

The demands of users and the publishing world: printed or online, free or paid for?

Hilary Nesi

Accepted manuscript PDF deposited in Coventry University's Repository

Original citation:

'The demands of users and the publishing world: printed or online, free or paid for?', in *The Oxford Handbook of Lexicography*, ed. by Philip Durkin, pub 2015 (ISBN 9780199691630)

Publisher: Oxford University Press

Reproduced by permission of Oxford University Press -

<https://global.oup.com/academic/product/the-oxford-handbook-of-lexicography-9780199691630?cc=gb&lang=en>

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.

Nesi, H. (2015) The demands of users and the publishing world: printed or online, free or paid for? In: Durkin, P. (ed) *The Oxford Handbook of Lexicography*. Oxford: Oxford University Press pp. 579-589

Pre-publication version

Abstract

This chapter examines the extent of the digital migration of reference works from print to screen, and the effect this is having on dictionary publishers and dictionary users. It discusses the place of the human lexicographer, and possible new sources of e-dictionary revenue in the new ‘give-away’ internet environment. It also considers the automatic and collaborative generation of dictionary content, quality issues, and the needs and preferences of dictionary users around the world.

Keywords

e-dictionary; cd-rom; pocket electronic dictionary; wi fi; internet; web corpora; natural language processing, web analytics, social networking; search engine optimization

1.1. Introduction

In his ‘Guides to tomorrow’s English’ (1998), Tom McArthur considered English dictionaries of the past, present and future. ‘Today’s English’, starting from the early 19th century, reflected the influence and range of the language resulting from the growth in international trade, the Industrial Revolution, the expansion of the British Empire and the increasing power of the United States. In this period English lexicography was primarily associated with just three locations: Oxford (Oxford University Press), Edinburgh (Chambers), and Springfield, Massachusetts (the Merriam publishing company).

The period of ‘tomorrow’s English’ had already begun for McArthur, however, with the increase in the use of English as an international language, and the development of English dictionary resources specifically for foreign language learners. McArthur identified eight pragmatic developments associated with the English dictionaries of ‘tomorrow’. Alongside *globalization* of the market and *localization* to meet the language learning needs of a particular country or group of countries, *thematization* marks a move away from alphabetical formats, *bilingualization* and *semi-bilingualization* involve the inclusion of data from languages other than English, and *nationalization* and *regionalization* mean that dictionaries are being developed outside Britain and the USA, to cater for the interests and needs of users who may or may not be ‘traditional native speakers’. The eighth pragmatic development McArthur identified, *electronicization*, began with the use of computers to assist the lexicographical process, and at the time of McArthur’s article had not yet fully overturned the methods of storage and retrieval long associated with print dictionaries.

All the developments on McArthur’s list reflect our shifting relationships with the multiple language varieties and the multiple environments in which dictionary users operate. However it is undoubtedly *electronicization* that has had the greatest impact since 1998, influencing all other aspects of dictionary creation and use. CERN opened up the internet as a possible site for e-dictionaries in 1993, but in the 1990s lexicographers’ focus was largely on the collection and exploration of digital language data, and in 1998 there were still only about 400 English dictionaries on the World Wide Web (Li 1998: 21). Most e-dictionaries were still being distributed on cd-rom or on stand-alone mobile devices (pocket electronic dictionaries). In the early 2000s, however, internet access gradually became faster and more reliable as high

speed broadband connections became available. As a result, thousands more digital versions of print dictionaries went online. At the same time wiki software became an open source tool, leading to the rapid expansion of ‘collaborative lexicography’ where dictionary information is created and edited by users (Nesi 2008, Krek 2011, Meyer & Gurevych 2012). In this new environment wi fi connected mobile devices are overtaking pocket electronic dictionaries containing licensed dictionary content, and publishers are having to rethink their entire marketing strategy. The new web technology seems to promote a give-away culture where files are shared, anonymous amateur editors construct reference works, and open source operating systems are created and distributed free of charge (Hall 2008: 206). Yet this same technology is capable of deriving ‘big data’ with enormous commercial potential from social media, internet archives, and internet search indexing. Whether such data can be harnessed to the advantage of dictionary publishers and lexicographers remains to be seen.

Recently, a survey conducted by Müller-Spitzer et al. (2011) asked English and German participants to rate ten aspects of online dictionary usability. The results indicated that *reliability of content*, *clarity*, *up-to-date content* and *speed* were regarded as the most important features. *Long-term accessibility* seemed to be of medium importance to users, and there was far less enthusiasm for *links to other dictionaries*, *adaptability*, *suggestions for further browsing*, and *multimedia content*. The researchers noted that ‘the classical criteria of reference books (e.g. *reliability*, *clarity*) were both ranked and rated highest, whereas the unique characteristics of online dictionaries (e.g. *multimedia*, *adaptability*) were rated and ranked as (partly) unimportant’ (2011: 207).

