Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art

Hilary NESI, Warwick, UK

Abstract
This paper attempts to define, describe and categorise the electronic dictionary, and consider the skills associated with its use. Prior research is briefly reviewed, and new research approaches are delineated.

1 Introduction
As lexicographers and language educators explore new ways to present information about word meaning and use, the traditional distinctions between different categories of reference work are becoming increasingly blurred. Many recently published dictionaries are hybrids, merging features associated with more than one kind of wordbook, and taking on some of the duties of encyclopaedias, pedagogic grammars and teaching materials, and this is particularly true of electronic dictionaries which do not have the organisational and spatial constraints of hard-copy dictionaries, and can retrieve and combine information according to the specifications of the user. However, as the traditional image of the dictionary changes, new distinctions between dictionary types arise. Most obviously, electronic dictionaries can be distinguished from their paper-based counterparts, but perhaps more surprising are the distinctions that have arisen between electronic dictionaries accessed via a hand-held device, a PC or the World Wide Web.

This paper first of all describes and categorises the range of electronic dictionaries, and then goes on to investigate the extent to which electronic dictionary use differs from conventional dictionary use. The findings of the first empirical studies suggest that there are some important differences, and the last part of this paper will briefly review this prior research, and point towards some future research issues.

2 What is an electronic dictionary?

The term electronic dictionary (or ED) can be used to refer to any reference material stored in electronic form that gives information about the spelling, meaning, or use of words. Thus a spell-checker in a word-processing program, a device that scans and translates printed words, a glossary for on-line teaching materials, or an electronic version of a respected hard-copy dictionary are all EDs of a sort, characterised by the same system of storage and retrieval.

It is the retrieval system, rather than the information content, which makes electronic dictionary use such a revolutionary experience compared to the consultation of a hard-copy dictionary. Books 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, their valency and collocational patterning,
or simply their letter combinations (as in the humble spell-checker). The A-Z sequence places
headwords in an order which is virtually meaningless, shedding no light on the relationships be-
 tween words that are alphabetically distant, and complicating searches for phrases and idioms.
A text search of an ED, on the other hand, provides almost instant access to groups of words in
any information field created during its development.

Thus the Oxford English Dictionary on CD-ROM can act, among other things, as a synonym
finder, a dictionary of quotations, or a terminology bank, simply through the rearrangement of
the information which is provided in the original paper-based text. Similarly, several learners’
dictionaries in electronic form offer filtered searches through examples, definitions and gram-
mar codes, while WordNet, a dictionary purpose-built for electronic retrieval, offers a full range
of links from the search word to its synonyms, co-ordinate terms, hypernyms and hyponyms.
Pocket electronic dictionaries (or PEDs) permit poor spellers to type in a "sound alike" version
of their search word and select from a range of near matches the dictionary suggests, while
cd-rom dictionary users can "hyperlink" and "interface" to call up entries for unknown words
within the entry they are consulting, or in an unrelated on-line text [Nesi 1996,1999].

The range and convenience of such search routes in EDs are, of course, no guarantee of the
quality of the information content. Many EDs are still just versions of lexicographically out-
dated hard copy dictionaries, converted without any change or addition to the original entries.
Some of the Internet dictionaries that are linked most frequently to language study websites,
for example, are Roget’s Thesaurus, based on the copyright-free 1911 edition, and Webster’s
Revised Unabridged Dictionary, first published in 1913 and criticised by modern users for its
"offensive or politically incorrect" entries [File 1999]. Similarly, although in some parts of the
world there is a huge market for (usually bilingual) PEDs, relatively little seems to have been
spent on lexicographical input as opposed to software design, and they too have tended to repro-
duce old hard-copy dictionary text [Taylor / Chan 1994]. Interest in the lexicographical aspect
of some electronic dictionary projects is so weak that the provenance of the dictionary is down-
played or even ignored. Many Internet dictionaries and PEDs have no named hard-copy source,
and in recent studies of the use of dictionaries to support on-line teaching materials, the source,
quality and appropriacy of the definitions is seldom discussed, all attention being focused on
the user interface and technical innovation [Nesi forthcoming].

