with instant and comprehensive accounts of word behaviour. Similar accounts could only be gathered with great difficulty from paper sources. It should also be noted that all five learners' dictionaries on CD-ROM allow cross-referencing within a single source; information on any word in any component can be obtained by highlighting the word and calling up a new window (a process known as "chaining" or "hyperlinking"). This feature may largely solve one of the greatest problems of monolingual learner lexicography - the fact that learners are sometimes unable to understand the definitions for the words that they have looked up.

Another exciting feature of electronic dictionary design is their potential to allow "fuzzy" or complex searches for groups of words containing the same phonological, syntactic or semantic features. In contrast, the only feasible searches in a monolingual A-Z printed dictionary are for words that the user already knows, and searches in a semantically organised paper-based picture dictionary or lexicon can often only be made via a superordinate term. Only a few hard-copy monolingual dictionaries, such as the Longman Language Activator (1993), the Word Routes series from Cambridge University Press (starting 1994), and the Oxford Learner's Wordfinder (1997) allow for "onomasiological" searches directly via word meaning, and because of space restrictions these dictionaries only define a relatively small number of words.

Dodd (1989: 89) foresaw that electronic dictionaries of the future might permit search routes such as:

"sounds like A"
"rhymes with B"
"is spelt like C"
"has an etymology of D"
"dates from year/century E"
"is used in style of F"
"is used in technical field G"
"is an antonym of H"
"is a synonym of I"
"is a hyponym of J"
"is a superordinate of K"
"includes the word(s) L in its definition"
"is of grammatical class M"
"has syntactic valency or pattern N".

A few of Dodd's routes have been available for a number of years in electronic versions of the Oxford English Dictionary, and more recently in collections belonging to the Oxford Reference Shelf series. We have also already noted that some handheld electronic dictionaries offer a variety of search routes, via synonyms, antonyms, a "sound-like" spelling, and first language equivalents. Learners' dictionaries on CD-ROM, however, permit an even wider variety of routes, although not all types of search are easy to accomplish, and some yield disappointing results because the print dictionaries on which the electronic dictionaries are based do not contain enough information in the specified search categories.

Dodd's searches A, B and C can be performed by means of wildcards. LIED, LIAD and OALD allow missing letters to be indicated at the beginning, middle or end of words, but COBUILD only allows the "trailing wildcard", representing word-final missing letters. LIED also has a Spelling Note box, which lights up automatically for all words beginning with c, f, i, j, k, n, r, and s, and suggests possible alternative letters. Although this provides some limited help for shaky spellers, it only
Guillot and Kenning (1994) found that interfacing was one of the most popular features. In a rare study of learners while they actually worked with an electronic dictionary (the Robert Electronique), they found that interfacing was one of the most popular features.

None of the learners' dictionaries provide etymology, but LIED and LIAD allow users to search by date for people listed in the Longman Dictionary of English Language and Culture and the Longman Dictionary of American Language and Culture, a route that has some of the characteristics of Dodd’s “dates from year/century E”.

Dodd's search route L (“includes the word(s) L in its definition”) can be achieved in all the learners' dictionaries on CD-ROM by conducting a full text search with the Boolean operators OR and AND. The more elaborate searches for style, register, antonyms, synonyms, hyponyms, superordinates, word classes and valency patterns (Dodd's routes F, G, H, I, J, K, M and N) are best conducted by a “field search” in COBUILD, or a “filtered search” in OALD. COBUILD offers ten fields to search separately or in combination: Headword, Inflections, Meaning, Examples, Grammar, Synonyms, Antonyms, Superordinates, Phrases and Derived Words. OALD users can apply up to five filters: Geography, Subject, Specialism, Register, Word Class and Word Class Subcategory.