Users’ attitudes towards online dictionaries could be changing, however. A second survey (Koplenig, 2011) found that users were more inclined to value

multimodal and user-adaptive interfaces once these features had been explained to them more clearly. Moreover Kilgarrieff (2005) notes the move away from ‘status symbol’ monolingual dictionaries, and Rundell (2011) claims that amongst digital native users (in their late teens and early twenties) dictionaries are no longer regarded as authoritative in the same way as before.

Dictionary portals such as *OneLook* and *Allwords* now include *links to other dictionaries* as standard, serving as metasearch engines across multiple sources which might contain the information the user requires. The *OneLook* site, for example, indexes 1062 dictionaries of varying types, ages, and reliability, including bilingual dictionaries and dictionaries for learners of English. More radically, portals may supplement their dictionary resources with *suggestions for further browsing* in the form of examples from media websites (Nesi 2012) and/or pedagogical materials (Campoy Cubillo 2002).

1.2. Computer generated lexicographical input

Grefenstette (1998) asked whether there would still be lexicographers in the year 3000. He listed the four steps undertaken by the ideal lexicographer, as identified by Kilgarrieff (1992):

- 1) gather corpus of citations for a given word
- 2) divide the citations into clusters
- 3) decide why the cluster members belong together
- 4) code the conclusions into a dictionary definition.

Steps 1 and 2, Grefenstette argued, could be achieved by computers, but steps 3 and 4 would remain the work of humans because they entail ‘drawing distinctions and

contrasts between shared experiences and expectations, explaining what makes this group different from other groups that the human user knows' (1998: 38).

Rundell (2009) reflects on Grefenstette's predictions in terms of the way technology has decreased the need for intervention in the lexicographical process, and how the advent of large corpora mean that dictionaries are no longer solely concerned with words, but also with language systems and syntagmatic networks. The four-step lexicographical process has now become too simple, a view that Leroyer (2011:122) seems to support when he argues that lexicography should no longer be regarded as a subset of applied linguistics, but as 'a unique discipline at the crossroads of social and information sciences and technology'.

In 2002 Esposito envisaged a 'new class of lexical applicationsbased on machines talking to machines'. These applications would eliminate the need for human mediation, but would still draw on the dictionary databases that human lexicographers had produced. An example of this type of application is the now defunct *Casey's Snow Day Reverse Dictionary*, which used 'n-gram analysis' ('a method of matching documents based on the statistical similarity of occurrences of ... combinations of letters') to match a meaning provided by the user to entries in the *Hypertext Webster Interface* (Nesi 1998). A modern equivalent of *Casey's Snow Day Reverse Dictionary* is the reverse dictionary provided by *Onelook*, which searches the full text of hundreds of online reference sites to find definitions conceptually similar to the words the user types in.

As part of this trend away from alphabetical searches and towards meaning-related queries – *thematization*, as predicted by McArthur (1998) - e-dictionaries are also using large lexical databases known as 'wordnets' to provide onomasiological search routes. Wordnets group words into interlinked sets of cognitive synonyms

(Fellbaum 2005) and were originally developed for use by artificial intelligence systems. The commercial online dictionaries *a2zDefined*, *Bee Dictionary*, *Memidex*, and *Wordnik* all draw data from the English WordNet developed at Princeton University, and the online *Danish Dictionary* (Trap-Jensen 2010) offers searches for related words through DanNet, the Danish wordnet. Wordnets are developed by human teams with reference to dictionaries produced by human lexicographers, but the information that e-dictionaries draw from wordnet sources is not manually edited. Trap-Jensen (2010) compares the thesaurus function in the online *Danish Dictionary* with that of the *Macmillan Online Dictionary*, which is compiled by lexicographers. He finds strengths and weaknesses in both systems. A wordnet can generate too many options, whereas manual editing can select those candidates which are most likely to be relevant to the user. On the other hand wordnets can identify relevant semantic links which cut across categories in a traditional manually-constructed thesaurus.

Language problems can also be addressed by automatically interrogating web corpora, without human intervention at any stage in the process. Whilst it might not be possible to derive old-style dictionary definitions directly from corpus data, algorithms are being developed to identify definition-like explanatory sequences within large collections of text. Dictionary sites such as *Wordnik* mine corpora for sequences which provide information about word meaning, rather than simply examples of word use (McKean 2011). These sequences are imported to *Wordnik* in place of definitions, as the *Wordnik* team do not define words themselves, and do not accept definitions contributed by users.

Automatically extracted explanations will be adequate in some consultation contexts but not in others, depending on the task and levels of user expertise and language knowledge. McKean (2011) admits that data mining techniques are not

useful as a means of uncovering word etymology, for example, because of the unreliability of the folk etymologies to be found in non-specialist texts.

