

DICTIONARY News

Dictionary *n. Obsolete?*

Ilan Kernerman

Kernerman Dictionary News has offered various views on the changing world of dictionaries and lexicography since the turn of the century (cf. p.4), and more recently the related processes and impacts have accelerated and become clearer. The topic has preoccupied me for some time and last year I put together a paper called 'Dictionary *n. Obsolete?* Before and afterwords', first presented at the Dutch Institute of Lexicology in November and since then this year on several other occasions. Also last November, Macmillan's announcement of ending dictionaries in print and going only digital stirred animated debate on email lists and expressed a new awareness. The main focus of this discussion and others seems to be print vs. electronic – basically revolving around practicality, economics and innovation as opposed to culture, personal taste and force of habit – but the question is not so much whether there will be dictionaries *in print* but would there *be* dictionaries, and how.

Paraphrasing Frank Zappa's quote on jazz (1974), one could say *dictionaries are not dead, they just smell funny*. While more are available (freely and easily) than ever before, dictionaries also lose their autonomous identity and disappear in language technology. Machine translation, word processors, search engines, learning aids and the like incorporate dictionary content and apply it in new forms and tools that go beyond the

dictionary as end-product per se. There will of course always be language enthusiasts and others who look up *a dictionary*, but mainstream usage without linguistic passion might be satisfied by broader communication and information solutions, such as talking on the phone in one language and hearing in another. As big dictionary names die out or dwindle and others struggle to survive, lexicography blooms in the technology world that recruits experts and invests in language R&D. We also see language companies selling specialized services to corporate, education and public sectors, beside national bodies (as well as academic and private initiatives) involved in innovative projects that are offered for free. Still, the major revolution today is perhaps linked to the reader-turned-user who gets involved in the creative process with the advent of crowd-sourcing in relation to expert authority (reflecting worldwider social, technological and political changes). At the background is a devaluation of the dictionary as *book of words* against growing needs for *full language* lexicography and new generation computational lexicographers.

The full paper is finally not published anywhere, but its list of main trends appears on p.7 of this issue, which is mostly devoted to more views from dictionary makers and lexicographers. Thank you to all the authors.

- 1 Dictionary *n. Obsolete?* | **Ilan Kernerman**
- 2 The future of dictionaries | **Judy Pearsall**
- 5 Redefining the dictionary: From print to digital | **Michael Rundell**
- 8 Make me a match: Putting learners in touch with dictionaries | **Colin McIntosh**
- 9 Free dictionary & Free lunch
- 10 Trade reference | **Sander Bekkers**
- 11 ASIALEX 2013 in Bali | **Deny A. Kwary**
- 12 A sweetshop for lexicographers | **Jaap Parqui**
- 13 X International School on Lexicography – Life Beyond Dictionaries | **Olga Karpova**
- 14 Paragon Software: Turning gigantic print reference content into light-weight apps with cutting-edge features | **Alexander Zudin**
- 16 A brief account of Dutch lexicography | **Tanneke Schoonheim**
- 22 Matthias de Vries Genootschap | **Anne Dykstra**
- 24 The new *Lighthouse* dictionary | **Kaoru Akasu**
- 27 Why do we need pattern dictionaries (and what *is* a pattern dictionary, anyway)? | **Patrick Hanks and Jane Bradbury**
- 30 Fortunes of National Cultures in Globalisation Context | **Vera Budykina**
- 32 Patrick Hanks. *Lexical Analysis: Norms and Exploitations* | **Orion Montoya**
- 33 *KERNERMAN Japanese* dictionaries by Seiko
- 35 Launching of *LEXOCOGRAPHY: Journal of ASIALEX* | **Ilan Kernerman**
- 36 **Paul Bogaards** 1940-2012 | **Wolfgang Worsch** 1953-2013

Editor | **Ilan Kernerman**



KDICTIONARIES

© 2013 All rights reserved.

K DICTIONARIES LTD

Nahum 8 Tel Aviv 63503 Israel

Tel: 972-3-5468102

Fax: 972-3-5468103

kdn@kictionaries.com

http://kictionaries.com

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.

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 only¹. 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 Kilgariff, 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 topic⁵, 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.

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/>

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

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

Make me a match: Putting learners in touch with dictionaries

Colin McIntosh



Colin McIntosh is the Publisher of *Cambridge Dictionaries Online* and editor of the *Cambridge Advanced Learner's Dictionary, Fourth edition*. In a previous life he edited new editions of the *Oxford Collocations Dictionary* and *Phrasal Verbs Dictionary*, as well as developing the *Oxford 3000* vocabulary list for the *Oxford Advanced Learner's Dictionary*. He is also an experienced language teacher, and has written self-study materials for learners of Italian. mcmintosh@cambridge.org

In the new, Fourth edition of the *Cambridge Advanced Learner's Dictionary*, the entry for *book* has been rewritten. The first sense is no longer “a set of pages fastened together inside a cover”, but “a written text that can be published in printed or electronic form”. The “set of pages” is destined to become a historical sense, much like *codex*, *scroll*, or *tablet* (the Mosaic rather than the Jobsian variety).

A dictionary, too, is a book, that is to say a written text in printed or electronic form. In undergoing the transformation from a set of pages to a text, the dictionary is in the process of achieving its destiny. In fact, the dictionary was electronic before its time, with its hypertext (cross-references), non-linear progression, and easy access to any piece of data that the user required. Online, alphabetical order is no longer relevant. Dummy entries, placeholders that cross-refer to other entries, are no longer needed. Running heads, alpha starts, run-ons, and all the rest of the time-honoured catalogue of dictionary furniture are consigned to the bonfire.

Much thought has been given to how to convert dictionary codices into electronic dictionary texts; perhaps less to how users encounter dictionaries in the real world.

One thing that strikes me in my role as dictionary publisher (as opposed to lexicographer) is that the distinctions between different types of dictionary are being eroded by online dictionaries.

Learners' dictionaries developed in the middle of the last century, born of a realization that the needs of learners of English (or of any language, for that matter) were very different from those of native speakers. Native-speaker dictionaries rarely trouble themselves with grammar, because things like complementation patterns and restrictions on plural use are obvious to natives. With its grammatical information at the lexical level and model example sentences, the learner's dictionary filled a gap in a way that nothing else has been able to do so far.

ELT bilingual dictionaries developed later, and continued the same approach to the type of information presented as the learners' monolinguals had done, setting them apart from the traditional type of

bilingual dictionary, which tends to be more useful as a decoding rather than an encoding tool.

ELT publishers and editors have always made great efforts to pass on the message to teachers and learners that learners' dictionaries are an almost inexhaustible mine of information and advice about English that can help with encoding as well as decoding. It has to be said that this message has got through only sporadically. To the vast majority of users of language (i.e. virtually everyone), a dictionary is still a monolithic fountain of knowledge that is not susceptible to variation or adaptation, and not the highly flexible tool that can be recreated endlessly to suit its users which lexicographers know it to be.

In the days of print, the distinction between learners' dictionaries and native-speaker dictionaries was made clear-cut by their physical separation in the bookshop. If your local bookshop was in São Paulo, it is likely that it stocked English learners' dictionaries and bilinguals, not native-speaker English monolinguals. If your local bookshop was in Brighton, you would find your learners' dictionaries in the ELT section. You would be guided by the realities of distribution or the layout of the shop into choosing the most appropriate dictionary for your needs. The question of level would largely be decided by looking at the dictionary on the shelf and opening it at random. If the 2000-page doorstopper looked too forbidding with its sheer bulk and closely packed print, then an intermediate or elementary dictionary would do the job.

In the digital age, these distinctions no longer apply. In the absence of the physical cues of the bookshop layout or the printed books themselves, a certain degree of sophistication is required of the user, a sophistication that we cannot necessarily expect them to have. When a student in Italy wants to know the meaning of *upcycling*, the most likely course of action is to google “upcycling definition” or “upcycling meaning” or “upcycling traduzione”, and wait for the avalanche of sources (Wikipedia, Urban Dictionary, learners' and native-speaker dictionaries, monolingual and bilingual dictionaries), all jostling for space on the first page of results.

Faced with this bewildering array, the user is likely to choose the first result, or the first name that he or she recognizes. The type of dictionary will be largely immaterial, and the whole historical back story of ELT dictionaries, from Hornby's flight from Japan in 1942 with his *Idiomatic and Syntactic English Dictionary* in his suitcase all the way up to the latest innovations in corpus lexicography, becomes invisible.

The result of this is that it becomes more difficult for the publisher to unite the user with his or her ideal dictionary. Where previously the publisher was the village matchmaker, the situation now can be more like a free-for-all online dating service, if not something rather more promiscuous. While SEO is a very useful tool for publishers, it does not necessarily do the job of putting the right dictionary at the user's fingertips.

Of course, not everyone comes to the dictionary via Google. An important part of our job as publishers will be to build up our online brands, so that the user comes straight to our site, whether via a bookmark or a widget. Cambridge's API is an attempt to do this at a group level rather than an individual one. A website with a particular community of users will be able to supply the datasets that most suit their users' profile.

Localization and channelization are other ways that dictionary publishers can get closer to their customers. Thanks to geo-targeting, specific local-language bilinguals be offered; advice about choosing the right tool for the job can be offered in the local language; the metalanguage of entries can be translated (and the user is always at liberty to reject the default option if that is not appropriate to his or her needs). If the level of words being looked up is intermediate rather than advanced, we can suggest an intermediate dictionary as the default. If the lookups tend to belong to a particular domain, we can suggest an ESP dictionary. And user profiles can help in hiding or showing particular elements of the entry, such as IPA, translations, or extra examples.