These searches can help the user discover sets of words that share the same semantic features, syntactic patterns, morphological structure or context. It is possible, for example, to create a list of all the hyponyms of vehicle in COBUILD, or of all the nouns marked “Australian” in OALD. However, it must be noted that the searches are not easy to conduct, they often throw up a number of irrelevant references as well as those that are useful, the accompanying manuals do not explain the search procedures adequately, and the user has to understand the structure of OALD and COBUILD in order to frame appropriate search commands. Perhaps these finer details of electronic dictionary use are more likely to be mastered by teachers and materials writers rather than by learners.

The fact that there are more complex search options available on CD-ROM is not so much due to the storage capacity of the CD-ROM, but rather to the sophistication of the book dictionaries on which the electronic versions are based. It is only possible to search for examples of a given register, syntactic structure etc. if the text explicitly provides this information. The current crop of electronic dictionaries were all published several years after their hard-copy equivalents, and now all but OALD and the Oxford Interactive Wordpower are based on editions that have been superseded in print. The next generation of dictionaries on CD-ROM may offer a wider range of information and search options if they are derived from the most recent corpus-based learners’ dictionaries, but still better products will be created if publishers are prepared to invest in purpose-built electronic texts rather than those adapted from hard-copy sources. Electronic dictionaries would be most effective if they were designed from scratch with computer capabilities and computer search mechanisms in mind.

One further feature of electronic learners' dictionaries is dependent on the storage capacity of the CD-ROM. All but COBUILD are multimedia applications, offering access to spoken headwords, pictures, photos, and, with the exception of OALD, video clips. LIAD also contains some animated sequences. The spoken headwords are a very useful addition, as they provide a much better guide to pronunciation than the IPA symbols in the printed dictionary, which many users are unable to interpret. Although the print dictionaries also contain still visuals, there are more pictures and photos in the electronic dictionaries, and OALD and the Oxford Interactive Wordpower Dictionary both permit interaction (the user can click on an image or part of an image to see a label and hear how the word is pronounced). The scripted video sequences in LIED and LIAD are less integral to the design of the dictionaries as a whole, and because very few words in the Dictionary of English Language and Culture and the Dictionary of American Language and Culture cross-reference to the clips, they are not really an effective way of providing the learner with examples of the way search words are used in context. Nevertheless Longman has made a selling point of the video component, and reviewers comment on its popularity with learners (Eastment 1996, McCorduck 1996).

Handheld devices and dictionaries on CD-ROM offer different capabilities for users, and will be used in different environments. The hand-held device is highly portable, and is therefore useful for receptive applications such as reading and listening; one of the Warwick University users, for example, commented on its value as a tool for lecture comprehension. CD-ROMs, on the other hand, have to be accessed at a computer workstation, possibly via an internal network at the educational institution, and are therefore more suited to productive use, as an aid to writing or work on other software applications. All but COBUILD allow direct access from Windows-based word-processing programs (known as "interfacing"), and LIAD also permits the same access from pages on the World Wide Web. Access by hot-key to an authoritative learner's dictionary is clearly an improvement on the thesaurus facility offered in most word-processing packages; in a rare study of learners while they actually worked with an electronic dictionary (the Robert Electronique), Guillot and Kenning (1994) found that interfacing was one of the most popular features.
Guillot and Kenning also comment very positively on the educational value of electronic dictionary use. They found that the *Robert Electronique* encouraged browsing and vocabulary development, and also made it possible to work in pairs and groups on dictionary skills activities, because a view of the screen was more easily shared than a view of the page.

The English learners' dictionaries on CD-ROM have been designed with teaching and learning in mind. LIED, LIAD and OALD allow learners to create their own personalised learning files for access in conjunction with the CD-ROMs, and *Oxford Interactive Wordpower Dictionary* and OALD have sections containing interactive educational games. All four dictionaries have print and copy options, so that pictures and information can be incorporated into paper-based learning materials, and LIAD even provides a thirteen-page file of teaching suggestions that can be copied onto hard disk and accessed as a word-processing file.