"Bottom-up lexicography" [Carr 1997] poses another threat to the quality of reference material
on the Internet. A number of dictionary sites invite collaboration from their users to supplement
their basic dictionary database, with the obvious risk that contributions may be inaccurate [Nesi
2000]. There are ways to prevent the spread of information in collaborative dictionary projects,
however. New words and their meanings must be selected and verified before being added to the
database, and collaborators must be prepared to take responsibility for their contributions. At the
website for Cambridge International Dictionaries Online, for example, a failed search prompts
the appearance of a contribution form with space to type in the missing word, its meaning, an
example sentence and the contributor’s name and e-mail address (information which helps to
inform the final editorial decision).

Dictionaries for learners of English on the Internet (Collins Cobuild Student’s Dictionary, Cam-
bridge International Dictionaries Online and Longman Web Dictionary) all signal that they are
flexible structures which can be expanded and updated according to users’ needs. Publicity for
Electronic Dictionaries in Second Language Comprehension

Longman Web Dictionary promises that "Regular updating online means it will NEVER go out of date!" and Cambridge and Cobuild exploit the interactivity of the Internet by inviting their users to comment on dictionary entries. Updating a cd-rom is already a less costly business than producing a new hard-copy dictionary, but the Internet is the medium that permits continuous development. [Storrer and Freese 1996], who make quite negative comments about the public domain Internet dictionaries available in the mid 1990s, still argue that they are ideal in areas where there is rapid terminological growth, because a new word can be added to an Internet dictionary as soon as it is coined.

Apart from enabling faster and more varied retrieval pathways to updatable (although not always updated) dictionary information, the storage of data in digital form has had huge repercussions in terms of the quantity of information available to dictionary users. Because enormous amounts of electronic data can be stored in a tiny space, there is plenty of room to add to basic dictionary information, either by combining existing reference sources, including other types of non-reference material, or creating new reference material to complement the information provided in the original text.

On the Internet, where available space is virtually limitless, learners’ dictionaries have all expanded to contain more entries than their corresponding editions in hardcopy or on cd-rom, and some one-stop websites allow simultaneous searches of hundreds of native speaker and bilingual dictionaries [Carr 1997, Storrer and Freese 1996] (although quantity should not be confused with quality - as one such website warns: "We do not edit and are not responsible for the content of other dictionaries" [FILE 1999]). The storage space in a handheld device is obviously more restricted, but PEDs are so small that they can be packaged within handheld personal organisers, complete with calendars, clocks, and calculators [Taylor / Chan 1994], and more recently fax transmission and voice recording facilities. [Sharpe 1995] laments the fact that PEDs have not used the extra space to increase dictionary coverage or provide extra language information, but many recent PEDs have cut down on gadgetry, and offer specialist terminology banks, grammar guides, synonym and antonym searches and vocabulary building activities instead.