In addition, publishers can actually take advantage of the absence of physical separation. Where previously a conscientious writer, editor, or translator would need various monolinguals and bilinguals, a thesaurus, a collocations dictionary, a spelling dictionary, and a usage guide, all of this information can be combined in one dataset. I may want to know a meaning of a word; but how do I say it in Turkish? And what are its collocations and synonyms? So instead of having to come out of one book to see what another book is saying about the same word, I can display as much or as little information about it as I like, all on the same screen.

Thus, a one-size-fits-all approach does not have to be the logical outcome of online dictionaries. The difference is that, before, the differences between dictionaries were explicit and enforced by the physical product, whereas now they are hidden below the surface so that the dictionary-using experience is a streamlined, almost automatic one. One thing that has not changed is that we still have to get to know our new customers extremely well, something that has always been the dictionary publisher's job.

Fortunately, we are now in a position to get to know our users better than ever before. Whereas in the past the feedback on our dictionaries was limited to small surveys, individual lookup observations, and letters from individual customers, we can now track users' journeys (with their permission), so that we can form a much clearer picture of the words they are looking up, the domains that these words fall into (and therefore their interests), the level of English that their lookups indicate, and even the type of word information that they are seeking. Our surveys can reach vastly increased numbers of people, with the result that we are closer to our users than ever before, and for the first time we can really start to understand their language-learning habits and preferences so that we can often answer their questions even before they have started to formulate them.

Some users may miss the old serendipity of browsing the pages of a print dictionary (the forerunner of the user journey), but the new technology used by the best online dictionaries has the potential to offer an experience that is of far higher quality in terms of the targeting of the information, and which can adapt itself to the user's backgrounds, needs, and interests in exciting and surprising ways.

Cambridge Dictionaries Online

Cambridge University Press is a leading global publisher of English language learning and teaching materials. The first Cambridge learner's dictionary was published in 1995.

Originally entitled *Cambridge International Dictionary of English*, this developed into the *Cambridge Advanced Learner's Dictionary*, and has just been published in its fourth edition. It was followed by dictionaries for intermediate and elementary learners, dictionaries of American English, and semi-bilingual editions for Chinese, Turkish, Polish, and Russian speakers.

Cambridge Dictionaries Online (CDO) was launched in 1999, the first dictionary website with learners of English in mind, and became one of the most popular dictionary sites on the Web. It remains the pre-eminent online learner's dictionary, with millions of users around the world.

Cambridge and K Dictionaries have just announced their decision to partner up to provide bilingual English learner's dictionary content for a range of languages, which will be added to CDO over time.

<http://dictionary.cambridge.org/>

Free dictionary & Free lunch

TheFreeDictionary.com welcomed its five billionth visitor in April 2013, just months before its 10th birthday. In 2007 the website's owner, Farlex Inc, launched Definition-Of.com as a community-based dictionary that allows users to generate, access, and rate its content. For every two definitions approved on the site, Farlex donates one school lunch to a hungry child through the United Nations World Food Program's (WFP) school meals program. In this way, users are afforded the opportunity to share their knowledge with others and, at the same time, fight global hunger. In addition to its monetary donations to the WFP, Farlex has provided it with development resources and advertising space on its websites. To date, these combined efforts have helped provide over 100,000 meals to children around the world. As noted by the WFP, these meals do more than just nourish the children physically; they help nourish their minds, encouraging them to attend school and improving their academic performance.

<http://farlex.com/>

Trade Reference

Sander Bekkers



Sander Bekkers is Publisher at Prisma language and dictionaries, Department of Unieboek|Het Spectrum, part of the Lannoo publishing group. He has rich experience in dictionaries and reference publishing, and has previously worked with Bert Bakker, Amsterdam; Van Dale Publishers, Utrecht; Educatieve Partners, Houten; and Reed-Elsevier, Amsterdam. sander.bekkers@unieboekspectrum.nl

Commercial dictionary publishing is changing, for several different reasons publishers have encountered a loss of product relevance to their public. One of my former colleagues once mentioned: language and communication is very much alive, everyone communicates continuously and most of it is done in writing, however dictionary publishers have trouble meeting the needs of today's writers and finding a niche in the present writers' environment where people communicate. The reality, however, is that dictionaries are by and large not being used as an active writing or learning tool, but as a passive reference tool.

If language learning is defined as combining a grammar database with definitions and relevant phrases within, a dictionary could be an essential construct. If you define a dictionary as a functionality to assist you in writing texts in your own or another language, or as a tool you can use to acquire a language, there are several solutions, none of which include it being in book form. But the fact remains that for dictionary makers books are still the primary revenue base.

It seems as though, for the last thirty years digital reference sources have always held the promise that the market would shift from dictionaries as books to digital media. Up to now this promise has not been fulfilled, however recently in Germany the situation has shifted. Duden Verlag has gone online, publishing their main dictionary *Die Deutsche Rechtschreibung* on the internet for free, and other publishers such as Macmillan and Pons have done the same. The number of internet hits indicates they are doing fairly well, but they were more reserved when inquiries were made about their revenue. The fact is, however, that doing nothing means an ultimate decline and finally – discontinuation.

Examining the matter more closely, it is clear that there are some divisions to be made. There are the different platforms upon which you can publish dictionaries, including online, smartphones and tablets, desktop computers or integrated dictionaries, as well as adding dictionary functionality to other applications. But it is also interesting to consider different users: a student or pupil, learning the language, is a different user than a professional translator or a staff member in an internationally-operating company – they all have different needs.

The international community increasingly uses the English language in global business communication, and this tendency is the most obvious reason behind the increase in the trade in English bilingual dictionaries and the decline in demand for German, French or Spanish bilingual dictionaries. Added to this are also internal issues, concerning the way dictionary publishers were organized in the past. Most traditional dictionary publishers still originate from book publishing and lexicographers, while the real innovations are now coming from computer programmers. This all depends of course on how dictionaries are defined.

Dr. Robert Amsler, an American computational linguist, wrote the following on this matter (in the email discussion following Macmillan's announcement on replacing print entirely by digital publication):

The work I did on the analysis of dictionary definitions demonstrated that there was an imperfect, yet intriguing, taxonomy of definition texts and showed that the alphabetic organization of dictionary entries was outmoded except under special circumstances. I.e., for example, you had to know how to spell a word to look it up; you had to know a word existed that dealt with the meaning you were trying to express to know how to look it up; and when you did look a word up you were given a tiny view of the dictionary's contents that didn't show you the other words whose definitions were related to the entry you were examining in terms of taxonomic relatives. Sure, some dictionaries did an excellent job of including information on synonyms... but NONE gave taxonomic or part/whole related headwords.

In many discussions there is no clear differentiation between the market position of bilingual and monolingual dictionaries, but the differentiation is there, especially in the Netherlands and Belgium. Monolingual dictionaries have lost considerable market share to spell-checkers, Google and free online dictionaries, spell-checkers have taken over the spelling function that not so long ago were provided by dictionaries. Google is also used as a resource for looking up word definitions, another function formerly provided by dictionaries. Bilingual dictionaries have a *must have* aspect, as they have become essential while writing

This article is based on the presentation made at the Matthias de Vries forum, at the Dutch Institute of Lexicology (INL), on 28 November 2012.

and translating texts in languages other than your native one.

There is also the promise of the dictionary being included within a computer application that translates or otherwise helps users with different language tasks. For instance, we frequently spoke of the opportunities in selling dictionaries in digital form with Microsoft products. Looking at past initiatives, the most successful enterprises were having our content combined with software, packaged with hardware, or offered with that of other publishers. While we probably sold over a million digital dictionaries, the revenue per product was always extremely low. On the other hand, we also observed that software products, especially when developed and marketed as a single product, had only a very short life-cycle.

If we look at all the available platforms there is a difference:

eBooks: you cannot switch easily between the book you are reading and the dictionary. A dictionary for this platform is therefore not efficient. In the beginning it was also difficult to have a useful search function, but nowadays the possibilities have evolved, including widgets, and a user-friendly interface is workable on certain platforms.

CD: not as a stand-alone product, we combine cd-rom with books. The interest of the public is waning, however, and selling a product as a download seems a better solution.

Apps: developing for tablets and smartphones is an interesting development. Prisma recently entered this market with a steep sales rise, I am curious to see how it will continue.

Online: there is potential with subscription for organizations, whereas offering content behind a login (even free) has limited commercial value for private users. The only digital products people seem to want to spend money on are those that can be placed on their own device as an app, but most dictionaries on the internet are free.

Books: we still earn 95% of our income through books, and not digital dictionaries. The number of books sold is still the same as several years ago. This is led by the number of students, and not the economy, which however forces us to use different distribution channels, with correspondingly different prices.

With an eye to the future, it is clear that there will always be a need for well structured and well edited lexical data. The way we earn our money, though, will be completely different. The development of free online dictionaries by the traditional dictionary publishers (as well as newcomers) is unstoppable, and we will be selling more of our products in the slipstream-online and increasingly less often in stores. What we need is a new entrepreneurship to create new products for new users, doing what we have always done: helping people to write, learn and understand language, working closely together with scientists and programmers to finally step into the digital future.

