

The future of dictionaries

Judy Pearsall



Judy Pearsall is Editorial Director, Dictionaries, at Oxford University Press, where she is responsible for the editorial programme of the *Oxford English Dictionary* as well as that of bilingual dictionaries and lexical content across a large range of languages. She has a background in English lexicography and was Co-Editor, with Patrick Hanks, of the first edition of the *New Oxford Dictionary of English*, published in 1998.
judy.pearsall@oup.com

Dictionaries of one sort or another have been around for many centuries and longer. For the English language, the traditional starting point is Robert Cawdrey's *A Table Alphabeticall* (1604) but, from a world perspective, the antecedents are considerably older. In fact, dictionaries go back several millennia perhaps to the early use of writing itself, and certainly to the early civilizations of the Akkadian, Babylonian, and Greek Empires. In the modern context, dictionaries – understood here broadly as a book containing a list of words in a particular language with definitions or translations (in another language), designed to help with understanding or using the language(s) in question – are found in more or less every written language in the world. For dictionaries to have lasted within so many human cultures for so long, it is surely not presumptuous to suggest that dictionaries must fulfil some essential human need, and that therefore, by implication, their future is secure.

And yet, from the perspective of the early twenty-first century, it is hard not to wonder aloud about the future of dictionaries, and whether it is conceivable they can survive not a few more millennia, but even another half century. This article considers this question of the Future of Dictionaries, why it is a question worth asking, and what those who curate and are involved with lexical content might do next.

The signs that dictionaries are under threat are all around us. The sale of print dictionaries is declining in all but the developing markets of the world; between 2007 and 2012 the total UK market declined by around 15%¹, and this was played out to a greater or lesser extent in developed markets including the United States, western Europe, Japan, Canada, and Australia. Respected and long-standing dictionary publishers, such as Chambers in the UK and Random House in the US, have all but disappeared while others, such as Langenscheidt in Germany, have greatly reduced their operations. While sales of print dictionaries are still growing in developing nations, for example in India and large parts of Africa, this is perhaps only because the relative wealth of these parts of the world means that the technology revolution affecting the rest of us is lagging behind.

All of this can, of course, be seen through the lens of publishing in general, as explained by a shift in format and medium: print books may be in decline, but that doesn't mean that people are not reading, or using the printed word. In fact, the opposite seems to be the case. The last decade has seen the explosion of new written forms (blogs, social media) and the internet's embrace of new writers of all kinds (via formal and informal self-publishing), while e-books are one of the fastest growing areas of the digital market, driven by the massive growth in ownership of e-readers and tablets in the last 3 years.

It is certainly true that dictionary publishers and others have boldly attempted to adapt to these changes by shifting formats and means of delivery, with some success. There are many dictionaries available online, especially free ones, and some of the English language-based ones, such as Dictionary.com and TheFreeDictionary.com, are reaching many millions of users and making good income from advertising revenue. There are thousands of dictionary apps available for smartphones and tablets and developed for a range of mobile platforms such as iOS and Android. In November 2012, the British-based ELT publisher Macmillan announced that it would no longer publish print dictionaries, bravely citing this as “a cause for celebration” rather than concern, with Stephen Bullon, Macmillan Education's Publisher for Dictionaries, confidently asserting: “[T]he message is clear and unambiguous: the future of the dictionary is digital.”

There is no doubting these successes but, in the midst of such fundamental external changes, it is hard not to read these current activities as merely a shoring up of current business models, rather than a positive leap into the future. To put it bluntly, many digital dictionaries are free and most of them are cheap. A combination of disintermediation and freely available digital resources means that, despite a few exceptional cases such as OED Online, the large sales to libraries of high-end dictionaries are not being replaced like for like by digital sales. Moreover, the fundamental idea of the dictionary, as a standalone volume encapsulating various types of information for individual words via the format of a dictionary entry (definition, orthography, morphology,

¹. Source: Nielsen Bookscan proprietary data for the UK book trade 2007, 2008, 2009, 2010, 2011, 2012.

usage, related words, etc.), remains basically untouched. Many apps simply display the dictionary entry organized much as in a print book, with maybe a few colours added and abbreviations removed. If digital dictionaries are the future, this doesn't look very exciting.

On the other hand, seemingly paradoxically, 'language' is big business. The English language is booming and, depending on which forecast you read, the number of English learners is expected to double in the next 15 years. The effects of globalization and growth in technology mean language technology is more and more in demand and there is a profusion of translation services, apps, and language learning packages. While English is the most widely used as a second or foreign language, other languages, particularly Chinese, Arabic, Spanish, are growing in prominence on the world stage. But it's much more than this, too. Language technology is invisibly all around us facilitating us in the most usual of tasks: every time we send a text, every time we search online, every time we receive 'suggested links' while shopping online or reading a blog, every time we use autocorrect functions in an email or a document, every time we use any aggregating service for news or other information, every time we use voice-activated technology. Businesses and technologies such as web advertising, spam filtering and parental control, sentiment analysis, and content management are all powered by language technologies at all but the most basic levels. The list could go on and on.