The first learners’ dictionaries on cd-rom made use of the additional space by amalgamating multiple types of reference material, including a variety of different dictionaries and grammar books, photographs, pictures, and video clips, and sometimes also learning activities and games. This approach was not a total success; as [Seedhouse 1997] comments of Collins Cobuild on CD-ROM: "the rationale... seems to have been roughly "Stick all the products we already have on a cd-rom and let’s hope somebody can find a use for it"". Some of the cross-referencing between sources in the first learners’ dictionaries on cd-rom was incomplete and confusing [Nesi 1996], and although the addition of games and multimedia sequences may have pleased the punters [Eastment 1996, McCorduck 1996] these features were not integral to the design of the dictionaries as a whole, and did not always suit the users’ backgrounds and abilities [Battenburg 1998, Nesi 1996]. The most recent learners’ dictionaries on cd-rom have just one hard-copy source, however, and make more precise use of multimedia. For example, whereas the Longman Interactive English Dictionary (1993) and the Longman Interactive American Dictionary (1997) contained video files of several "mini-dramas" largely unconnected with the dictionary look-up process, the new Oxford Advanced Learner’s Dictionary on CD-ROM (2000) only uses short video sequences to illustrate the meaning of hard-to-define verbs, and Longman
As electronic dictionaries become more widely used, and users become more discerning and less impressed by technological wizardry, lexicographers and dictionary designers are refining those features that take best advantage of the medium, and are producing more lexicographical material expressly for the Internet and cd-rom, without any hardcopy equivalent. Thus recorded pronunciation models, which make it unnecessary for the user to interpret IPA, are now provided by many PEDs and cd-rom dictionaries, including the Concise Oxford Dictionary and the latest learners’ dictionaries on cd-rom. Collins Cobuild Students’ DictionaryOnline also offers downloadable pronunciation files, but sound is not yet a widespread feature of Internet dictionaries. Field searches, too complex for all but the most skilful of first generation electronic dictionary users [Nesi 1996], now tend to be much simpler to conduct; one click from an entry in the Longman Web Dictionary, for example, leads straight to all "related entries" (those containing the search word), while CollinsE-Dict (1998) offers easily activated searches through words with similar grammatical behaviour, but has dispensed with the Word Bank and searches of semantic relations (antonym, synonym, hyponym and superordinate) offered by Collins Cobuild on CD-ROM (1995).

In the interests of simplicity and user-friendliness, some valuable features of first generation learners’ dictionaries on cd-rom have been lost. None of the new EDs offer "raw" language data to replace the Word Bank, for example, although this was a convenient way of providing users with additional examples of their search word in authentic contexts - the "most useful feature" of Collins Cobuild on CD-ROM according to [Seedhouse 1997]. On the other hand the role of the dictionary as a teaching and learning tool has been emphasised in some products, with interactive exercises included on the Cambridge and Oxford cd-roms.

3 Dictionary typology for the electronic age

In the last decade we have seen the emergence of four main categories of ED for language learning: the Internet dictionary, the glossary for on-line courseware, the learners’ dictionary on cd-rom and the pocket electronic dictionary. These categories were at first fairly distinct. Internet dictionaries were non-profit-making, and therefore used copyright free material which was generally derived from outdated sources or users’ contributions. Glossaries in on-line courseware were also not primarily commercial ventures, but were written by the staff of university language departments, often as part of a research project. Learners’ dictionaries on cd-rom were the products of the major dictionary publishing houses, and drew on the resources of reputable hard copy reference books. PEDs, on the other hand, tended to be developed and marketed by the makers of electrical goods, and were inspired by market forces rather than by pedagogical or lexicographical research.

These distinctions were never entirely clear cut, however, and in recent years have become even less so. A number of highly innovative dictionaries have been developed for the Internet as part of lexicographical or software design projects, such as The Free Internet Lexicon and Encyclopedia (a product of the Dictionary Development Group), Collins Cobuild Students’ Dictionary Online (developed with the Sprachwissenschaftliches Institut, Ruhr-Universität Bochum [Hoelter/Wilkens 1998] and WordNet (developed by the Cognitive Science Laboratory at Princeton
University [Miller 1998]). And whereas almost all dictionary applications on the Internet used to be free, we are now starting to see the advent of Internet dictionaries by subscription - the Longman Web Dictionary, like the Oxford English Dictionary, became available to the paying public in spring 2000.

Dictionaries of a kind can also be purchased as part of online courseware. English language teaching textbooks, complete with glossaries, are being converted to cd-rom, or written especially for the electronic format, while online courseware originally developed for internal use is now becoming commercially available.

The past few years has also witnessed the spread of learners’ dictionaries from disk to the Internet, starting with the Collins Cobuild Student's Dictionary Online in 1998, and followed by the Cambridge International Dictionaries Online in 1999 and the Longman Web Dictionary in 2000.

