If dictionaries are free, who will buy them?

Adam Kilgarriff

If dictionaries are free, who will buy them? The question looms over publishing houses like a slow-motion tsunami. Dictionaries are now free, on the web and bundled with Microsoft Office and other products, so where is the publishers' income stream going to come from?

Articles in the last four editions of this newsletter have addressed the title question. Charles Levine (KDN9) opened the debate in optimistic mode, seeing signs of growth in English-language lexicography despite the web. Joseph Esposito's response (KDN10) was a:

"grim vision … when I complained about Microsoft bundling a spell checker, with its limited dictionary, into Word ages ago, the techies I knew all laughed at me. Now that most of them have burned through their venture capital after Microsoft "integrated" the gist of their products into Windows, we all cry into our lattes together."

Then, following Levine's still-optimistic response (KDN11), the villain of the piece bravely entered the fray. Microsoft's Julian Parish (KDN12) argued that Microsoft saw publishers as partners, not competitors. There were many new e-opportunities, for example, as add-ons to Microsoft products.

History: a dictionary in every household

In the twentieth century, a number of European and North American publishers occupied the fertile coastal strip of "a dictionary in every household". Dependable as the cycle of one generation growing up and handing over to the next, it was a large and enviable market, in harmony with the grand and noble agenda of universal education. To be sure, the coastal strip was sometimes crowded with competitors, but the soil was good: there were always more households to buy dictionaries. They don't need to buy them any more.

There is no use lamenting the lost market. It may disappear with varying speeds: as Esposito notes and Levine confirms:

"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. In the short term, this will redound to the benefit of market leaders, such as Merriam-Webster and Oxford University Press…"

But disappear it will.

The market which is collapsing is the monolingual, emblematic "dictionary-at-home" market (the role of which has always been complex: status symbol for spelling, scrabble and – sometimes – schoolwork). Different markets, notably the boom EFL one and bilinguals that people need for travel and language-learning, have different trajectories.

The future

For the regular monolingual centerpiece, away from that lush dictionary-in-each-household coastal strip, what is there?

The key lies in quality. Most free dictionaries are not very good. Most people don't care: a dictionary is a dictionary is a dictionary, good or bad, and one is plenty. Some free ones are even quite good; Esposito and Levine note the quality of the Encarta dictionary, possibly the first of a new breed of market-swamping, "good-enough" dictionaries.

But the minority of people for whom language is their trade do care. They are the translators and academics, etc. The numbers are tiny compared to the golden age but, in this, 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. The up side is that customers can be found all over the globe and, once found, they are the right customers for the product so are likely to be willing to spend more.

The nice thing about this is that making good dictionaries, as opposed to bad ones, is what every lexicographer wants to do. There is usually tension between lexicographer and publisher – better vs cheaper – and the change in the market gives more weight to the lexicographer's case. While Esposito despairs at the traditional publishers being left "to focus on the scraps Microsoft leaves on the floor", we note that the market for the most accurate, the most consistent and the most current account of a language (or source-target pair) is far more than a scrap.

Of course, language professionals will be online. Lexicographically, this is exciting as it means the dictionary can be

Adam Kilgarriff is a Director of Lexicography MasterClass Ltd and Research Fellow at University of Sussex. His specialisms are language corpora, computational linguistics and lexicography. He has worked for Longman and as consultant to Oxford University Press, Cambridge University Press, Bloomsbury and Macmillan. He is on the executive board of Euralex. adam@lexmasterclass.com

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/
newsletter/kdn9-1.html

Joseph J. Esposito, Dictionaries, another Netscape?
Kernerman Dictionary News
10, July 2002
http://kdictionaries.com/
newsletter/kdn10-1.html

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

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

far better than any that went before: it is not constrained by space, and we can open our vision to the dictionary as an object integrated with the underlying corpus resources (as in Word Sketches [1]). But that is a different topic: here, our concern is for income streams.

Many of the language professionals are associated with universities and libraries. They are traditional customers for dictionaries, have substantial budgets, and, with physical space ever at a premium, are often enthusiastic about services which do not incur extra demands on space or personnel.

For example, Oxford Reference Online [2] is an online subscription service, sold almost exclusively as a site licence to institutions, incorporating a wide range of Oxford University Press's reference materials. It is very successful. Extensions which focus on language resources are planned. Of course, OUP has a wonderful brand, and has so many resources that it is able to offer a very broad resource, a one-stop-shop which is attractive to libraries. Others probably need to assemble into consortia (branding according to the best-known brand in each market). It is a route out of the path of the tsunami.

Dictionaries for computers?

All of the above is about dictionaries for people to use. Esposito, writing in 2002, says:

"The real game for Microsoft is using lexical databases within computer algorithms, as in natural-language processing."

Parish, too, stresses that Microsoft is an energetic customer for dictionaries for NLP (aka Language Technology, Computational Linguistics). As an NLP researcher, I'm a little sanguine here. To be sure, most NLP applications need dictionaries as inputs. In the short term most will probably be derived from dictionaries as we know them, where there are good ones available at reasonable cost. But consider, for example, Prinsloo and de Schryver's spellcheckers for African languages [3]. The wordlists are corpus derived.

Across NLP, researchers are finding ways of solving problems using corpora. While high-quality, well-structured, hand-crafted resources currently support technologies that corpus-derived resources don't, the list is shrinking. Even three years ago, Esposito's remarks looked right, but now, as NLP has changed, and while it may often be a short-term convenience for Microsoft and others to take publishers' resources, it is not an income stream for

the long term. While post-editing corpus-derived resources is a job that will need doing for some time yet, it is less than a glorious future for the grand old names of dictionary publishing.