So, if the language business is booming while dictionary publishers face an uncertain future, what does this mean and what can dictionary makers learn as a result?

In 1960, a Harvard Business Professor called Theodore Levitt published an influential article called 'Marketing Myopia'². His central point was that companies paid too much attention to producing products and too little to satisfying customer needs. One of the practical examples he gave was drills. But what he said was this: 'People who buy drills don't need drills; they need holes.' His point was to illustrate the importance of focusing on the need rather than on products or features. His further point was that businesses continue to focus on the product at their peril – because if someone else thinks beyond the product and finds a better way to fulfil the need, then the existing business is in trouble.

This is important for dictionary publishers,

because it's easy to take for granted that they are in the Dictionary business, and to think that the product is dictionaries. But, really, dictionary publishers are in the business of fulfilling a need, that of providing resources to help users to understand or use language. It just so happens that dictionaries have been a good vehicle in the past to fulfil the need. But this may not be true for the future, or for the present. If we focus on what is needed, where it is needed, who needs it, and why it is needed, we might well redefine what we are doing and even dispense with the idea of the Dictionary itself along the way.

To do this means going to the very heart of what we do – the content – and coming up with a content strategy that exploits lexical content to the full, making it always relevant and useful. It is interesting to note that the term 'content strategy' is quite recent: it emerged as the term to describe the technologies, methods, and systems (such as SEO, Content Management Systems, metadata) that were deployed from the 1990s to deal with the mass of 'content' on the web. Content strategy is based on the deceptively simple idea that content must be created, managed, and disseminated according to such criteria as relevance (to people), usefulness (to machines), reusability, efficiency, sustainability, and comprehensiveness. A good content strategy demands a relatively stable core content hub that is flexible, reusable, connected, sustainable, and efficient. Whatever is going on externally – new digital trends, changing user behaviour, emergence of new technologies and new markets – the content is sufficiently robust and flexible to adjust itself to the changing needs.

The principle of a new content strategy can be applied to rethinking traditional dictionary content. Traditional dictionary formats store many different types of information, much of it elliptically. They are typically designed to serve multiple needs: to help with both decoding and encoding, as well as including information on usage, pronunciation, related words, derived words, collocates, and so on. Users need a certain amount of expertise and familiarity to navigate an entry quickly, passing over information they do not need to get to the information they do need. The dictionary may be accurate and coherent in its own terms, but as a means of providing the information needed at the moment of use, it is just not very efficient. And while digital access may help, this doesn't really get around the central issue.

Moreover, today we need to think not only about the human user, but beyond, to the machine as user, and beyond that to uses we don't even know. The dictionary

Oxford Dictionaries

Oxford University Press has been involved in producing dictionaries for around 150 years; in 1879, James Murray was appointed as the first Editor of the *Oxford English Dictionary*, a dictionary on historical principles whose mission was (and continues to be) to research and record the entire history of the lexicon of the English language. Throughout the twentieth century, Oxford Dictionaries grew from a largely British-based programme to a global programme across English and bilingual titles published in more than 20 countries, including titles such as the *Oxford Advanced Learner's Dictionary* (first published 1948) and the *Oxford Dictionary of English* (1998). Increasingly, the focus of the Oxford programme is now moving beyond traditional dictionary formats and is on using language technology to develop ever richer and more useful lexical data for a wide variety of language solutions, services, and products.

<http://oed.com/>

<http://oxforddictionaries.com/>

² Theodore Levitt (1960), 'Marketing Myopia'. *Harvard Business Review*.

Charles M. Levine.
The Coming Boom in
English Lexicography:
Some Thoughts about the
World Wide Web
(Part One)

*Kernerman Dictionary
News* 9, July 2001
[http://kdictionaries.com/
kdn/kdn9-1.html/](http://kdictionaries.com/kdn/kdn9-1.html/)

Joseph E. Esposito.
Dictionaries, another
Netscape?