If hand-held dictionaries are designed and marketed for the non-academic consumer, learners' dictionaries on CD-ROM are clearly aimed at the teaching profession. Although some individual learners may buy the CD-ROMs to consult at home, they are far less evident in bookshops than their hard-copy sources and are rather too expensive for most full-time students to buy. Instead, the CD-ROMs tend to be ordered directly from the publishers' catalogues for use by schools and universities in their classrooms and self-access centres.

C. Dictionaries on the Internet

Like hand-held dictionaries, most dictionaries on the Internet lie outside the control of the pedagogues. Internet searches are too costly and time-consuming to play a regular part in conventional classroom teaching, and publishing houses that have invested heavily in dictionary development are unwilling to distribute their products via a platform which traditionally provides its services for free, and does little to guard against copyright infringement.

As Carr (1997) comments: "It is ironic that the pioneers in computerised editing and CD-ROM books are struggling against their technologies spreading onto the Net" (Carr 1997: 210) but the Internet is in many ways an excellent medium for the storage and retrieval of large amounts of dictionary data, and some publishing pioneers have found ways to exploit it by exacting payment from registered users. *Oxford English Dictionary Online*, for example, is in preparation at the time of writing and will be available by subscription. The size of the OED, its cost in book form, and the serious and scholarly habits of those who consult it all justify the establishment of a payment scheme, but the case is very different for learners' dictionaries, which are tiny by comparison, relatively cheap to buy, and are used by young students of many different nationalities (and currencies!).

At the time of writing there are no learners' dictionaries for non-native speakers of English available on the Internet. English dictionaries there are aplenty, but none are specifically written for the language learner, and many in fact require excellent linguistic judgement from their users, who must separate useful entry information from the useless or misleading.

Broadly speaking dictionaries on the Internet fall into three categories: those available by subscription, those in the public domain because they are out of copyright or are unfinished products, and collaborative projects in the process of construction by contributing users.

The most popular Internet dictionaries are in the public domain. Two that have links to most dictionary and language Websites are the *Hypertext Webster Interface*, which has, according to Carr (1997) "an obscure copyright status", and *Roget's Thesaurus*, based on the copyright-free 1911 edition. Hundreds of others have no named hard-copy source. Storrer and Freese (1996) comment on the unreliability of public domain on-line dictionaries as compared to dictionaries in book form; they point out that nobody takes responsibility for the accuracy of the information Internet dictionaries provide, and that both the Web addresses and the page contents are constantly changing. Nevertheless Storrer and Freese admit that the dictionaries are useful in areas where there is rapid terminological growth, because new words can be added to an Internet dictionary far more quickly than they can be admitted to a hard-copy book. Carr (1997) claims that none of the public domain dictionaries available on the Internet contain full definitions of a full range of headwords. He too is complementary, however, about the coverage of technical words, and comments on the range of technical translations available via bi-and multilingual Internet dictionaries such as *EURODICAUTOM*.

Websites which permit simultaneous searches of a large number of dictionaries such as *EURODICAUTOM* and *OneLook Dictionaries* put the storage capacity of the Internet to good use, although obviously the value of a search is still dependent on the quality and coverage of the multiple volumes. A search for *andrology* in *OneLook*, for example, did not result in a single hit despite the fact that hundreds of dictionaries had been accessed. (This was a genuine query;
androgogy is a technical term in the field of Education and is not listed in the OED. It means “teaching adults.”

As well as permitting multi-dictionary searches, some of the public domain dictionary Websites offer unusual search routes. Casey’s Snow Day Reverse Dictionary, for example, uses "n-gram analysis" ("a method of matching documents based on the statistical similarity of occurrences of ... combinations of letters") to match the definition the user types to entries in the Hypertext Webster Interface. Unfortunately the "matches" provided by Casey’s Snow Day Reverse Dictionary would be of little use to language learners, and if taken seriously could result in some strange mistakes. Androgogy is not one of the forty-eight words suggested for “teaching adults”. The first ten are, in descending order: extramarital, vendor, catlike, ascription, dreamlike, saw mill, childhood, pulpit and instruct.