The inclusion of the latest new words may not greatly improve the usefulness of a dictionary, but is important from a marketing perspective, as Rundell and Kilgarriff (2011) note, and there is clearly a public appetite for information about words reflecting new phenomena and societal change. The web analytics site *Google Insights for Search* (www.google.com/insights/search/#) for June 2011 to June 2012 reveals that many of the most frequent searches for the term ‘English dictionary’ led to articles about the acceptance of new terms in well-known authoritative publications, for example ‘mumpreneur’ (Cameron 2011) in the *Collins English Dictionary*, and ‘LARPing’, ‘scratchiti’ (Taylor 2012), and ‘squeezed middle’ (Zafar 2011) in the *Oxford English Dictionary*. Nowadays a new word can be added to an e-dictionary in a fraction of the time it takes a lexicographical team to compile and publish a print dictionary entry. Rundell (2011) contrasts modern e-dictionary practices with the flow chart on the Oxford Dictionaries site (n.d.) showing the elaborate and increasingly outdated process by which a word is considered for admittance in an Oxford Dictionary. Human web-page editors can now by-pass much of this process whilst still setting criteria for word inclusion. The collaborative e-dictionary *Wiktionary*, for example, allows contributors to add new entries, but requires that neologisms should be attested through widespread use in ‘a well-known work’ or ‘permanently recorded media, conveying meaning, in at least three independent instances spanning at least a year’.

Direct links to news databases and social media sites also mean that new words can be identified and analyzed computationally, however, and rapidly incorporated into some e-dictionaries with little or no editorial intervention. This may

lead to tension between the demands for up-to-date content and reliability, of course. Systems for the automatic extraction of neologisms have to overcome many problems, as Halskov and Jarvad (2010) point out, because expressions that the software will flag as new may actually be transparent, transient and/or idiosyncratic, and therefore lexicographically insignificant. Without humans to assess the lexicographical worth of potential new entries, e-dictionaries can rapidly become populated with words and expressions that have no real currency. Problems resulting from the lack of human intervention are particularly noticeable in the less prestigious varieties of ‘alternative’ bilingual e-dictionary which are popular with English language learners in East Asia. Nesi (2012) identifies in such dictionaries archaisms and nonce formations that are not differentiated in any way from items essential to an English language learner. It is sad to think of users wasting their time memorizing vocabulary that is relatively worthless to them from a communicative perspective.

Some bilingual e-dictionary sites also supplement their dictionary entries with automatically generated illustrative text. For example the *Jin Shan Ci Ba*, an enormously influential e-dictionary in mainland China, works with the machine translation device *Jinshan Kuaiyi* (Nesi 2012). Mair (2007) blames this system for the production of ‘absurdly crude English mistranslations in bizarrely inappropriate contexts’. A further problem is the incidence of meaningless machine generated sentences, originally posted on internet discussion sites to fool filters into accepting spam messages, but now sometimes automatically incorporated in online dictionary sites. Postings to USENET, the Internet discussion system, are used to supplement entries in the *Doosan Dong-a Prime* English dictionary on the *Daum* South Korean web portal, for example. Some of the postings are genuine, but others, such as ‘What

did Francis arrive the cup before the dark coffee?’ are simply random sequences which make no sense (Nesi 2012: 367).

1.3. Digital migration and the fragmented dictionary market

Of course, automatic translation and the automatic extraction of lexicographical material are techniques that bring commercial benefit to the companies that run e-dictionary sites. They add to the size of the dictionaries on offer (size being a simplistic but common means of evaluating dictionary worth), and they give the impression that the dictionary material is up-to-date, and that the site is technologically advanced. Most significantly, automation reduces development costs, just as a web platform reduces the costs of production and distribution, which is why companies such as Google and Amazon have embraced the move from print to digital and have persuaded the public to follow suit - Amazon started to sell more e-books than print books in May 2011 (Krek 2011).

It seems that the print to digital migration has particularly affected reference materials. People typically consult maps, encyclopaedias, and dictionaries while they are doing something else, for example whilst driving, writing, reading, listening, or conversing, and under these conditions the electronic format can improve accessibility and ease of consultation (‘usability’ in Laufer and Kimmel’s terms, 1997: 362). Thus paper maps have largely given way to e-maps delivered via satellite navigation systems, with the result that old-style cartography companies have shrunk or closed down, while a new e-cartography industry has grown up (Parish 2004). The print edition of the *Encyclopedia Britannica* ceased production in 2010, and most dictionary publishers now accept that print-based dictionaries will also largely disappear as content migrates to e-dictionaries of various types.

Levine (2001) realised that digital migration was responsible for the decline in sales of encyclopaedias, but commented optimistically about the commercial future of dictionaries, at least English language ones: ‘a boom in English usage and commerce fostered by the World Wide Web.... seems to be having just the opposite effect on lexicography as it did on “encyclopédiography”’. By 2003, however, he noted that the American monolingual dictionary business was showing little growth. Electronic dictionaries had been marketed alongside print dictionaries, but this had not resulted in an increase in overall sales because e-dictionaries were being bought instead of hard-copy dictionaries, or were being bundled with print editions (Levine 2003). Esposito (2002) was even more gloomy about the outlook for commercial dictionary publishers. Given the increasing availability of free e-dictionaries, ‘all current attempts (except Microsoft’s) to put dictionaries into electronic form are nothing more than a limp attempt to extend the life of a failing business model’.