*Kernerman Dictionary
News* 10, July 2002
[http://kdictionaries.com/
kdn/kdn10-1.html/](http://kdictionaries.com/kdn/kdn10-1.html/)

Charles M. Levine.
The Coming Boom in
English Lexicography –
Reconsidered (Part Two)
*Kernerman Dictionary
News* 11, July 2003
[http://kdictionaries.com/
kdn/kdn11-1.html/](http://kdictionaries.com/kdn/kdn11-1.html/)

Julian Parish.
Microsoft and Dictionary
Makers: Defining
Partnerships
*Kernerman Dictionary
News* 12, July 2004
[http://kdictionaries.com/
kdn/kdn12-1.html/](http://kdictionaries.com/kdn/kdn12-1.html/)

Adam Kilgarriff.
If dictionaries are free,
who will buy them?
*Kernerman Dictionary
News* 13, June 2005
[http://kdictionaries.com/
kdn/kdn1307.html/](http://kdictionaries.com/kdn/kdn1307.html/)

John M. Morse.
Merriam-Webster and
the future of
dictionary-making
*Kernerman Dictionary
News* 16, July 2008
[http://kdictionaries.com/
kdn/kdn16/kdn1601
morse.html/](http://kdictionaries.com/kdn/kdn16/kdn1601morse.html/)

format is limited in that it is essentially fixed, standalone, and designed for human users. Inconsistencies between and even within texts abound. Information is deliberately missing, truncated, or implied in a way that may be acceptable (even desirable) for a human user but represents only incompleteness and failure to the computer as user. The reason, typically, is that many dictionary conventions arose due to a need to economize on space in a print volume. But in the digital age these print-driven conventions are not only unnecessary, they actively undermine the machine interface. Imagine trying to use a standard dictionary, without the addition of metadata, as the backdrop to look up words from an e-reader, or to find meaningful results from a search engine. For every time a so-called regular variant or morphological form is missing from the dataset, every time a variant is recorded in intractable syntax (*encyclopedic, be of one/be of the same mind (about)*), every time a definition follows a non-standard style, and every time there is variability in spelling, hyphenation, or capitalization, a gap in information, a failure to connect, a failure to deliver what is needed will be the outcome.

One way of expressing the transformation that is needed by the new content strategy is to consider it in terms of *quality*. Established organizations, businesses, and individuals which are under threat from new competitors or new models often cite quality as a factor differentiating themselves from newcomers. Publishers are certainly no exception. As a result, quality has become a kind of shibboleth of traditional publishers. But the same organizations may not question their notion of quality and how it might need to change according to customer behaviour and needs. Whatever its merits or demerits, they may not acknowledge that it is a new type of quality brought by the newcomers that is creating the disruption in the first place. Whereas quality of lexical content in a pre-digital age may have been measured principally in terms of local detail and accuracy, quality of lexical content for the digital age may be measured also by macro factors such as discoverability, speed, or availability, underpinned by full morphology, semantic metadata, breadth of coverage, and frequency information. In this context, user-generated content, adequately curated and differentiated from core content, can be a viable force for enhancing existing quality content, rather than being seen as merely a marketing strategy or judged as of dubious value. In the end, for the average user, being able to find some reasonable results quickly and at the point of use may be more important than

having a single perfect result that takes a long time to find, is located in a standalone application somewhere else, or, because of insufficient quality of morphology or metadata, is never found at all.

It is, of course, always easier to state the challenge than to articulate the opportunity. On the other hand, there is plenty to be positive about. Even if the Future of Dictionaries *qua* dictionaries is uncertain, it is clear that lexical information is very much in demand, especially by the new technologies. All of the uses mentioned in the paragraphs above – and many more – are underpinned by lexical information. There is an opportunity for those expert in handling lexical content to continue to do so, by developing a strong content strategy that can serve machine users as well as human users, and which can be focused on needs and finding solutions to those needs rather than improving the features or performance of existing products. While much present-day dictionary content is structured (as xml, for example) so that it can be processed by a machine, the new lexical content goes further; it is structured and semantically annotated in such a way that it can be read intelligently by a machine, and new products, links, or information are automatically produced as a result. This type of new content is segmented and modular so that each type of information is separable, while also linking into a central concept, allowing for specific needs to be addressed directly and efficiently. It starts with the simple hub, but supports the creation of a scalable language resource, allowing related types of content – text from corpora, taxonomies, and synonyms, for example – to be included or added incrementally as time and need dictate. Frameworks are consistent across languages to enable interlinking. Linking to other similar types of content or content of similar meaning creates context that further enriches the information for future use. The new lexical hub is format and platform independent, and built within a flexible technology that allows new combinations quickly to be prototyped and produced.

None of this sounds very much like building a dictionary. But such models are starting to be modelled and produced, and they are very possibly the Dictionaries of the Future. Keeping focused on the content, and its purpose, and being able to jettison modes of the past will be key to making the transition. But as globalization, digitization, and use of technology to communicate and transmit language and information continue to grow, I would argue that the future for lexical development, where it is transformed in these ways, can be very bright indeed.