Only the PEDs remain, in fact, a distinct category. There is little evidence of the transfer of lexicographical material from the Internet or cd-roms to the handheld device, and the PED does not seem to have had much influence on the design of other kinds of electronic dictionary. The only major learners’ dictionary publisher to dabble in the PED market so far has been Longman, with the Handheld Longman Dictionary of Contemporary English (marketed by Seiko in 1995). PEDs may be ignored or regarded with suspicion by some researchers and members of the teaching profession, but their portability and flexibility is exceptional, and they may ultimately replace hardcopy dictionaries entirely, as [Bolinger 1990] predicted.

4 The skills of the electronic dictionary user

Can ED users rely on the skills they employ when consulting a paper dictionary, or are some of these skills irrelevant in an electronic context, and are new skills needed?

The first, and perhaps most fundamental dictionary skill is that of selecting an appropriate dictionary for a given task, and in this respect the advent of EDs has placed much greater demands on dictionary users. Different dictionary formats suit different language activities, so now the choice is not just between a learners’ dictionary or a native speakers’ dictionary, general or specialist, bi- or monolingual, but also between a dictionary that requires access to a networked computer, a dictionary with hotlinks to a word-processing program, or a portable dictionary to consult while speaking and listening, or while reading hardcopy text (the Seiko Quicktionary, for example, sometimes called "the reading pen", translates and pronounces words scanned in directly from the printed page).

Dictionary skills books tend to be linked to one specific title, and do not discuss the pros and cons of alternative publications. Nevertheless it is not hard to find comparative information about hardcopy dictionaries, as they are well known to teachers, reviewed in teaching journals, and available for inspection in bookshops and libraries. In contrast there are very few sources of unbiased advice available to the language learner who needs to select the right ED for the job.

Dictionary skills also involve an understanding of the micro- and macro-structure of the dictionary, its cross-referencing system and the contents of appendices and study pages. This kind of knowledge is particularly difficult for an ED user to acquire because only one screen page can
be viewed at a time, and the spread of contents can only be visualised, rather than viewed and handled as with a hardcopy book. Some EDs provide demonstration tours or “how to use” sections, but explanatory documentation tends to be poor compared to the lengthy front matter of hardcopy dictionaries. This is particularly true of the first generation of learners’ dictionaries on cd-rom, which came with instruction leaflets of just a few pages and gave minimal explanations of complex search procedures (see, for example [Seedhouse 1997]).

While some aspects of dictionary use pose particular problems for ED users, other skills which are frequently practised in dictionary workbooks and dictionary skills teaching sessions are more or less irrelevant to the process of ED consultation. ED users do not need to worry much about alphabetical ordering and letter distribution, neither do they need to learn IPA if the dictionary offers recorded pronunciations of all entry words. The ability to identify the appropriate form of the look-up item may not matter much either, as many EDs accept “searches by letter” (type in the first letter(s), and select from a list of lemmas), or “sound alike” searches, or wildcards, and some automatically display the lemma of whatever inflected form the user types in. Instead, learners need to acquire the skills to use these dictionary functions correctly, as they are not always obvious. In an experiment comparing OALD in hardcopy and on cd-rom, for example, subjects failed to find appropriate headwords because they had keyed in the exact form of the word found in their reading passage, and ignored the base form which appeared in a separate window [Nesi forthcoming].

Electronic dictionary use is often so quick and convenient that there is a danger that users will be less inclined to query entry information or adjust its meaning according to the context in which the word appears. [Taylor / Chan 1994] found that some teachers were worried that easy dictionary access would prevent the development of their students’ reading skills. Similarly [Sharpe 1995] comments on the possibility that the most easily extracted information may require least thought, and be soonest forgotten. Unthinking acceptance of dictionary information is particularly dangerous when the ED is based on an inaccurate, outdated or limited source. Although they are rarely thought of as such, the most widely used EDs are the spellcheckers attached to our word-processing programs, capable of creating nonsense of well-written text if their every suggestion is activated without query. In a recent survey of dictionary skills training in higher education, however, only one out of 35 university lecturer informants named spellchecker use as a training need [Nesi 1999b].