Prisma Dictionaries

Prisma dictionaries and language publishers are a market leading school dictionary publisher and language learning publisher in the Netherlands and Belgium. Most important products are a series of pocket dictionaries and a series of language learning books for self study.
<http://prisma.nl/>

ASIALEX 2013 in Bali

The 8th International Conference of the Asian Association for Lexicography (ASIALEX) has attracted participants from 35 countries. It will be held in Bali, Indonesia, on 20-22 August 2013, and its theme is **Lexicography and Dictionaries in the Information Age**.

The conference is expected to contribute to the development of lexicography and dictionaries in Indonesia, as well as in Asia in general, where the number of languages is significantly higher than that in other parts of the world. According to data from Ethnologue.com (accessed on 21 May 2013), out of 7,105 living languages in the world, 2,304 are in Asia. Indonesia has 706 living languages, of which only 21 are institutional, 97 are developing, 248 are vigorous (i.e. used for face-to-face communication), 265 are in trouble, and 75 are dying.

The conference topics include the evolution of dictionaries into multiple information sources, lexicography as an interdisciplinary research field, development of online and mobile dictionaries, features of pocket electronic dictionaries, printed dictionaries in the information age, dictionaries for minority languages,

advances in dictionary user surveys, and dictionaries as language learning resources

Featured speakers include Dr. Diah A. Arimbi (Indonesia), Prof. Henning Bergenholtz (Denmark), Dr. Adam Kilgarriff (UK), Prof. Robert Lew (Poland) and Prof. Yukio Tono (Japan), and there are 75 parallel papers and software demos.

The conference is jointly organized by ASIALEX and the Faculty of Humanities of Airlangga University in Surabaya. The proceedings will be published by Airlangga University Press and distributed to selected universities in Indonesia and worldwide.

The conference will be held at Bali Dynasty Resort, a premier beachside resort at Kuta Beach, renowned for its warm Balinese hospitality. The conference room is very spacious and has high teakwood ceilings, a permanent stage, and the latest in audio-visual equipment – to ensure the comfort of the presenters and the participants. For further information see <http://asialex2013.org/>.

Deny A. Kwary
deny_ak@yahoo.com

A sweetshop for lexicographers

Jaap Parqui



Jaap Parqui studied psychonomy and Italian linguistics and literature at the University of Leiden. At the beginning of his studies he saw a note calling for trainees to work on an Italian-Dutch dictionary project and knew that this was what he wanted to do. A couple of years later he entered this project as a trainee, eventually rising to become an editor. He then moved on to Van Dale, first as an editor and then as a publisher. He is still fascinated by the idea that in a dictionary the whole world is collected and neatly ordered. jaap.parqui@vandale.nl

Eatable marshmallow pillows
Lickable wallpaper for nurseries
Hot ice creams for cold days
Cows that give chocolate milk
Fizzy lifting drinks
Square sweets that look round
(Roald Dahl, *Charlie and the Chocolate Factory*, 1964. NY, Knopf)

Since the outbreak of the economic crisis in 2008, sales of reference books have fallen considerably in Holland. Electronic dictionaries are meanwhile flourishing. Print dictionaries have been perfected over the centuries within the tight limits of the printed page. It is hard to introduce new ideas to them. Electronic dictionaries, however, are a recent innovation and there are still numerous ways of improving them. There are so many platforms to develop, so many functionalities to invent, so many more sources from which to harvest material, so many new ways to get in touch with the public...

For both publisher and user this is a sweetshop of new possibilities. There are so many interesting, useful and inspiring possibilities that the biggest challenge now is to find a focus and pick the right ideas. Because of the myriad of options offered by the electronic revolution, it is hard to predict which road lexicography in general will take. I will now discuss some of the new developments Van Dale has chosen to work on.

Update cycle

Is it the changing view of the lexicographer or do new words arise (and disappear) faster nowadays? Although the word *hype* was invented in the nineteen twenties, it now seems more alive than ever. With online subscription models, customers expect their data to be right up-to-date. This expectation is challenging. What should be the policy for admitting new words into dictionaries? And will the back office be able to maintain this speed for so many dictionaries simultaneously?

In the *Dikke Van Dale*, our monolingual flagship, we used to include only established words. In practice, this meant that a word needed to have been used in general sources for at least three years. And even now, in the online era, we do not want our dictionary to be cluttered with ephemera and flashes in the pan, especially those that have already

become dated. That is why we decided to give new words a temporary status. New words are accompanied by the message 'temporarily added' plus the year of their appearance. After a couple of years we decide which of these 'temporary' words have survived the test of time.

Only a couple of years ago, if we wanted to publish a new edition of a book, we needed at least three months to perform all the checks. The file would then go to a typesetting company, which would take another couple of weeks, and it is only then that printing would start. In order to publish more often, we therefore needed to increase efficiency greatly. We transformed our data files into XML, made them more uniform and invested in automatic typesetting from XML. Most of the checks are now done along the way. We are still fine-tuning the process. This year our most important online dictionaries will be updated twice and we are working towards an even higher frequency.

Crowd sourcing

With better access to all kinds of corpora the quality of our dictionaries has improved considerably. It can still happen though that we are slow to spot a new word, or that there is an even better translation or definition. We do not cover highly specialised terminology. And you can never have enough phraseology.

We are developing an option that will allow users to add new words, new phrases and new translations or definitions to our dictionaries. These will appear, with their own very distinct look and feel, alongside the 'real' dictionary.

I see many advantages. It could be a very good way to include much more information in our dictionaries. It is a nice way of developing a closer tie with our users. And it can be a good source for the editors, who will promote information from the 'people's dictionary' to the official dictionary.

It is also a leap in the dark. Will the information the crowd adds make sense? Will users actually make use of this option? How will they perceive the difference between official and non-official content? That is why we will give this *Van Dale Wiki* a soft launch. We will start with just one dictionary and a selected group of participants. And from there we will slowly expand this exciting experiment.

Search functionality

We polled our users. The first reason they consult our explanatory dictionary is to look up the meaning of a word. The second reason is to look up the spelling. At present you can only find a word in our online dictionaries if you already know how it is spelled. That is why we are working on a 'suggestion module' that will offer a short selection of words for which you may have been looking. This will be similar to what Google or spell checkers currently do.

It is already possible to look for phrases, by entering one or more words without worrying under which headword you should look. Many other smart-search options are also imaginable.

Flexion forms

Many users will need conjugations of French verbs, the flexion of German nouns or plurals of Dutch nouns. Print dictionaries have very little space for flexion forms and usually limit themselves to key forms, or to a reference to a model verb or noun. In electronic dictionaries we have abundant space. Producing all these forms in so many languages is a major project. We do this portion by portion and try to find the cleverest balance between automatic expansion and handiwork to complete the paradigms.

In paper dictionaries flexion forms are typically given after the headword. When a user looks up a translation in a foreign language, however, that is typically the moment when he or she needs to see its inflected forms. We will therefore place a small button next to the translation that will pop up a window with the forms. We will do something similar for spoken pronunciation.

Images

It is a cliché, but sometimes a picture does say more than a thousand words. Some print dictionaries do have pictures, of course, but once again electronic dictionaries offer much more space. The production of pictures is very expensive and so we are considering placing links to pictures on the Internet. And we are not limiting ourselves to stationary images. For words that describe movement – dances, for example, or a horse's gait – film is much more appropriate.

Language platform

Translators often tell us that they look up a word in a bilingual dictionary and then check that translation in an explanatory dictionary. They might also want to check specialised dictionaries or read paragraphs on usage. I am doing it myself, as I write this article in a language that is not my own. It is therefore important that our users

should have a choice of many products and be able to switch effortlessly between them. We built an online service, into which it is relatively easy for us to incorporate new products. We are developing new products ourselves, but we also collaborate with other publishing houses – *Oxford English Dictionary* and *Oxford English Thesaurus*, for example. In this way we are building a complete online platform that will provide users with many different kinds of reliable language information.

Books

Does this mean we will not make books anymore? We love the possibilities of electronic dictionaries, but we also love the smell and the romance of books – and the fact that you can use them without your computer. And so do many of our users. Once you have invested in the data, it is not a huge investment to publish it in print too. I expect we will publish books for many years to come.

Conclusion

So where will we be in ten years? As the saying goes, it is hard to make predictions, especially about the future. If we reason along the lines I have sketched here, in ten years time we will have high quality dictionaries on many devices and platforms, with lots of data, both official and user-generated, not just words, but also sounds and images that are very easy to search for on many devices and platforms. I think dictionaries will then be more reliable, richer and easier to use than ever before.

Van Dale Publishers

It is more than 150 years since Johan Hendrik Van Dale died a slow and horrible death of smallpox just before finishing his *Van Dale Groot woordenboek der Nederlandse taal*. It was finally completed by his student, Jan Manhave. If only Johan Hendrik had known the status his *Dikke Van Dale*, 'Fat Van Dale', would go on to attain in Holland and Belgium! In the nineteen seventies a publishing house was established around this title and a series of large translation dictionaries evolved. Over the years the portfolio has grown, and nowadays Van Dale publishes dictionaries in many shapes and sizes as well as grammars, usage books and books for children. In the last couple of years Van Dale has also started in-class language courses and a translation service. <http://vandale.nl/>

X International School on Lexicography – Life Beyond Dictionaries

Ivanovo State University is celebrating the tenth anniversary of its International School on Lexicography by holding the current event in two different places simultaneously: Ivanovo (Russia) and Florence (Italy). In Florence we plan to discuss questions connected with cultural tourism and heritage dictionaries, cultural aspects of learner lexicography and new dictionary projects, while in Ivanovo participants will focus on dictionaries for special purposes, historical lexicography and problems of the modern lexicographic scene. Online video conference will give us the opportunity to have personal interaction without being at the same location.

