Redefining the dictionary: From print to digital

Michael Rundell

Last November, Macmillan Dictionaries announced that it was abandoning the print medium, and would henceforth publish dictionaries in digital formats only[1]. Around the same time, I heard a great story from my friend Jim Ronald, a professor in English linguistics working in Japan. Jim had taken a set of (printed) learner's dictionaries into a class, and noticed one of his students picking up a dictionary and nostalgically leafing through it, before declaring "Ah, this brings back memories!" Two months earlier, when Jonathon Green wrote a critical piece about crowdsourced dictionaries in *The Guardian*, one (British) reader added a comment saying:

> "The three things no young person owns or uses and often don't realise exist: an alarm clock, an address book and a dictionary. … At university i didn't meet a single person who owned any of them"[2].

Anecdotal evidence, yes, but what both stories suggest is that, for younger people living in developed economies, the print dictionary is already history. This should come as no surprise: most people currently entering higher education are effectively digital natives, and for their general reference needs, the Web will always be the first (and, usually, only) port of call.

A very different attitude towards the physical book can be seen in this review of a new edition of the *American Heritage Dictionary* from 2011: "I confess I still get a psychic satisfaction from fumbling with a balky dust jacket wrapped around a real 'live' book, while taking in that distinctive new-book fragrance, and experiencing the subtle, yet futile resistance of the book spine on its very first opening".[3] This touching display of bibliophilia may strike a chord with readers of a certain age. But for most people a dictionary is a practical tool for resolving immediate communicative problems, and as such, a dictionary accessed on a computer or mobile device has huge advantages over its analogue predecessors.

This is not to say that the migration of dictionary content from print to digital media has met with universal approval. Macmillan's announcement sparked a lively debate (notably on the Euralex discussion list) on the pros and cons of digital dictionaries, and there were plenty of dissenting voices. The tone was occasionally elegiac: "a sad day for dictionaries", and similar sentiments. Much of this is pure nostalgia (rather like mourning the passing of steam trains), but two recurrent concerns deserve to be addressed. First, the idea that an online dictionary can't match the "browseability" of a printed one – where you can skip from entry to entry or from page to page, making serendipitous discoveries. There isn't much substance to this argument. In many online dictionaries, every word in an entry (including inflected forms), whether in a definition or example sentence, is hyperlinked to its *own* entry. Many also have some kind of 'related words' panel, typically listing compounds or phrases that include the word you are looking up. Thus the entry for *dog* in the Macmillan Dictionary provides links to items such as *hot dog, top dog, dog-eared, dog eat dog*, and *you can't teach an old dog new tricks*. Experience with Wikipedia suggests that, if people really want to while away their leisure hours leafing through works of reference, the digital medium provides abundant opportunities.

A more significant concern is the question of connectivity. Unless you are using a standalone dictionary app, you need to be connected to the Web to search an online dictionary. And it remains the case that there are many places where connections are slow, unstable, expensive, or non-existent. It quickly became clear, from exchanges on the Euralex forum, that connectivity isn't a simple case of rich-countries-connected, poor-countries-not. Somewhat to my surprise, Geoffrey Williams pointed out that many of the students at his French university didn't have internet access when they went back to their parental homes after a day at college. Conversely, David Joffe gave an upbeat assessment of the situation in Africa, whose mobile phone revolution – one of the most astonishing developments of the last two decades – is now being consolidated by infrastructure

Michael Rundell has been a lexicographer since 1980. He has edited numerous learner's dictionaries, and after working at both Longman and COBUILD, he became Editor-in-Chief of a new range of learner's dictionaries for Macmillan. He has published extensively in the field of corpus-based lexicography, and is co-author (with Sue Atkins) of the *Oxford Guide to Practical Lexicography* (2008). With Sue Atkins and Adam Kilgarriff, he set up the Lexicom workshops in lexicography and lexical computing, which are now in their 13th year. In the last ten years, he has been at the forefront of applying computational technologies to the development of dictionaries. Michael's career is bookended by two major lexicographic revolutions: the arrival of corpora in the 1980s and – more recently – the transfer of reference resources from print to digital media, a process in which Macmillan has been a leading player. With this second revolution still unfolding, he is engaged in exploring the opportunities it offers and pondering its implications for the principles and practice of dictionary-making. http://lexmasterclass.com/people/michael-rundell/

[1.] http://macmillandictionaryblog.com/bye-print-dictionary/

[2.] http://guardian.co.uk/books/booksblog/2012/sep/13/dictionaries-democratic-crowdsourcing/

[3.] http://weblogs.baltimoresun.com/news/mcintyre/blog/2011/09/infatuated_with_a_book.html/

Macmillan Dictionary

The *Macmillan English Dictionary for Advanced Learners* was first published in 2002, and a second edition followed in 2007. A free online version was launched in February 2009, and in November 2012, Macmillan announced its intention to phase out printed dictionaries and focus entirely on digital publishing.