At the highest level, dictionary skills require an understanding of the principles and processes of dictionary compilation, and the ability to criticise and evaluate lexicographical works. Few language courses have the time to devote to this kind of study, although of course many professionals acquire such competence through experience. Dictionary writing projects have been posited as a means of helping school-age children gain higher-level dictionary skills [Whitfield 1993, Department for Education and Employment 1998], and such projects are clearly much easier to manage when computer resources are employed. For one such scheme the Australian National University encourages the use of Cardfile, a Microsoft Windows Application, to store and edit dictionary entries [Schools Dictionary Project 1997]. This combination of IT and lexicography seems to be a very appropriate way to develop the critical and practical skills of future generations of electronic dictionary users.
5 Prior studies of electronic dictionary use

The study of electronic dictionary use is still in its infancy. Although [Roby 1999] reports a few 1980s experiments involving the use of online glosses, the linking of language learning materials with dictionary definitions only really became feasible in the 1990s, as hypertext became a standard feature of computer systems. To date, most ED research has been concerned with the use of glossaries relating to modern language course materials, such as the German Cyberbuch [Chun / Plass 1996], Spanish magazine articles [Aust et al 1993, Knight 1994, Roby 1999], and excerpts from a French novel and poem [Lyman-Hager et al 1993, Lomicka 1998]. The results of these experiments strongly favour the use of computer-based lexical information. In Knight’s study, for example, where subjects read text with and without access to online glosses, those with glosses took longer to read the passage but achieved higher vocabulary and reading comprehension scores. [Aust et al 1993 and Roby 1999] found that users of online glosses looked up significantly more words than subjects with hardcopy glosses, and although in these two studies differences in comprehension scores were found to be insignificant, in [Lyman-Hager et al 1993] the subjects who used online glossing achieved significantly higher vocabulary quiz scores than those using glosses on paper. [Chun / Plass 1996] and [Lomicka 1998] also experimented with multimedia glossing online, and found that access to glosses with still pictures as well as text improved word recall.

But although these studies point to the value of creating hyperlinked glosses to key texts that students are required to study in depth, they ignore the fact that language users need to decipher many other spoken and written texts which cannot possibly be glossed in advance. Dictionary entries differ from glosses in that they provide the word’s "signification", applicable in many different contexts, rather than its "value" in a given text [Roby 1999], but this means that interpretation of a dictionary entry is usually harder than interpretation of a textual gloss, involving consideration of the word’s grammatical and syntactic behaviour, and possibly some adjustment of meaning.

So far there have been very few studies of ED use, however, to compare with studies of online glossing. In one qualitative study [Guillot / Kenning 1994] report on the behaviour of students engaged in the use of the Robert Electronique (1992). They note that the electronic format encouraged exploratory browsing and multiple searches, and they also comment on a pedagogical advantage of ED use in a classroom setting: on-screen look up is a more public event than book consultation, providing opportunities for teachers and fellow students to discuss dictionary entry information together.

The findings of my recent study of the use of OALD on cd-rom and in book form [Nesi forthcoming] accord with those of earlier experiments. The OALD definitions on screen and in print were identical, yet the ED users looked up more words, found the process easier, and were more satisfied with their dictionary consultations.

These first studies thus provide conclusive evidence for the value of the electronic medium, but have hardly begun to explore the potential of this medium to transform the look-up process. Research into the use of glossing has largely concentrated on the effects of technology and the development of online courseware, rather than the effects of defining style and entry organisation; research into ED use has focused on the comparison between existing EDs and their
paper-based counterparts, rather than the value of possible alternative approaches to dictionary design. Electronic dictionaries are certainly capable of offering the user choices of source material, search routes, and level of detail [Leech / Nesi 1999], yet we still know little of the benefits of different potential options. An exciting new medium demands creative new approaches to dictionary design, and more thorough exploration of dictionary users’ wants and needs.

References


http://polyglot.cal.msu-edu/llt/vol1num2/article2/default.html

Electronic Dictionaries in Second Language Comprehension