Florence was chosen to co-host this conference in view of the cooperation that our university has had since 2004 with the local Romualdo Del Bianco Foundation – Life Beyond Tourism, which promotes international events, including our annual international student workshops. This year the workshop participants will present their projects as part of the conference.

Invited speakers include Linda Mugglestone (Oxford University) and Sandro Nielsen (Aarhus University) in Ivanovo, and Paolo del Bianco (Romualdo Del Bianco Foundation), Janet De Cesaris (Universitat Pompeu Fabra) and Geoffrey Williams (Université de Bretagne Sud).

<http://en.lex2013.16mb.com/>

Olga Karpova

olga.m.karpova@gmail.com

Paragon Software: Turning gigantic print reference content into light-weight apps with cutting-edge features

Alexander Zudin



Alexander Zudin graduated from Moscow Institute of Physics and Technology in 1998 with a degree in computer science. He co-founded Paragon Software in 1994, and is the company's CEO. alex_zudin@penreader.com

The digital dictionary market is maturing, providing new opportunities not only for large and prominent publishing houses, but for smaller ones as well. Independent authors come up with rare reference editions adding new languages, unique dialects and very specialized vocabulary to the wide spectrum of already established traditional reference and learning brands. Paragon Software encompasses all languages and dialects, in the aim of creating digital applications that satisfy all mobile user reference and learning needs.

Since its launch in 1994, Paragon has brought to the mobile market more than 2,000 titles based on its proprietary Slovoed compression technology, which compresses even the largest-known print volumes (such as *Britannica Concise Encyclopedia* and *Webster's Third New International Dictionary, Unabridged*) to the tiniest apps taking minimal memory space and offering fast look-up functions. This advanced technology has already been employed by dozens of publishers around the globe. One of the latest joint projects, to be accomplished in August 2013, is the release of the K Dictionaries' *PASSWORD* series apps, consisting of over 40 semi-bilingual English dictionaries.

Slovoed makes it possible for publishers to offer their titles for all available platforms – whether desktop, mobile or web-based – broadening the appeal to a range of consumers while decreasing time-to-market. The essentials of the digital development process are presented in this article.

1. Reliable conversion technology to re-create the brand

The publisher's first steps in the mobile world begin with converting print content from PDF, MS Word, InDesign, MS Excel or other formats into the engine-compatible digital format. The biggest challenge and the end goal of successful conversion is to communicate all brand elements into a mobile version and to present the user with similar design, look, colors, logo and other visual components as those of the print edition.

In order to provide a complete brand correspondence as well as utmost comfort and benefits for the end-user, the Slovoed technology restructures dictionary entries

by separating them into components such as examples of usage, translations, phonetics, parts of speech, tables, pictures, and other groups. The structuring operation permits the developer (in agreement with the publisher) to add to the app new advanced features, such as colored markup, hyperlinks between articles and directions (for example, English-French or French-English), the possibility to hide or display specific information (examples, comments, etymology, etc) in the entry for a more customized view on a mobile device, and many other features.

2. Simultaneous multi-platform development

The key point a publisher should consider when entering a digital dictionary market, is that it has a highly variable structure, where the revenue level depends on popularity of a certain platform (i.e. operating system). Platforms continuously replace each other, winning market share, bringing unexpected breakthroughs or ruining hopes. Thus, some titles for Mac OS may bring more revenue than those for Android, or vice versa. It's all relative. This instability of the digital dictionary market raises tricky questions for publishers, such as which platform to choose, and how to secure one's investment in case of poor performance of a new platform (e.g. Bada, Windows Phone, Blackberry) and avoid such risks.

Slovoed attempts to ensure a stable revenue level by providing multiple platform support. The engine supports more than 15 platforms – from Palm OS, Java, UIQ and Symbian to Android, iOS, Mac OS and Windows 8. Paragon's 18-year experience has proved that a cross-platform approach is the core of a long-term strategy in digital dictionaries. To implement this idea, Slovoed technology uses Device Agnostic Data Format (DADF), which is a digital data format that does not depend on the operating system into which it is implemented.

Using DADF simplifies the product development to shell migration only and reduces development costs per title (e.g. by incorporating the keyword list, articles, sound, morphology, grammar, etc). For example, a database developed for iOS

can be used on another platform (e.g. Android) without the database recast. In conjunction with multiple development teams (one per platform) that can work on many dictionary projects simultaneously, Slovoed's cross-platform approach decreases a product's time-to-market and enhances stable revenues.

3. Innovative end-user features to take advantage of the device's capabilities

Recent services and technologies, such as from Google and Apple, have raised user expectations so highly, that developers have to create apps with features that meet these new standards. Slovoed technology for mobile and desktop platforms offers premium opportunities for users to translate and learn languages, featuring intelligent search tools, full-sentence audio pronunciations by native speakers, grammar tables, vocabulary tests, sophisticated input options, friendly UI, etc, and guarantee very fast and intuitive operation. Here are the essential features:

- A morphology module analyzes and provides translation for words entered in any grammatical form (plural nouns and adjectives, verb inflections, etc). Since most words occur in speech in inclined or conjugated forms, it may be tricky for a language beginner to figure out an infinitive and to get a proper result.
- When a user enters a query into the input field, the engine automatically builds a list of different entries through the entire dictionary that contain the query.
- Slovoed finds precise word translations containing all usage examples.
- Full text search allows users to look up a word not only in headwords, but also through all available translations and examples of usage in the entire dictionary. The app delivers comprehensive results that offer users instant access to all available content, along with a thorough analysis of all findings. For example, by typing the word **taste**, the user will see not only the entries and links to the most relevant entries but also others such as *discriminating* and *educated* that happen to provide examples of phrases with this word. As a result, the translation becomes more comprehensive and detailed.
- Full text search involves the morphology module and 'Did you mean...?' functionality, and thus finds almost all search queries regardless of misspelling or grammatical word form.
- Search results are displayed in both languages (for bilingual dictionary apps).
- Along with powerful search options, users can write words or phrases by hand without using a keyboard. This new handwriting

recognition capability is enabled through the integration of Paragon's proprietary handwriting recognition PenReader technology that supports 30+ languages.

4. Trim the sails to the wind: marketing

Once the application development has been finished, the most important next step is to market the new product and submit it to online sales channels, such as the iOS App Store, Google Play Store, Amazon App Shop, Windows Phone Store, and others. Paragon is a leading expert in marketing services, successfully managing more than 2,000 reference applications in the company's own online distribution platforms for every major app store.

5. Web services

Looking closer at the user interaction patterns for mobile and desktop devices, we find that global giants like Apple and Google influence them too. Thus, iOS 6 with its iCloud Tabs allows users to sync browser tabs between different devices and continue viewing the page right where it was left on a previous device. A similar technology is supported now in a recent version of Google Chrome for iOS. As a result, the user gets a hitherto unknown mechanism of cross-platform content consumption on multiple devices and browsers with seamless access to the content of interest.

Paragon has applied the trend to mobile and desktop apps and introduced the Slovoed Online dictionary web service. Now users of mobile apps, powered by Slovoed, can get free access to their dictionaries online and, what's more important, synchronize their search history. No matter where the users are situated – in the office, in front of a laptop, or on the go with a smartphone in hand – they get the desired content in most comfort and context.

6. Open Dictionary API Initiative

The latest initiative by Paragon, aimed at integration of new opportunities for promoting dictionary apps, is an agreement with the Open Dictionary API Alliance (ODAA). ODAA is the first non-profit open community that provides open versions of in-app translation tools to software developers and organizations. Now third-party developers of e-books, journals, encyclopedias, office documents and social media apps for Android and iOS can integrate instant in-app word translation functionality and cutting-edge dictionary products into their own apps without any fees. Through this integration, mobile users can read foreign literature and documents in multiple languages without leaving the active application.

Paragon Software

- Operating in the mobile reference business since 1994
- Headquarters in Freiburg, Germany
- Development offices in Moscow, St Petersburg, Obninsk
- Sales offices in Irvine CA, Tokyo, Shengjeng
- 80+ employees in digital publishing team
- 15+ supported operating systems (OS) – including Android, BlackBerry, iOS, Mac OS, Windows 8, Windows Phone
- Dozens of publishing partners – including Collins, Duden, Encyclopædia Britannica, K Dictionaries, Langenscheidt, Merriam-Webster, Oxford University Press, Pons
- 2,000+ apps – including dictionaries, phrasebooks and encyclopedias
- 40+ languages
- 40+ features
- Millions of users worldwide
<http://paragon-software.com>

A brief account of Dutch lexicography

Tanneke Schoonheim



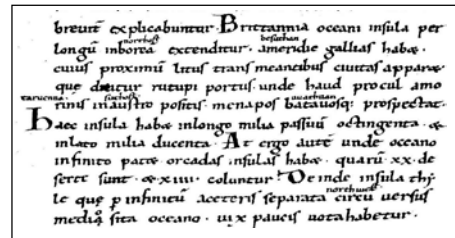
Tanneke Schoonheim has an MA in Dutch Linguistics and Literature and a PhD in the Arts from Leiden University. In 1988 she joined the Instituut voor Nederlandse Lexicologie (INL – Institute for Dutch Lexicology), and was editor of the *Vroegmiddelnederlands Woordenboek* (1988-2000) and *Oudnederlands Woordenboek* (2000-2005). From 2005 to 2009 she was head of the INL's department of Descriptive Historical Projects and editor-in-chief of the *Woordenboek der Nederlandsche Taal* online, *Oudnederlands Woordenboek*, and *Etymologisch Woordenboek van het Nederlands*. Since 2009 she heads the department of Description & Production and is editor-in-chief of *Algemeen Nederlands Woordenboek*. She is board member of the Dutch-Flemish Foundation of Dutch Dialects, chair of the Dutch-Flemish Spelling Commission, member of the Fryske Akademy, and on the editorial board of the *Journal of International Lexicography*. She was co-organizer of the 14th International Euralex Congress in Ljouwert and co-editor of its Proceedings (2010).
tanneke.schoonheim@inl.nl

Introduction

Modern Dutch lexicography is part of a longstanding tradition. The dictionaries we nowadays produce and use in the Netherlands could not be what they are without their predecessors. In this survey I would like to present some lexicographical works that can be counted as highlights in the history of Dutch lexicography.