Some of the publishers’ responses to this situation echo McArthur’s predictions in 1998 regarding the *regionalization*, *localization*, and *electronicization* of tomorrow’s dictionaries. Esposito foresaw some possibility for growth in the niche markets for sophisticated or specialized lexicographical products, for example dictionaries of obscure languages and dialects. Kilgariff (2005) agreed that there was scope for the marketing of smaller, more specific products to a world-wide customer base: ‘dictionary publishing is undergoing the same transformation as many other markets with the advent of the internet: the market fractures, and where there were a small number of products selling to millions, there are now millions of products – selling far smaller numbers – to billions’. Similarly Rundell (2011) talks of ‘a more fragmented landscape’ moving towards functionally diverse products for many different types of user.

This fractured market may offer some hope for further developments in dictionary content, continuing the progress made in the last decades of the 20th century. It seems that the changing business environment favours technical rather than lexicographical innovation, with many new developments in interface design and automatic data extraction but fewer developments addressing language learners' information needs. Small but pedagogically innovative dictionary ventures are, however, being undertaken by academics financed through research funding. A number of such projects are described in the *Proceedings of eLex 2009* (Granger and Paquot, eds. 2010) and the *Proceedings of eLex 2011* (Kosem and Kosem, eds. 2011). They include the *Louvain EAP Dictionary*, a project at the Université Catholique de Louvain (Granger and Paquot 2010), *Lang Yeast*, a dictionary to help biologists writing in English being developed at the Université Paris Diderot (Volanschi and Kübler, 2010), and *DAELE*, a Spanish learner's dictionary from Pompeu Fabra University in Barcelona (Mahecha Mahecha and DeCesaris, 2011; Renau and Battaner, 2011). The developers of these dictionaries aim to apply linguistic and pedagogical theory to lexicographical problems, drawing on the capabilities of the latest technologies. They are free to explore the effects of new search routes and defining methods because they do not have to sell their products, but at the same time resources are limited, and progress is therefore often slow.

1.4. Revenue sources

In the new business environment a few prestige e- dictionaries such as the *Oxford English Dictionary* can be sold to universities and libraries, and a few niche e- dictionaries can be sold to individuals, but for the most part people expect to use e- dictionaries for free. It continues to be common practice for publishers to offer e-

dictionary material as a means of adding value to their other products. Buyers of the *Collins COBUILD Advanced Dictionary* in book form, for example, can use *myCOBUILD*, an online version enhanced by the addition of specialist words. Similarly the *Longman Dictionary of Contemporary English Online* is a free online version of the CD-ROM, but users are urged to buy the full CD-ROM version to hear the pronunciation of 88,000 example sentences. Publishers may feel that they have to bundle electronic products in this way in order to maintain a competitive edge, but it is unclear whether it influences users' choice of dictionary title, or results in any additional revenue. Morse (2008) regards all free e-dictionary access as a form of bundling, but considers that 'so far, no bundled dictionary, whether with browser, search engine, operating system, or e-book reader yet looks likely to have a major impact on the dictionary business'.

It is also unclear whether the majority of print dictionary users really benefit from bundled e-content. Nowadays people continue to use print dictionaries in contexts where they do not have access to an electronic device, either because of school rules, or because of poor internet access or lack of equipment. Boonmoh and Nesi (2008), for example, surveyed 1,211 Thai university students who had been recommended by their teachers to buy the *Longman Active Study Dictionary* in book form. Only 28% of respondents claimed to own a monolingual dictionary on cd-rom, even though the *Longman Active Study Dictionary* cd-rom was included with the book, attached to the inside cover. Most of these students did not own computers, so the cd-rom was useless to them and was ignored.

Whilst e-dictionary publishers may not be paid directly by the end-user, they can derive revenue from licensing deals with manufacturers and commercial websites. Stand-alone pocket electronic dictionaries and web-based dictionary download sites

provide e-dictionaries from many sources, bilingual and monolingual, local and global (e.g. emanating from prestige publishing houses in Britain and the USA). Users may typically opt to consult the local bilingual source, but the inclusion of one or two prestigious sources adds credibility to the product and may help boost its sales. Oxford dictionaries are particularly highly regarded; the hardware companies AOnePro, Canon, Casio, Franklin, Seiko, Sharp, and Sony have all licensed content from Oxford University Press (Nesi 2008).

Stand-alone pocket electronic dictionaries are now being superseded by internet-enabled devices. Tuteja (2011), for example, reviews a recent Casio model released in India (the EW-B2000C) and wonders whether it is worth the price: ‘Can't I read (or even listen to) speeches on my internet-enabled smartphone or laptop? Can't I download dictionary apps, that too for free?’ He remains unconvinced of its value, although he concedes that the model ‘might be of some use to writers and students who don't have access to the Internet all the time, or find it a little bothersome to locate and launch an app on the phone or PC only to look up a definition’. In Europe interest in pocket electronic dictionaries may be growing, however, even as they lose ground in the Indian subcontinent and the Far East where they were originally most popular. In 2010-11 Casio Europe collaborated with researchers at the University of Osnabrück to conduct what they claim to be the first scientific study in Europe into the effects of pocket electronic dictionaries on learning (Ludewig et al. n.d.). A longitudinal project at the University of Wuppertal is also introducing pocket electronic dictionaries in a number of German schools, as a motivating alternative to dictionaries in book form (Diehr, n.d.).