Notes

[1] http://www.sketchengine.co.uk
[2] http://www.oxfordreference.com
[3] D.J. Prinsloo and G.-M. de Schryver 2003. 'Non-word error detection in current South African spellcheckers'. *Southern African Linguistics and Applied Language Studies* 21/4 (Special issue on 'Human Language Technology in South Africa: Resources and Applications'): 307–326. http://tshwanedje.com/publications/nlp-SPELL.pdf

Further Comments

Joseph Esposito

This is a very good article and I am delighted to have read it. A few comments, but all in the spirit of appreciation:

First, I do not and did not despair at the demise of Oxford and Merriam. I despaired at the complacency of the managements of these organizations for not addressing the maturity of their product lines and business strategy.

Second, the specialized dictionaries noted in the piece are precisely what I envisioned as the "crumbs" Microsoft would leave behind. As Bill Gates said apropos Sun Microsystems, you can only climb up that tree so far.

Third, the economic success of Oxford Reference Online is arguable. It's a good idea; if it generates a positive contibution to overhead (as I suspect it does), it's a very good idea. But it is not a sustaining idea. Library subscriptions cannot support the overall editorial and technical infrastructure of OUP's dictionary efforts. Other revenue streams must chip in, and those revenue streams are drying up.

Fourth, new product development (of innovative products, of course) is the answer to the business problems – but when is that not the case? The problem for the current crop of dictionary-makers is that the new products require additional investment, but the old products require ongoing maintenance. So the total cost of being in the business continues to rise, even as the market becomes more specialized.

Finally, and this is a matter of semantics, when I said that dictionaries would find a market in computational linguistics, I did not mean the current crop of general-interest dictionaries but new

forms of dictionaries, such as the ones derived from algorithmically generated corpora. The real point here is that there is a future for dictionaries, but there is no future for lexicographers. That's why we have machines. The future is only trivially post-print; more fundamentally, it is post-human. Do I despair? You could have asked that question of Cro-Magnon Man. But wasn't he, like ourselves, doing everything for his children?

Charles Levine

More than two decades into the digital revolution, and more than one decade into the Internet revolution, Adam Kilgarriff's assertion that the monolingual dictionary market is collapsing remains hyperbole – at least from the perspective of the United States.

There is no sign yet that sales of printed monolingual English dictionaries have declined in America; and, in fact, there seems to have been a modest increase over the past few years; possibly the result of an upturn in the economy and the appearance on the scene of more learners of English and more learners dictionaries to serve them.

For example, Merriam-Webster in 2003-2004 had one of its strongest launches ever of a new edition of its flagship Collegiate Dictionary – selling more than one million copies in the first twelve months. Merriam smartly bundled a free CD-ROM with the new print dictionary, but raised the retail price only one dollar.

Another example: An unnamed source at the country's largest book chain confirmed for me that the chain has experienced an up-tick in sales of print dictionaries over the past five years – which she attributed partly to an increase in market share by the chain, and partly to an increase in consumer purchases.

This is not to say that dictionary behavior is not changing rapidly, or that someday (sooner or later) digital lookups won't eclipse print lookups; but at the moment, the print monolingual English dictionary market in the States is chugging along with signs of modest growth.

Kilgarriff says that "Most free dictionaries are not very good," but this is misleading. For example, at the Merriam site (www.m-w.com) a visitor has free access to the previous edition of the Collegiate, which remains more than adequate for most lookups. And Merriam offers purchasers of its new print Collegiate a one-year free subscription to the corresponding updated fee-based Web site. Traffic on all its sites, I am told, has

increased significantly since 2003.

And www.dictionary.com (for whom I do some consulting) currently has free lookups to the latest American Heritage Dictionary of the English Language, which is one of the most popular of American monolingual dictionaries.

Not to nit pick further, but Kilgarriff mischaracterized my optimism about the near-term future of English lexicography, as something that would occur "despite the web." My optimism about English lexicography is because of the Web, and possibly the next phases of the Web, like the Semantic Web, we hear so much about.

By all (anecdotal) accounts, the Web has increased interest in language among students; has increased the worldwide use of English as a lingua franca; has given lexicographers more tools (like Google) with which to research and monitor language; and seems, as in the experience of Merriam, to have buoyed sales of print dictionaries, for the time being, at least, and not to have detracted from them.

Although I agree with Joseph Esposito that the print dictionary business may not be a growth business – i.e. not something you want to invest your money in if you are looking for a healthy return – and that this means that the dictionary business will be strained for investment capital, I am not one to predict the imminent decline of dictionary making – taking into account all sources of revenue, both print and electronic.

But, who can see clearly ahead ten, twenty, or fifty years? By then, we might have forms of e-paper and plastic look-a-like-books and pocket gadgets, all utilizing wireless broadband, on which we will be able to read entire libraries of information and content.

To me this suggests that some lexicographers, as much as science fiction buffs, might want to volunteer to be frozen cryogenically and revived in fifty years just to see how people look up words.

Unlike Esposito, I am confident that any such plucky lexicographers, once defrosted in the future, will find numerous co-workers still busily keeping up with a dynamic and changing English language.

Adam Kilgarriff

I hope Esposito's not right about the shrinking market failing to sustain lexicograph*ers* (as no-one will be investing in it), though I must admit there is a grim logic to what he says.