The medieval period

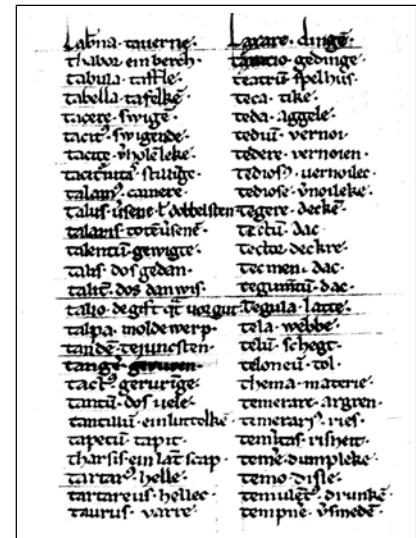
In the early Middle Ages, the need for dictionaries grew simultaneously with the increase of literacy in society. From this period onwards, we find Latin manuscripts in the libraries of monasteries with explanations of words written in the margin of the text or between the lines, for instance the so-called Orosiusglosses. These explanations or glosses were meant to increase the comprehensibility of the text and were also used for educational purposes. The glosses from these texts form the origin of our dictionaries, and soon collections of glosses were made into so-called glossaries and these were duplicated, reworked and manifoldly spread.



Orosiusglossen St.-Omaars (ca. 1150-1200)
BM 717 © Bibliothèque municipale, Boulogne-sur-Mer / IRHT

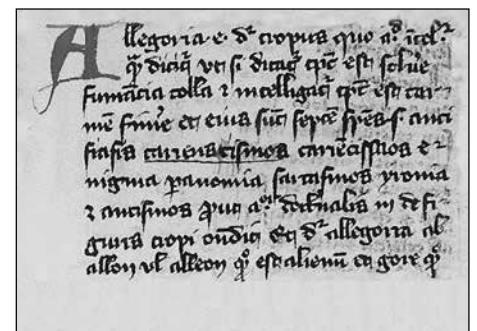
An example of such a glossary is the so-called *Reichenauer glosses*, a collection of a few thousand Bible glosses dating from the 8th century, explained in Middle Latin and Roman. In the first part of the manuscript, the glosses are ordered in the order of appearance in the Bible, the second part shows the glosses in a more or less alphabetical order, somewhat more suited for educational purposes, but still not ideal as the words were arranged under their initial capital without any further alphabetisation. The next step therefore was to put the words in the glossary in strict alphabetical order, as can be seen in the thirteenth-century *Glossarium Bernense*. This glossary shows the Latin entries in alphabetical

order, followed by explanations or short descriptions in Dutch. It is, however, not an original piece of work but an adaptation of an older Latin-Latin glossary. The same goes for the *Glossarium Trevirense* (14th century) and the *Glossarium Harlemense* (ca. 1450).



Glossarium Bernense (ca. 1225)
BB Bern Hs. 641

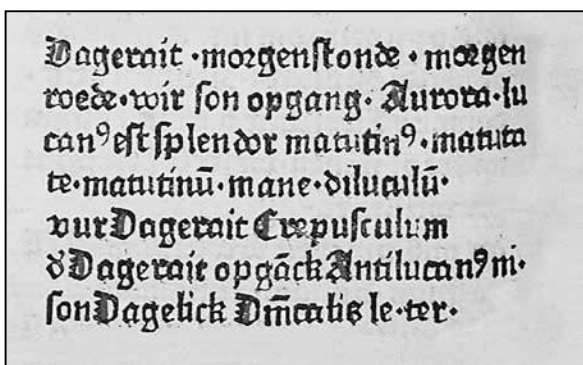
As time went by, glossaries became bigger and more elaborate, both with regard to the amount of entries as well as in the detailed elaboration of the definitions. The oldest known dictionary which has a clear Dutch input is the *Brachylogus*, made by Jan van der Meeren or Joannes de Mera from Brabant. This dictionary is the adaptation of the twelfth-century *Liber derivationem* by Hugucio de Pisa. Van der Meeren added a lot of entries and expanded the already existing entries as much as possible with extra information. His main goal was to produce an adequate compendium as an aid for reading and teaching Latin.



Joannes de Mera: *Brachylogus* (1390-1410)
UB Utrecht, Cat. 822.

The *Vocabularius copiosus*, which has been preserved in a copy from ca. 1480 but originally probably dates from around 1400, was meant for more advanced students of Latin. In the prologue, a number of authors are mentioned whose works were used in compiling the dictionary. This is for instance the case with the famous scholar Isidorus of Sevilla. The Latin entries are accompanied by an indication of inflection, a synonym or paraphrase of the word in Latin and/or in the vernacular and a short explanation. These vocabularies became very popular and were among the earliest books printed in press.

In 1477, the *Vocabularis qui intitulator Teuthonista* of Gherard van der Schueren appeared. This Dutch-Latin / Latin-Dutch dictionary was based on a collection of Latin sources including the above mentioned *Liber derivationem*. It was meant to serve administrative personnel, used more to Dutch than to Latin and satisfied the need for a dictionary that also provided Dutch entries. As such, it was the first Dutch dictionary that placed the vernacular at front. It was ordered alphabetically and etymologically, which meant that compounds and derivations were included in the description of the simplex words and that only these simplex words were alphabetised. Primarily the entries contain synonyms, Latin equivalents and antonyms and information on inflection. Subsequently the compounds and derivations are given, together with the same lexical and grammatical information as is listed for the simplex words.



Gherard van der Schueren: *Vocabularis qui intitulator Teuthonista* (1477)
UB Utrecht, B fol 130 Rar: ex 1.

Besides alphabetically-ordered dictionaries, systematically-ordered ones also began to appear. In these so-called *vocabularia rerum* the vocabulary is ordered and treated by subject. We distinguish two kinds of systematic dictionaries, those that take God and the creation as the starting point (*de deo*) and these that start from mankind and its

place in the world (*de homine*). In this last category the vocabulary is divided into topics such as 'the human', 'illness', 'clothing', 'animals', 'trees and shrubs', 'the house', 'instruments and tools', etc.

Medieval lexicography in the Low Countries consisted mainly of the production of alphabetical and systematical vocabularies for educational purposes, i.e. the study of Latin. These were based on the work of early-medieval authors such as Isidorus of Sevilla and Hugucio of Pisa. The target consisted of beginning and advanced students or, as in the case of Van der Schueren, of clerks who needed a practical knowledge of Latin for carrying out their duties.

The humanist period

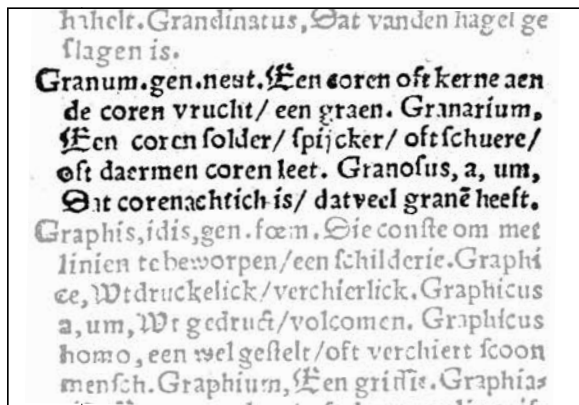
In 1531 a dictionary appeared in France that would be an example for many dictionaries in the surrounding countries: the *Dictionarium seu Latinae Linguae Thesaurus*, created by the Parisian printer and publisher Robert Estienne. His work meant a methodological innovation in lexicography; it contained a lot more examples and idioms than former dictionaries. The goal of the dictionary was to give a description of Latin that was as complete as possible. Although Estienne still used etymological classification, with the treatment of compounds and derivations directly under the simplex word, he added references at their alphabetical position in the dictionary as well. However, the most remarkable thing he did was to explain the Latin entries in French, which made them also comprehensible for non-Latinists. On the basis of these French word explanations he compiled the first translation dictionaries, the *Dictionarium Latinogallicum* in 1538 and *Dictionnaire francoislatin* one year later.

The example of Estienne was copied by humanists throughout Western Europe. Rapidly more bilingual dictionaries appeared, such as the *Dictionarium Latinogermanicum* (1535) by Peter Hasenfuss from Switzerland, in humanistic circles better known as Dasypodius. This dictionary, intended for use in schools, was re-edited for Dutch in 1542 by Fleming Antonius Schorus, maintaining the original title. A series of edited translation dictionaries followed, such as the *Dictionarium triglotton* (1545) and the *Pentaglottos* (1545) and the *Dictionarium trilingue* (1549).