A further source of revenue is the licensing of e-dictionary content for use with e-books and media websites, so that readers can access definitions of unknown

words directly from the text. The British version of Amazon Kindle comes with the *New Oxford American Dictionary* and the *Oxford Dictionary of English* pre-installed, for example, and the Indian online newspaper *DNA* (Daily News and Analysis) uses the *Macmillan English Dictionary* as a ‘plugin’ (Rundell 2011).

As predicted by de Schryver (2003) and Parish (2004), such applications mark a move away from the dictionary as a stand-alone product, and open the way to customization by licensees. This may involve adjusting layout and functionality, adapting content, and/or supplementing it with material dynamically generated from internet resources. As e-dictionaries are cut up and recompiled, content appears and disappears without warning and without trace. We lose standard bibliographical information such as the editor’s name and publication date, and we can no longer refer to a rationale and content overview of the kind traditionally provided in the front matter of a print dictionary.

All this has serious implications for dictionary evaluation and dictionary skills training. Customized e-dictionaries are difficult to review because of their instability and lack of front matter. In turn the lack of scholarly description and evaluation makes it difficult for teachers and students to interpret lexicographical content, and to choose which dictionary sites to use. Under these circumstances, perhaps the best kind of education for dictionary users is one which encourages a critical stance, and helps to dispel blind faith in the authority of all works entitled ‘dictionary’. Dictionary websites which invite users to comment and collaborate have the potential to support dictionary content evaluation, as can be seen from the active critical and scholarly discussion on the *Leo Dictionary* site, for example (Nesi 2012: 368-9).

Apart from selling dictionary content to the hardware manufacturers and to commercial websites, publishers can also sell advertising space on their own e-

dictionary pages. As Morse (2008) points out, ‘In the online world we don’t sell the dictionary; we sell the eyeballs that look at the dictionary’. Morse claims that the free Merriam Webster dictionary website attracts a great deal of advertising revenue, and Lannoy (2010) and Caruso (2011) both note that dictionary sites are attractive to advertisers because of the time and attention dedicated to dictionary consultation. Kilgarrieff (2009) is somewhat less optimistic, arguing that advertising works best for dictionaries based in the USA, and those which already have a strong brand name. He notes that in the UK, *Cambridge Dictionaries Online* have been the most successful in making money from advertising. Cambridge was an early adopter of the advertising strategy, however, and it could take time for publishers who started later to reap the same rewards, and to make enough to sustain a lexicographic team.

Rundell (2011) compares the *Macmillan* site, which at that time had relatively few advertisements (just a header and a column in the right margin of the webpage), to the noisier *Cambridge Dictionaries Online* where advertisements surround the definition on all sides. The noisier site may bring in more revenue, but the quieter site may be less distracting for the user. An informal test conducted by the blogger Marc Wandschneider (2010) seems to support this view. Wandschneider compared his own experiences of looking up the same words in four bilingual Chinese-English / English-Chinese dictionaries over a period of two or three weeks, and although he does not explicitly comment on the presence of advertising material, he states a preference for interfaces that are ‘clean’ and ‘spartan’, with ‘lots of blank space’. Dziemianko (2011) suggests that students might learn more from consulting an online dictionary that is advertisement-free. In studies comparing users’ retention of meaning and collocations, she found that those who referred to the online *Collins COBUILD Advanced Dictionary* (e-COBUILD6) (which did not have advertisements) had

significantly better scores than those who used the e-versions of the *Oxford Advanced Learner's Dictionary* (OALDCE7) and the *Longman Dictionary of Contemporary English* (LDOCE5) (which did).

Lannoy (2010: 174) proposes that publishers should establish business partnerships with companies selling complementary products or services on the web. Companies may be willing to pay e-dictionary publishers to perform an intermediary role, leading users from initial dictionary consultations to related products such as course books, novels, reference materials, news media, or learning materials.

Some sites, such as *Wordnik.com*, contain no overt advertising links but encourage users to sign up as members of an online community where they can tag words, create lists, and post comments. This kind of social networking activity has potential as an indicator of purchasing habits, and could lead to users being targeted as consumers at some later stage. Although members of such e-dictionary user groups are unlikely to view their personal data as having any commercial worth, according to the information on the *Wordnik* site the President of *Wordnik* takes a special entrepreneurial interest in 'opportunities in next-generation social commerce, community, crowdsourcing, and social media'. In these early days of social networking is impossible to gauge how such data might eventually affect e-dictionary use, design, and finances.

In the future it is conceivable that sensor technology could provide an even more sophisticated way of interacting with e-dictionary users. Films and games can now be customized to individual requirements by monitoring heart rate and sweat levels (see www.shimmer-research.com). The data is sent from a skin response sensor worn by the user to a smartphone app, which then adjusts on-screen content accordingly, in real time. In the same way, subtle changes in dictionary users'

physiology might one day trigger the provision of simpler or more complex dictionary definitions, more in-depth treatment of the look-up item, or links to activities related to the look-up item.