As well as providing the traditional learner's dictionary fare (definitions, example sentences, information on syntactic, collocational and text-type preferences, and so on), the online *Macmillan Dictionary* has a range of other components. Thesauruses are a common feature of online dictionary sites, and the one in Macmillan is fully integrated – with a link from every word, word-sense, or phrase in the dictionary.

Like many online dictionaries, Macmillan's engages with its users through social media, competitions, and the inclusion of "user-generated content" (UGC). The site additionally includes a games area (with a growing suite of language-learning games), a set of videos dealing with dictionaries and language change, a weekly column (*Buzzwords*) providing in-depth discussion of notable new words and meanings, and a blog (http://macmillandictionaryblog.com/). The blog posts four or five new articles every week, on topics such as language change, world Englishes, common learner errors, language technology, and metaphor.

http://macmillandictionary.com

improvements which will provide fast and affordable Web connections for increasing numbers.

But the direction of travel is clear. Eurostat reports that 76% of households in the EU27 (EU countries and a few neighbours such as Norway and Turkey) have access to the internet - and that was a year ago.[4] (In 2004, the figure was 41%.) In parts of east Asia, the percentage is much higher. We are, admittedly, still in a transitional phase, but the trend is unstoppable, and in deciding to focus only on digital dictionaries, Macmillan was merely anticipating a move that all dictionary publishers will have to make eventually (and probably sooner than most people think).

The benefits of moving from print to digital have been well-rehearsed, and don't need to be discussed in detail here. Several posts on the Macmillan Dictionary blog have been devoted to this topic[5], and in one of these Adam Kilgarriff described Macmillan's decision as "A day of liberation from the straitjacket of print". The fact is that printed books are not a very efficient medium for reference materials. Space constraints have made the dictionary a miracle of compression, as huge amounts of information are shoehorned into a limited space. Many lexicographic conventions – the abbreviations and tildes, the compressed defining styles, the truncated examples – can be seen as devices for maximizing the amount of data that will fit within the covers of a book. But it all comes at a cost: how well is the user served, for example, when, *expectant* is defined as "characterized by expectation", and *expectation* as "the act or state of expecting" (*Merriam-Webster's 11th Collegiate*, 2003)?

The corpus revolution gave us the tools and the data to provide a far richer account of word behaviour than was previously possible, but this has left printed books bursting at the seams. My old copy of the third edition of the *Oxford Advanced Learner's Dictionary* (1974) is small and portable, with just over 1,000 pages. The latest crop of learner's dictionaries come in a larger format, contain around 2,000 pages, weigh a ton, and are bundled with CD-ROMs (another ageing technology) to accommodate overspill data. This can't continue – and fortunately it doesn't need to.

Apart from space, other obvious benefits include hyperlinking, multimedia (providing audio pronunciations, animations, and games, for example) and the potential for regular updating. The old print model saw new editions coming out perhaps once every five years, leaving enormous gaps in the record. Macmillan now has several updates every year.

We are just beginning to grasp the possibilities of the medium, but the implications of going digital will be wide-ranging. For instance, traditional criteria for inclusion (which words get into the dictionary) were partly determined by how many pages your dictionary had. This in turn contributed to the dictionary's perceived role as "gatekeeper", because the imperative of keeping a lot of vocabulary *out* encouraged the popular view that admitting a word to a dictionary conferred some special status on it.

Many of these rules no longer apply, but we are still working out what to replace them with. What contribution, for example, could be made by crowdsourcing? Some experiments in this area have been less than impressive – the *Urban Dictionary*[6] being an egregious example (notwithstanding its value as entertainment). But there are plenty of counterexamples. *Wiktionary*[7] continues to grow, as subject-specialists add headwords or translations for terms in their own fields. Macmillan has its own crowdsourced dictionary (the *Open Dictionary*[8]), which already includes almost 2,000 items sent in by users from all over the world. We have found that this model works particularly well for "long tail" items like neologisms, regional usages, and technical terms. Typically these have only one meaning, so they don't require the kind of lexicographic skills you would need when compiling an entry for *set* or *place*. Some of this material is ephemeral, but some takes its place in the language, in which case we "promote" it to the main dictionary (and edit as necessary). Crowdsourcing has great potential, but to exploit this fully we need to develop clear guidelines and provide contributors with foolproof templates.