INL – the Institute of Dutch Lexicology

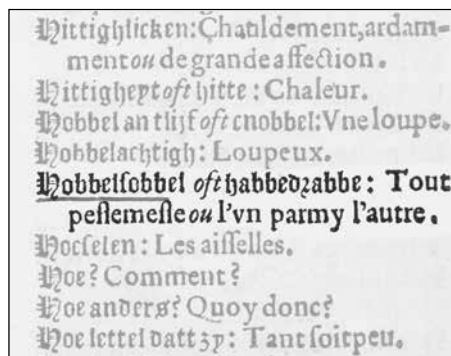
The initials INL stand for **Instituut voor Nederlandse Lexicologie**, the Institute of Dutch Lexicology. It is the place where anyone can go to learn about Dutch words, their spelling, form, meaning or usage throughout the centuries. INL collects and describes the Dutch language and its vocabulary from the 5th and 6th centuries to the present day, including all new words. One of its core tasks is therefore the collection and enrichment of texts in corpora. The language in which these texts are written ranges from the earliest Old Dutch to the most recent standard language as it is used in the Netherlands, Flanders and Suriname.

INL produces and maintains historical and contemporary dictionaries of Dutch, which are all available online: the *Oudnederlands Woordenboek* (ONW, Dictionary of Old Dutch), the *Vroegmiddelnederlands Woordenboek* (VMNW, Dictionary of Early Middle Dutch), the *Middelnederlandsch Woordenboek* (MNW, Dictionary of Middle Dutch) and the *Woordenboek der Nederlandsche Taal* (WNT, Dictionary of the Dutch Language) can be consulted separately or in combination via <http://gtb.inl.nl/>, whereas the *Algemeen Nederlands Woordenboek* (ANW, Dictionary of Contemporary Dutch) can be found at <http://anw.inl.nl/>. More information about INL, its products and its activities can be found at <http://inl.nl/> or via secretariaat@inl.nl.



Petrus Dasyposius & Antonius Schorus (ed.):
Dictionarium Latinogermanicum (1535/1542)
UB Utrecht X qu 83/1

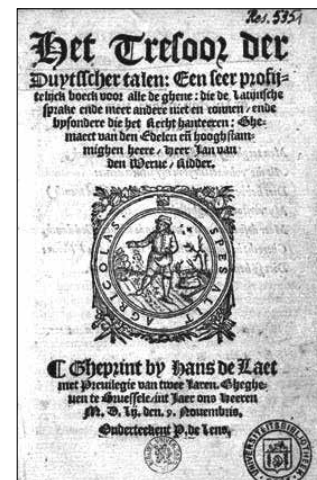
The first translation dictionary containing Dutch entries appeared in 1546. The *Naembouck* [Book of Names] by Joos Lambrecht, a printer from Ghent, was a Dutch-French dictionary meant for students and containing many words from everyday life. Loanwords were explicitly banned. The second edition appeared in 1562. The structure of the *Naembouck* was copied by Plantijn and Kiliaan, who also borrowed much of the content of the book for their own dictionaries.



Joos Lambrecht: Naembouck (1562)
Museum Plantin-Moretus Antwerp R 28.12

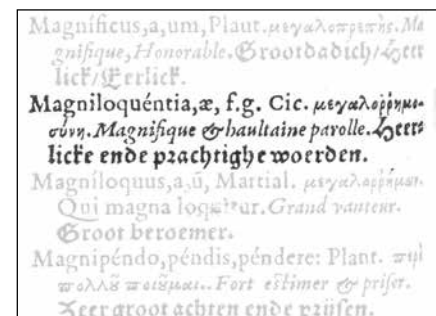
Language purification was an important subject in the Dutch scholarly circles of the sixteenth century. Many scholars believed that the influence of the French language had to be reduced and dictionaries played an important role in this. In 1553 the Antwerp lawyer Jan van den Werve published *Het Tresoor der Duytsscher talen* [The Treasure of the Dutch Language]. This was the first dictionary with an explicitly purist aim. Van den Werve wanted to purify the Dutch language from all foreign words. The *Tresoor* is an alphabetical list of French and Latin loanwords followed by one or more Dutch words which were recommended instead. However, many of the words that Van den Werve rejected still exist in present-day Dutch, such as *absoluut* [absolutely], *artiest* [artist], *bibliotheek*

[library], *congres* [congress] and *delicaat* [delicate]. It became a very popular book; the second edition appeared in 1559 under the purified name *Den Schat der Duytsscher talen* and until 1774 the book was reissued 18 times.



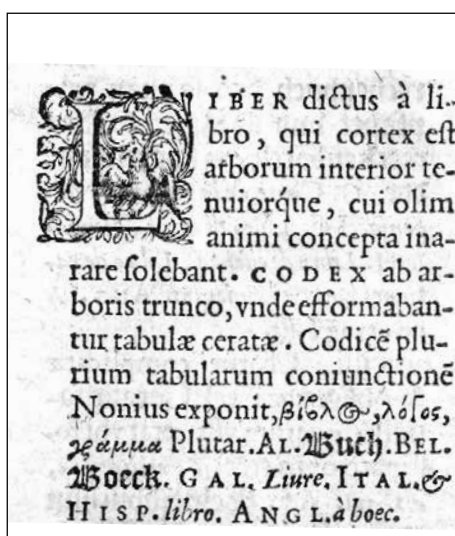
Jan van den Werve: Het Tresoor der Duytsscher talen (1553)
UB Ghent R 535

In 1555 the Frenchman Christoffel Plantijn started a printing and publishing house in Antwerp, *De Gulden Passer*. In the course of the sixteenth century, this company, also known as *Officina Plantiniana*, became one of the most important humanistic centres in the Low Countries. Being an immigrant himself, Plantijn recognized the importance of Dutch dictionaries of high quality. Since such dictionaries were hardly available, he decided to take this task with the dictionaries of Estienne for French as an example. In 1557 he published the *Vocabulaire francois-flameng* by the Antwerp schoolmaster Gabriel Meurier, based on the work of Estienne. Some years later in 1562, the *Dictionarium tetraglotton* appeared, a Latin-Greek-French-Dutch dictionary, also modelled on Estienne's work. The editor of this dictionary of approximately 20,000 entries was probably the famous etymologist Cornelis Kiliaan, also known as Kiliaan. Plantijn himself wrote the preface.



[Cornelis Kiliaanus:] Dictionarium tetraglotton (1562)
UB Ghent 121 H 44

In 1567, Plantijn published the *Nomenclator omnium rerum* by Hadrianus Junius. This dictionary orders the Latin entries systematically instead of alphabetically on subjects such as fishes, birds, trees or illnesses. In his *Nomenclator*, Junius introduced some innovations. He was, for instance, the first to add information on regional variation in Dutch and he also added equivalents in Dutch, as well as in other languages, such as Greek, German, French, Italian and Spanish. At the end of the *Nomenclator*, he added an alphabetical index. In 1928, the systematical approach of the sixteenth-century *Nomenclator* by Junius was still used (be it in an extended form) in L. Brouwers' *Het juiste woord*.



Hadrianus Junius: *Nomenclator omnium rerum* (1567)

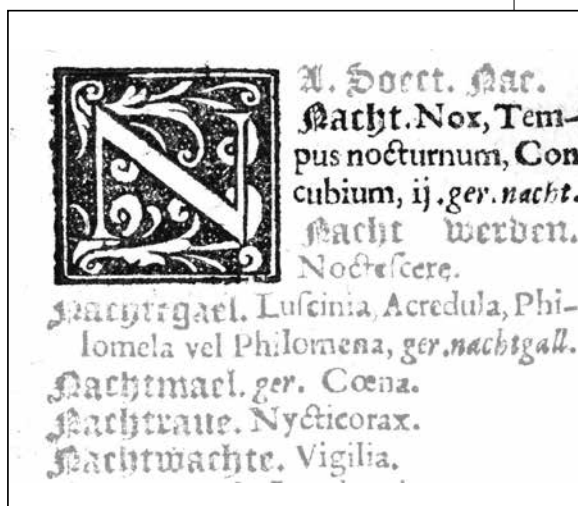
Museum Plantin-Moretus Antwerp A 442

The first translation dictionary from the publishing house of Plantijn with Dutch as the source language was the *Thesaurus Theutonicae linguae* or the *Schat der Neder-duytscher spraken* [Treasure of the Low German language] of 1573. French and Latin equivalents were used to explain the approximately 40,000 Dutch entries. These entries were collected from all existing alphabetically ordered Dutch dictionaries, such as Van der Schueren, Lambrecht, Meurier and Schorus. The dictionary appeared anonymously, but had according to Plantijn himself four co-authors amongst whom the aforementioned Kiliaan. It was the first attempt to give a full insight into the wealth of the Dutch language, following again the example of Estienne for French. In this huge dictionary we find Dutch words such as *giechelen* [to giggle], *hunkenen* [to yearn], *kiekendief* [duck-hawk], *overgrootmoeder* [greatgrandmother], *slager* [butcher] and *tegenstelling* [contrast].