1.5. E-dictionaries and web presence: the way forward

In order to maintain a competitive edge, dictionary publishers now work to maximize their web presence. Lannoy (2010) recommends the use of search engine optimization strategies to improve traffic to e-dictionaries and increase the speed with which they deliver the information users desire. The entry point for many e-dictionary consultations is a search engine such as Google, Yahoo, or Bing, and these techniques can dramatically increase the number of general dictionary searches that lead to a specific site.

Moreover many potential dictionary users have backgrounds in languages other than English, and specific local needs. For example *Google Insights for Search* (www.google.com/insights/search/#) reveals that in the 12 months leading to June 2012 the greatest number of Google searches for ‘English dictionary’ came from Pakistan, Cambodia, Afghanistan, Bangladesh, Myanmar, Mongolia, India, and Nepal (Google does not operate in mainland China). Lannoy (2010: 180) concludes that ‘internet strategy needs to be designed market by market’ and advises publishers to localize, by providing bilingual content, and by translating their dictionary interfaces.

This advice perfectly echoes the thoughts of McArthur (1998), who foresaw the need for *localization*, *bilingualization*, *nationalization*, and *regionalization* to meet the requirements of the next generation of dictionary users, in the global e-market. The days of most authoritative, monolingual print dictionaries may be

numbered, but there are exciting opportunities ahead for dynamic, adaptive bi- and multi-lingual local e-dictionaries.

Dictionary references

a2zDefined www.a2zdefined.com

Allwords www.allwords.com

Bee Dictionary www.beedictionary.com

Cambridge Dictionaries Online <http://dictionary.cambridge.org>

Collins COBUILD Advanced Dictionary 6th edition (2009) HarperCollins [= e-COBUILD6]

Diccionario de Aprendizaje del Español como Lengua Extranjera www.DAELE.eu
[=DAELE]

Danish Dictionary (Den Danske Ordbog) (2003-2005) Ed. Lars Trap-Jensen. Danish Language Council <http://ordnet.dk/ddo>

Doosan Dong-a Prime on the Daum portal <http://engdic.daum.net>

Jin Shan Ci Ba (Kingsoft Powerword) online <http://www.iciba.com>

Lang Yeast <http://ytat2.ijm.univ-paris-diderot.fr/LangYeast>

Leo Dictionary <http://www.leo.org>

Longman Active Study Dictionary. Ed. Della Summers (2004). Harlow: Pearson Education

Longman Dictionary of Contemporary English Online [= LDOCE5]
www.ldoceonline.com

Louvain EAP Dictionary www.uclouvain.be/en-322619.html

Macmillan Online Dictionary <http://www.macmillandictionary.com>

Memidex www.memidex.com

myCOBUILD www.mycobuild.com

New Oxford American Dictionary 3rd edition. (2010) Eds. Angus Stevenson and Christine A. Lindberg. Oxford University Press

Oxford Advanced Learner's Dictionary of Current English 7th edition. Eds. Michael Ashby, Colin McIntosh and Joanna Turnbull. Oxford University Press [= OALDCE7]

OneLook www.onelook.com

Oxford Dictionary of English 3rd edition. Eds. Catherine Soanes and Angus Stevenson (2010) Oxford University Press

Wordnik www.wordnik.com

References

Boonmoh, A. and Nesi, H. (2008). 'A survey of dictionary use by Thai university staff and students, with special reference to pocket electronic dictionaries.' *Horizontes de Lingüística Aplicada* 6 (2): 79-90.

Cameron, L. (2011). 'Mumpreneur leads Collins English Dictionary entries.' *The Independent*, 5 October 2011. Available online at <http://www.independent.co.uk/news/uk/home-news/mumpreneur-leads-collins-english-dictionary-entries-2365895.html>

Campoy Cubillo, M. C. (2002). 'General and specialised free online dictionaries.' *Teaching English with Technology*, 2 (3) 42-47

Caruso (2011). 'Online specialised dictionaries: a critical survey', in Kosem, I. and Kosem, K. (eds) *Electronic Lexicography in the 21st Century New Applications for*

New Users. Proceedings of eLex 2011, Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies. pp. 66-75

de Schryver, M. (2003) 'Lexicographers' Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography*, 16 (2) 143-199.

Diehr, B. (undated). 'Project Description: With MOBIDIC to better English.'

Available online at http://www.uni-wuppertal.de/fachbereiche/fachbereich_a/MOBIDIC/Project_description.html

Dziemianko, A. (2011). 'Why one and two do not make three – dictionary form revisited.' Paper presented at *Electronic Lexicography in the 21st Century New Applications for New Users*., Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies.