This is one of many areas of lexicography where we are still feeling our way during this exciting period of transition. The business model is another. The question we're all asking is: is it possible for dictionary publishers to make money if they don't sell books? The long-term viability

4. http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tin00134/

5. http://macmillandictionaryblog.com/no-more-print-dictionaries/

6. http://urbandictionary.com/

7. http://wiktionary.com/

8. http://macmillandictionary.com/open-dictionary/

of dictionary publishing is an issue that was debated in *KDN* over a decade ago. Writing in these pages, Joseph Esposito offered a gloomy vision for traditional dictionary publishers, who he saw being outflanked by Microsoft, adding: "In the absence of growth, the old business will be strained for capital, which will beget smaller investments, which will in turn hasten the decline" (2002). It's still too early to say how accurate this prediction will turn out to be, but – though some players will not survive – there are reasons for cautious optimism.

On the development side, technology is helping to drive down costs. Acquiring corpus data used to be a major expense, but billion-word Web corpora can now be assembled for a fraction of what it cost to create the BNC 20 years ago. Meanwhile, significant progress has been made in automating editorial processes such as extracting relevant information from corpora, selecting example sentences, and checking text quality (cf. Rundell and Kilgarriff 2011, Rundell 2012). In prospect now is a dictionary-compilation model where "the software selects what it believes to be relevant data and actually populates the appropriate fields in the dictionary database" (Rundell and Kilgarriff 2011.278), so that the whole process is streamlined…and therefore costs a lot less. Crowdsourcing (mentioned earlier) could also – if well managed – have a part to play in keeping a lid on editorial costs.

On the publishing side, several possible revenue streams are already in the frame, and others that we can't yet imagine will no doubt emerge. Apps, APIs, and licence deals to provide dictionary services to third-parties can all contribute. But it's a fluid situation, and there is bound to be a lot of trial and error before a robust business model takes shape. When we launched the online Macmillan Dictionary in 2009, for example, there was a debate about whether to adopt the so-called freemium model, keeping the more valuable content behind a paywall. Our conclusion was that for general reference this wasn't going to work (just as it doesn't work for general news: users have too many other options).

Over the four years that the Macmillan Dictionary has been online, the landscape has changed, and there are many more competitors out there. Despite this, our commitment to continuous improvements to every aspect of the site, including its look-and-feel, functionality, content, and currency. has paid off in terms of steadily growing traffic, and hence significantly improved advertising revenue. SEO (search-engine optimization) has an important role in attracting visitors to the site, but it is not the critical factor. Contrary to the way things looked a few years back, we're now increasingly convinced that appealing, relevant, high-quality content is what really draws users to the site and encourages them to come back. And "content" now means much more than a traditional defining dictionary.

Given the abundant corpus resources and powerful software now at our disposal, the opportunities offered by digital media are unlimited – and only just beginning to be explored. In this sense, it's an exciting scenario. At the same time, commercial dictionary publishers find themselves operating in a challenging and often uncomfortable environment. We used to know who our competitors were (and there weren't very many of them), but we now compete for attention with numerous online dictionaries (of wildly varying quality), automatic translation sites, language forums, text-remediation devices, and other resources. Publishers have to remain alert (you never know who is going to appear from nowhere to eat your lunch) and be aware that the environment can change rapidly – as shown, for example, by the dramatic growth in mobile devices, so that dictionaries are now more likely to be accessed from a smartphone or tablet than from a desktop computer. Just as we get used to the idea that "the dictionary" is no longer a printed book, we have to face the possibility that dictionaries will not survive at all in the longer term – at least, not as the autonomous entities they are now. It is just as likely that they will be embedded in other resources. But that is for another day. What is clear is that the migration of reference resources from print to digital media is going to be an even bigger game-changer than the arrival of corpora in the 1980s.

References

Esposito, Joseph. 2002. Dictionaries, another Netscape?, *Kernerman Dictionary News* 10.

Rundell, Michael and Kilgarriff, Adam. 2011. Automating the creation of dictionaries: where will it all end?. In Meunier, F., De Cock, S., Gilquin, G. and Paquot, M. (eds.), *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Amsterdam: Benjamins. 257-281.

Rundell, Michael. 2012. The road to automated lexicography: an editor's viewpoint. In Granger, S. and Paquot, M. (eds.), *Electronic Lexicography*. Oxford: Oxford University Press. 15-30.

Dictionary *n. Obsolete*?
Before and afterwords
Ilan Kernerman, 2012

Trends
- from print to digital
- from tangible to virtual (to invisible)
- from one dictionary for life to many simultaneously
- from one-size-fits-all to customized and personalized
- from a language product to (multi-)language services
- from paid to non-paid (to paid)
- from private(ly-owned) to public(ly-funded/ shared)
- from (passive) reader to (interactive) user
- from old to young
- from human to machine
- from content to technology
- from words (to phrases and structures) to language
- from dictionaries to lexicographicology