Thesaurus Theutonicae linguae (1573)
Museum Plantin-Moretus Antwerp R. 46.16

In 1574, the first dictionary appeared that formally carried the name of Kiliaan, the *Dictionarium Teutonico-Latinum*. This dictionary introduced a thorough renovation of Dutch lexicography. With no more than 12,000 entries, it was relatively small, but methodologically the progress was huge because of the fact that Kiliaan, for instance, included cognates from other languages. Kiliaan used Latin, the former language of science, to explain the meaning of the entries, but where his predecessors used these explanations for a better understanding of Latin, it was Kiliaan's goal to achieve a better understanding of Dutch.



Cornelis Kiliaan: *Dictionarium Teutonico-Latinum* (1574)
Museum Plantin-Moretus Antwerp A 3928

The second edition of his dictionary appeared in 1588 and counted almost three times as many entries as the 1574 edition. In this edition he implemented some more lexicographical innovations. He used other

dictionaries as source material as did his predecessors, but he added material from scientific and historical studies as well. He paid a lot of attention to the dialect of Brabant, the region he originated from. He took his own dialect as the starting point to which he related and contrasted words from other dialects, which he – more elaborate than Junius who had done this before – provided with regional labels such as Flemish, Hollandish, Frisian or Low Saxon. Finally, he took the Roman loanwords from the main text and placed them in a separate word list at the end of the dictionary.

In 1599, the third edition of the dictionary appeared bearing the new title *Etymologicum Teutonicæ linguæ*. This is actually the first scientific etymological dictionary of Dutch. All approximately 40,000 entries are systematically provided with cognates from French, English, German, Low German, Gothic, Spanish, Italian, Greek and Hebrew. This made Kiliaan the first European who systematically compared various languages. The main part of the entries contains an etymological explanation as well. During the following centuries, the dictionary was reprinted many times and the entries that Kiliaan selected were used in all bilingual dictionaries of the seventeenth and eighteenth centuries.

plücken. Carpere, decerpere, vellere, legere, *ang* plucke,
 plücken de bychten. Legere poma, legumina.
 pluck-graen. Legumen.
 pluck-haeren. Euellere capillos.
 pluck-voghel. *fris.* Parasitus, alienis
 inhians patinis: infamis propter rapacitatem, & impudentiam.
 pluck-bincken plocke binckē pluckede-bincken. *sax. fris. hol.* Minutal.
 pluckelinge i pluckfel. *Carpora.*

Cornelis Kilianus: *Etymologicum Teutonicæ linguæ* (1599)
 GA Duffel K2

In the humanist period, dictionaries were extended with idioms and more examples. Although education in Latin was still the main goal of the lexicographers, dictionaries also became more and more practical instruments for language purification, as can be seen in the works of Joos Lambrecht and Jan

van den Werve. The publishing house of Christoffel Plantijn played an important role in the emancipation of Dutch lexicography, by publishing dictionaries as the *Vocabulaire francois-flameng*, the *Thesaurus Theutonica linguæ*, the *Nomenclator omnium rerum* and of course the dictionaries of Kiliaan, in which the focus slowly shifted from Latin to Dutch.

From the Renaissance to present day lexicography

The work of Kiliaan inspired others to become lexicographers of Dutch. The first monolingual dictionary of Dutch appeared in 1681. The *Seeman* [Seaman] by Wigardus van Winschooten is a dictionary on the jargon in the field of shipping. Innovative is the fact that Van Winschooten describes all meanings of the keywords in his dictionary, both the literal and the figurative ones. He also adds collocations, idioms, dialect variants and etymological explanations. This dictionary contains far more linguistic information on Dutch than any other dictionary from the Renaissance.

SCHIJTVAL (k naa het getuigenis van de feelui, die om de noord vaaren) is een soort van roof voogels, de *feenvalken* niet ongelijk: deese *walk* heeft de naam daar van gekreegen, dat hij niet als andere *valken* aald op het vlees van het gevoogelte: maar (met verlof) op den drek van kokmeuwen: weshalven de feelui gewoon sijn te roepen, als sij sien, dat een meeuw van den soogenaamde *schijtwalk* gegreepen is:

Schijt meeuw! schijt! soog raakje de walk quijt.

hier uit is nu ligtelijk af te meeten: waarom formige pronkertjes (als bij gelijkenis tot Amsterdam) *schijtvalken* genaamd werden: te weeten, om dat sij de uiterlijke gedaante hebben van rijke luiiden, en sij hebben qualijk en dult in de mars.

Wigardus van Winschooten: *Seeman* (1681)
 UB Ghent Acc 691

In the beginning of the 18th century, a revival of lexicography took place in the Netherlands. Many large bilingual dictionaries appeared, containing not only more entries than their predecessors, but also more information per entry. One of the lexicographers who introduced some important innovations in his work was William Sewel, who in 1708 published *A Large Dictionary, English and Dutch*.

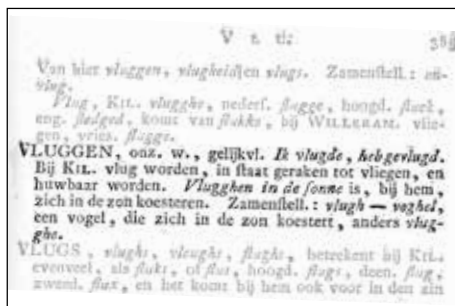
He is, for instance, the first to add information on the genus of substantives. He also clarifies the difference between the various meanings of a word by adding a definition in Dutch or a synonym, a compound or an example sentence. This is a new step on the way to a monolingual dictionary of Dutch. Sewel also mentions the style register of the words and he especially marks the words that are outdated or have a foreign origin.



W. Sewel: *A Large Dictionary, English and Dutch* (1708)
KB The Hague 651 E 18

In the second half of the 18th century the demand for a monolingual dictionary of Dutch arose. Such dictionaries were available in Italy since 1612 and in France since 1694, and the Netherlands could no longer fall behind. All kinds of private initiatives were undertaken to fulfil this demand, but the plan appeared to be too ambitious for that time. All that was made was a collection of source material for such a national dictionary that was to be put under the care of the *Maatschappij voor Nederlandse Letterkunde* [Society of Dutch Literature], which was founded in 1766.

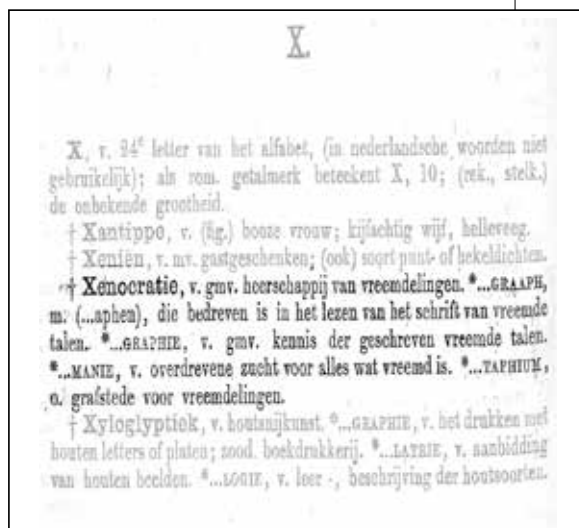
When at the end of the 18th century the Netherlands came under the government of France, the regulation of the language was supported officially. Reverend and linguist Petrus Weiland took advantage of these new circumstances and published between 1799 and 1811 the *Nederduitsch Taalkundig Woordenboek* [Dutch Linguistic Dictionary], a dictionary of eleven volumes, modelled on the *Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart* of Adelung (1793-1801). Weiland made use of the source material of the *Maatschappij* for editing his dictionary.



Petrus Weiland: *Nederduitsch Taalkundig Woordenboek* (1799-1811)
UB Ghent 121 H 44

In 1864, brothers-in-law I.M. Calisch and N.S. Calisch published the first monolingual concise dictionary of Dutch, the *Nieuw woordenboek der Nederlandsche taal* [New Dictionary of the Dutch

Language]. This dictionary marks the beginning of a long tradition. The second edition (1872) was revised by Johan Hendrik van Dale, and it is under his name that the dictionary still exists. In the course of time, the dictionary grew to three huge volumes, to which it owes its nickname 'dikke Van Dale' [fat Van Dale]. The 15th edition is due to appear in 2015. Nowadays the dictionary is not only available in print, but also on CD-ROM and, of course, on the internet (<http://vandale.nl>).



I.M. & N.S. Calisch: *Nieuw woordenboek der Nederlandsche taal* (1864)
SB Antwerp C 13310 [S3-893 f]

Since the dictionary of Weiland did not meet the scientific lexicographical demands, the participants of the first Dutch Congress of Language and Literature in 1830 decided there was need for a national dictionary of the Dutch language. This *Woordenboek der Nederlandsche Taal* [WNT; Dictionary of the Dutch Language] was to be modelled on the *Deutsches Wörterbuch* of the Grimm brothers (1838-1960). In 1851, professor Matthias de Vries accepted the task of compiling this dictionary. Prior to its publication, the first need was to design a standard spelling for Dutch. This task was fulfilled by L.A. te Winkel, who in 1863 published *De grondbeginselen der Nederlandsche spelling* [Principles of Dutch Spelling] and in 1866 the *Woordenlijst voor de spelling der Nederlandsche taal* [Glossary for the Spelling of the Dutch Language], which is still used as a guide for official Dutch spelling, be it regularly revised and updated (new edition to appear in 2015).