Other types of search route are available through WordNet, where users can specify a search word, a word class, and a search type (for example Synonym, Hyponym or Antonym), to achieve results similar to those of a COBUILD field search. The Wordbot Website is reminiscent of the interfacing facility provided by LIED, LIAD, and OALD, because users have access to a range of bilingual and multilingual dictionaries on a split-screen, alongside the document they are translating. Wordbot is a good example of a collaborative project, or what Carr (1997) describes as "bottom-up lexicography". Although basic dictionaries have been provided since the start of the project, these have been added to by successive users, and are continually in the process of development. The Website information page explains:

If everyone contributes just a little, then everyone will gain a lot. Please help fill in the gaps in the dictionary. If you find that a word is missing, and you know the correct translation, please take the time to supply it, so that everyone can benefit from your knowledge.

The obvious problem with this system is that users may supply inadequate or even false information. On sites such as the collaborative Alternative Dictionaries project for slang and bad language many of the entries seem to be humorous inventions rather than recorded instances of use. Perhaps users of technical dictionaries are less likely to make up words, but if they do not rely on their own word knowledge they may turn to more authoritative sources. The Website compilers warn against indiscriminate copying:

If you do enter information from a physical dictionary, be sure not to type it in verbatim, since that could be a copyright violation.

Nevertheless, plagiarism seems an inevitable hazard of projects of this kind.

Bottom-up lexicography is still in its infancy, and is experiencing plenty of teething problems. For the time being users should approach collaborative dictionaries with caution, but Carr (1997) is optimistic about their future, and with appropriate editorial control they may add significantly to the sum of human word knowledge, particularly in new areas of Science and Technology.

III. CONCLUSION

Computer-based dictionaries offer three great potential benefits for users: they are quick and easy to use, they can provide access to large amounts of data, and they are interactive. This is true for all three types of electronic dictionary: hand-held devices, dictionaries on disk and CD-ROM, and dictionary Websites. It is also true for most of the dictionaries that are currently available, even though there remains enormous scope for development and improvement.

At present the best electronic dictionaries for learners seem to be those that are published on CD-ROM, but hand-held devices and Websites are probably better storage and retrieval systems; the former because of their portability, and the latter because of their immense capacity and easy links to other computer-based applications. Market forces are all that prevent the major publishers from transferring their products to IC cards and the Internet, but markets will change, especially as the Internet stabilises, and copyright law is reinterpreted to deal at last with problems raised by the electronic transfer of information.

Perhaps there will always be a place for the fat hard-copy dictionary. There is something reassuring about paper-based products; we can touch them, write on them, and when we pay for them we acquire something very tangible in return. But
computer-based dictionaries are gaining the ascendancy. As we receive more and more of our daily supply of information on-line, it will become an increasingly obvious alternative to access word knowledge on-line too, using whatever search routes suit our needs.

References


APPENDIX

Some English Native-Speaker Dictionaries on CD-ROM


Some English Learners' Dictionaries on CD-ROM

*Collins COBUILD on CD-ROM* 1994 Worthing: HarperCollins


Some Dictionary Sites on the Internet

Alternative Dictionaries http://www.notam.uio.no/~hcholm/altlang
Casey's Snow Day Reverse Dictionary (and Guru) http://www.c3.lanl.gov:8075/cgi/casey/revdict
EURODICAUTOM http://www.echo.lu/echo/databases/en/eu92.html
Hypertext Webster Interface http://cgp.cs.cmu.edu:5103/prog/webster
An Index of On-line Dictionaries http://www.bucknell.edu/~rbeard/diction.html
OneLook Dictionaries http://www.onelook.com/index.html
Online Rhyming Dictionary http://www.csx.cmu.edu/~dougb/rhyme.html
Roget's Thesaurus http://humanities.uchicago.forms_unrest/ROGET.html
WordNet http://www.cogsci.princeton.edu/~wn/w3wn.html

Download this paper as a doc-file.

Download this paper as a rtf-file.