Esposito, J. (2002). 'Dictionaries, another Netscape?' *Kernerman Dictionary News*, Number 10. Available online at <http://kictionaries.com/newsletter/kdn10-1.html>

Fellbaum, C. (2005). 'WordNet(s)', in Brown, K. (Editor-in-Chief) *Encyclopedia of Language and Linguistics*, Second Edition. Oxford: Elsevier. pp 665-670

Granger, S. and M. Paquot (eds.) (2010). *eLexicography in the 21st Century: New Challenges, New Applications*, Proceedings of eLex 2009, Louvain-la-Neuve, 22-14 October 2009. Louvain: Presses Universitaires de Louvain.

Granger, S. and M. Paquot (2010). 'Customising a general EAP dictionary to meet learner needs', in S. Granger and M. Paquot (eds.) *eLexicography in the 21st Century*:

New Challenges, New Applications, Proceedings of eLex 2009, Louvain-la-Neuve, 22-14 October 2009. Louvain: Presses Universitaires de Louvain. pp. 87-96.

Grefensgtette, G. (1998). 'The future of linguistics and lexicographers: will there be lexicographers in the year 3000?', in: Fontanelle, T. (eds) *Proceedings of the Eighth EURALEX Congress*. Liege: University of Liege. pp. 25-41.

Hall, G. (2008). *Digitize This Book! The Politics of New Media, or Why We Need Open Access Now*. Minneapolis: University of Minnesota Press.

Halskov, J. and Jarvad, P. (2010) 'Automated extraction of neologisms for lexicography', in S. Granger and M. Paquot (eds.) *eLexicography in the 21st Century: New Challenges, New Applications*, Proceedings of eLex 2009, Louvain-la-Neuve, 22-14 October 2009. Louvain: Presses Universitaires de Louvain. pp. 405-410

Kilgarriff, A. (2005). 'If dictionaries are free, who will buy them?' *Kernerman Dictionary News*, Number 13. Available online at <http://kictionaries.com/kdn/kdn1307.html>

Kilgarriff, A. (1992). *Ploysemy*. DPhil thesis, University of Sussex.

Koplenig, A. (2011). 'Understanding How Users Evaluate Innovative Features of Online Dictionaries – An Experimental Approach', in: Kosem, I. and Kosem, K. (eds) (2011) *Electronic Lexicography in the 21st Century New Applications for New Users*. Proceedings of eLex 2011, Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies. pp. 147-150

Kosem, I. and Kosem, K. (eds) (2011). *Electronic Lexicography in the 21st Century New Applications for New Users*. Proceedings of eLex 2011, Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies.

Krek, S. (2011). Round table talk at *Electronic lexicography in the 21st century: new applications for new users (eLex2011)*. published: Dec. 2, 2011, recorded: November 2011. Available online at http://videolectures.net/elex2011_krek_table/

Laufer, Batia. and Kimmel, Michal. (1997). 'Bilingualised dictionaries: how learners really use them.' *System* 25 (3) 361-369

Lannoy, V. (2010). 'Free on-line dictionaries: why and how?', in S. Granger and M. Paquot (eds.) *eLexicography in the 21st Century: New Challenges, New Applications*, Proceedings of eLex 2009, Louvain-la-Neuve, 22-14 October 2009. Louvain: Presses Universitaires de Louvain. pp. 173-182.

Leroyer, P. (2011). 'Change of paradigm: from linguistics to information science and from dictionaries to lexicographic information tools', in Fuertes-Olivera, P.A. and Bergenholtz, H. (eds) *E-Lexicography: The Internet, Digital Initiatives and Lexicography*. London: Continuum. pp. 121-140.

Levine, C. (2001). 'The coming boom in English lexicography: Some thoughts about the World Wide Web (Part One)'. *Kernerman Dictionary News*, Number 9. Available online at <http://kdictionary.com/newsletter/kdn9-1.html>

Li, L. (2011). 'Neologisms: Challenge to EFL learning and EFL dictionaries', in: Kaoru Akasu and Satoru Uchida (eds.) *Lexicography: Theoretical and Practical Perspectives*. Proceedings of Asialex 2011 Kyoto 303-312

Ludewig, P., Evert, S. and Gust, H. (n.d.) *Die CASIO-Wörterbuchstudie*. Universität Osnabrück. Available online at www.ex-word.de/global/downloads/CASIO-Woerterbuchstudie.pdf

Mahecha Mahecha, V. and DeCesaris, J. (2011). 'Representing nouns in the Diccionario de aprendizaje del español como lengua extranjera (DAELE)', in Kosem, I. and Kosem, K. (eds) (2011) *Electronic Lexicography in the 21st Century New Applications for New Users*. Proceedings of eLex 2011, Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies.

Mair, V. (2007). 'The etiology and elaboration of a flagrant mistranslation.' *Language Log*. Available online at <http://itre.cis.upenn.edu/~myl/language-log/archives/005195.html> (last accessed April 29 2011)

McArthur, T. (1998). 'Guides to tomorrow's English.' *English Today* 14:3 21-26

McKean, E. (2011). 'Wordnik: Notes from an online dictionary project'.. Paper presented at *Electronic lexicography in the 21st century: new applications for new users (eLex2011)*. published: Dec. 2, 2011, recorded: November 2011. Available online at http://videolectures.net/elex2011_mckean_notes/

Meyer, C. M. and Gurevych, I. (2012) Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography, in S. Granger & M. Paquot (eds.) *Electronic lexicography*. Oxford: Oxford University Press.

Morse, J. (2008). 'Merriam-Webster and the future of dictionary-making.'

Kernerman Dictionary News, Number 16. Available online at

<http://kictionaries.com/kdn/kdn16/kdn1601morse.html>

Müller-Spitzer, C., Koplenig, A. and Töpel, A. (2011). 'What makes a good online dictionary? – Empirical insights from an interdisciplinary research project', in Kosem, I. and Kosem, K. (eds.) *Electronic Lexicography in the 21st Century New Applications for New Users*. Proceedings of eLex 2011, Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies. pp. 203-208.

Nesi, H. (2012). 'Alternative e-dictionaries: uncovering dark practices', in S. Granger & M. Paquot (eds.) *Electronic lexicography*. Oxford: Oxford University Press.

Nesi, H. (2008). 'Dictionaries in electronic form', in Cowie, A. P. (ed) *The Oxford History of English Lexicography*. Oxford: Oxford University Press 458-478

Nesi, H. (1998). 'Dictionaries on computer: how different markets have created different products', in Wilson, A. and Schmied, J. (eds) *Proceedings of the Chemnitz Symposium*. 20-21 February 1998, University of Chemnitz. Available online at <http://www.tu-chemnitz.de/phil/english/chairs/linguist/real/independent/llc/Conference1998/titel.htm>

Oxford Dictionaries (n.d.). 'Have you ever wondered how a word earns its place in an Oxford Dictionary?' Available online at

<http://oxforddictionaries.com/page/newwordinfographic/how-a-new-word-enters-an-oxford-dictionary>

Parish, J. (2004). 'Microsoft and dictionary publishers: Defining partnerships.'

Kernerman Dictionary News, Number 12. Available online at

<http://kdictionary.com/kdn/kdn12-1.html>

Renau, I. and Battaner, P. (2011). 'The Spanish learner's dictionary DAELE on the panorama of the Spanish e-lexicography', in Kosem, I. and Kosem, K. (eds)

Electronic Lexicography in the 21st Century New Applications for New Users.

Proceedings of eLex 2011, Bled, 10-12 November 2011. Trojina: Institute for Applied Slovene Studies. pp. 221-6.

Rundell, M. (2011). 'Many a mickle maks a muckle.' Round table talk at *Electronic lexicography in the 21st century: new applications for new users (eLex2011)*

published: Dec. 2, 2011, recorded: November 2011. Available online at

http://videlectures.net/elex2011_rundell_table/

Rundell, M. (2009). 'The road to automated lexicography: first banish the drudgery....

Then the drudges?' Paper presented at *eLexicography in the 21st century. New*

challenges, new applications, Louvain-la-Neuve 22-24 October 2009. Centre for English Corpus Linguistics

Rundell, M. and Kilgarriff, A. (2011). Automating the creation of dictionaries. Where will it all end? in Meunier, F., de Cock, S., Gilquin, G., and M. Paquot (eds) *A Taste for Corpora. In honour of Sylviane Granger*. Amsterdam: John Benjamins. pp. 257–282

Taylor, L. C. (2012). ‘LARPing’, ‘scratchiti’ declared words by the Oxford English Dictionary.’ *The Star*, April 3 2012. Available online at <http://www.thestar.com/news/world/article/1155340--larping-scratchiti-declared-words-by-the-oxford-english-dictionary>

Trap-Jensen, L. (2010). ‘One, Two, Many: Customization and User Profiles in Internet Dictionaries’, in Dykstra, A. and Schoonheim, T. (eds.) *Proceedings of the Fourteenth EURALEX International Congress*, Leeuwarden, Netherlands, 6 - 10 July 2010. Fryske Akademy NL. pp. 1133-1143

Tuteja, A. (2011). ‘Review: Do we really need the Casio EW-B2000C e-dictionary?’ *IBNLive*.. Posted on Nov 28, 2011. Available online at <http://ibnlive.in.com/news/review-casio-ewb2000c-edictionary/206622-11.html>

Volanschi, A., and Kübler, N. (2010). ‘Building an electronic combinatory dictionary as a writing aid tool for researchers in biology’, in S. Granger and M. Paquot (eds.) *eLexicography in the 21st Century: New Challenges, New Applications*, Proceedings of eLex 2009, Louvain-la-Neuve, 22-14 October 2009. Louvain: Presses Universitaires de Louvain. pp. 343-54.

Wandschneider, M. (2010). 'The Great Chinese Online Dictionary Showdown.'

Available online at <http://chipmunkninja.com/The-Great-Chinese-Online-Dictionary-30@>. Last accessed 15 June 2012.

Zafar, A. (2011). 'Oxford English Dictionary picks 'Squeezed Middle' as Word of the Year.' *Time NewsFeed* November 28 2011. Available online at

<http://newsfeed.time.com/2011/11/28/oxford-english-dictionary-picks-squeezed-middle-as-word-of-the-year/>