Work on the dictionary took considerably more time than the forty years that de Vries originally calculated. The first instalment appeared in 1864, and the first volume in 1882. In 1967, the project was institutionalised in the

Matthias de Vries Genootschap



In 1851, Matthias de Vries presented his Plan for a Dutch dictionary. This was the beginning of a project that was to last 147 years. De Vries belonged to the first team of editors of the *Woordenboek der Nederlandsche taal* (WNT), the first instalment of which appeared in 1864. The first volume saw the light of day in 1882. In 1998 the publication of the final volume was celebrated. Nine years later, in 2007, the largest dictionary in the world went online, on <http://gtb.inl.nl/>. None of this would have been possible without the efforts and perseverance of Matthias de Vries. The society of lexicographers that was established in the Dutch-speaking area in the late twentieth century could therefore only be named after this founding father of Dutch lexicography: the Matthias de Vries Genootschap (MdV, Matthias de Vries Society). The society began to publish a journal called *Trefwoord* [keyword, entry word], which is now an e-magazine available at <http://www.fryske-akademy.nl/trefwoord/>. The e-magazine is open to new articles, but it also contains all articles previously published in the paper journal. On top of that, it re-publishes

Instituut voor Nederlandse Lexicologie [Institute of Dutch Lexicology]. During one and a half centuries, five generations of editors worked on the dictionary, which contains more than 300 000 entries to be found in sources dating from 1500 until 1976. In 2001, the last three volumes appeared and the dictionary was completed. In the meantime, the dictionary had been made available on CD-ROM and since 2007 it can also be found on the internet (<http://wnt.inl.nl>). The internet version has certain advantages over its predecessors. It contains for instance keywords in modern spelling and links to sources, pictures and etymological and dialectal information.

WARTEL (IV), znw., gesl. onduidelijk en mv. niet aangetroffen. Herkomst duister. Veroud. — Lendendoeck. || Voor kleine misgrepen zet men hun (*t.w. de schepelingen*) een wartel om het lijf, en laetze zoo tot spot van anderen, het schip schoon vegen, WITSEN, *Scheepsb.* 392 a [1671]. *Zy* (*t.w. de bewoners van het dorp Savacca*) zyn van aart even eens, als de Manadosche berg-boeren, gaande de mans mede naakt, maar met een wartel om den middel, VALENTIJN, *O.-I.* 1, 2, 77 a [1724]. De Ternataanen, die als de Wilden nog maar met een Tsjidako, dat is, een Gordel, of Wartel van oude lappen, om 't lyf liepen, I, 2, 140 b [1724].

M. De Vries (e.a.): *Woordenboek der Nederlandsche Taal* (1864-2001)
Instituut voor Nederlandse Lexicologie, Leiden

The *Woordenboek der Nederlandsche Taal* was succeeded by the *Algemeen Nederlands Woordenboek* [ANW; Dictionary of Contemporary Dutch], also a project of the Instituut voor Nederlandse Lexicologie. In this online dictionary (<http://anw.inl.nl>), which is still under preparation, the vocabulary of contemporary Dutch from 1970 onwards is described. The lexicographical work is no longer done

alphabetically, but thematically. In 2009, the first entries were published online and since then new entries are added regularly, providing the users access to the most recent lexicographical information. At the moment, many words concerning occupations, animals, drinks, vessels and diseases can be found in the dictionary, alongside many neologisms.

The information in the entries of the ANW takes into account the new search possibilities computers offer. With the help of definitions and so-called semagrams, it is not only possible to find meanings to words but also to find words on the basis of their characteristics. A user who is, for instance, looking for words that indicate a crossing for wildlife, does not only find the entry *wildviaduct* but also a lot of synonyms such as *cerviduct*, *dierenpassage*, *ecoduct*, *ecoviaduct*, *natuurbrug* and *wildwissel*.

Summary

In the course of time, Dutch lexicography has developed into a mature branch of science, sometimes following then anticipating trends in other countries. Out of the earliest glosses and glossaries, vocabularies and dictionaries arose which offered more information than the meaning of a single word only. From the Renaissance onwards, etymological information was added and derivations and compounds based on the keyword were included and explained. The focus shifted from Latin to Dutch entries and information on parts of speech and inflections were added.

Kiliaan brought about a great leap forward by systematically comparing Dutch words with their equivalents in other languages. He gave the dictionary a scientific status through the critical

The screenshot shows the ANW website interface. The main heading is 'lexicologie'. Below it, the entry for 'lexicologie 1.0' is displayed, including its definition as 'wetenschap die de woordschat bestudeert'. The entry also lists 'Uitspraak' (Aantal lettergrepen: 5; Plaats hoofdklemtoon: 5de lettergreep) and 'Semagram' (is een wetenschap; wordt beoefend om bv. een woordenboek te kunnen maken; wordt meestal door taalkundigen beoefend; betreft de woordschat). A 'Combinatiemogelijkheden met adjectief ervoor' section lists 'cognitieve lexicologie; historische lexicologie; Nederlandse lexicologie'. The right sidebar provides 'Woordsoort' (Type: substantief; Naamtype: soortnaam; Geslacht: vrouwelijk; Lidwoord: de; Getal: geen meevoud; Betrekkingsklassen: abstractum) and 'Spelling en flexie' (Enkelvoud: lexicologie [lexi.zo.3o.ge]).

Algemeen Nederlands Woordenboek (2009 -)
Instituut voor Nederlandse Lexicologie, Leiden

and deliberate use of lots of new source material. In the 17th and 18th centuries, renewals of the description of all meanings of a word, the bright and clear separation of these meanings and the addition of example sentences and quotes contributed to the evolution of the dictionary. In the course of the 19th century, the demands of a scientific historical dictionary of Dutch were formed on the basis of all these elements, which resulted in the *Woordenboek der Nederlandsche Taal*. For the *Algemeen Nederlands Woordenboek*, these demands have been modified and adapted to the requirements of modern times and contemporary users where good use is made of the new possibilities of the digital age.

References

- Reichenauer glossen end 8th century.** Badischen Landesbibliothek Karlsruhe, codex Augiensis CCXLVIII.
- Orosiusglossen ca. 1150-1200.** Bibliothèque municipale Boulogne-sur-Mer, BM 717
- Glossarium Bernense ca. 1225.** Bürgerbibliothek Bern, ms. 641.
- Glossarium Trevirensis 14th century.** Stadtbibliothek Trier, ms. 1125.
- Glossarium Harlemense ca. 1450.** Stadsbibliotheek Haarlem, codex 183 D2. *Vocabularius ex quo* 1479. Zwolle: Peter Os van Breda.
- Vocabularius Copiosus et singularis unus ex diversis, diligentissime theutonicatus ca. 1481-1483.** Leuven: Jan van Westfalen.
- Johannes de Mera, Brachylogus 1390-1410.** Universiteitsbibliotheek Utrecht, Cat. 822.
- Gherard van der Schueren, Vocabularis qui intitulator Teuthonista 1477.** Universiteitsbibliotheek Utrecht, B fol 130 Rar: ex 1.
- Robert Estienne, Dictionarium seu Latinae Linguae Thesaurus 1531.** Paris: Robert Estienne.
- Robert Estienne, Dictionarium Latinogallicum 1538.** Paris: Robert Estienne.
- Robert Estienne, Dictionnaire francoislain 1539.** Paris: Robert Estienne.
- P. Dasypodius, Dictionarium Latinogermanicum 1535.** Strassbourg: Wendelinus Rihelius.
- P. Dasypodius, Dictionarium Latinogermanicum 1542, ed. Antonius Schorus.** Antwerpen: Gulielmus Montanus.
- J. Servilius, Dictionarium triglotton 1545.** Antwerpen: Christophorus Plantinus.
- A. Calepinus, Pentaglottos 1545.** Antwerpen: Joannes Gymnicus.
- Johannes Brechtanus Custos, Dictionarium trilingue 1549.** Antwerpen: apud Martinum Nutium.
- Joos Lambrecht, Naembouck 1562.** Gent: Joos Lambrecht.
- Jan van den Werve, Het Tresoor der Duytsscher talen 1553.** Antwerpen: Joannes Latus.
- Jan van den Werve, Den Schat der Duytsscher talen 1559.** Antwerpen: Joannes Latus.
- Gabriel Meurier, Vocabulaire francois-flameng 1557.** Antwerpen: Christophe Plantin.
- [Cornelis Kilianus], Dictionarium tetraglotton 1562.** Antwerpen: Christophorus Plantinus.
- Hadrianus Junius, Nomenclator omnium rerum 1567.** Antwerpen: Christophorus Plantinus.
- L. Brouwers, Het juiste woord 1928.** Turnhout: Brepols.
- Thesaurus Theutonicae linguae 1573.** Antwerpen: Christophorus Plantinus.
- Cornelis Kilianus, Dictionarium Teutonico-Latinum 1574, rev. ed. 1588.** Antwerpen: Christophorus Plantinus.
- Cornelis Kilianus, Etymologicum Teutonicae linguae 1599.** Antwerpen: Officina Plantiniana-Joannes Moretus.
- Wigardus van Winschooten, Seeman, behelsende een grondige uitlegging van de Neerlandse Konst- en Spreekwoorden voor soo veel die uit de Seevaart sijn ontleend. 1681.** Leiden: Johan de Vivie.
- William Sewel, A Large Dictionary, English and Dutch 1708.** Amsterdam: Jacob ter Beek.
- Petrus Weiland, Nederduitsch Taalkundig Woordenboek 1799-1811.** Dordrecht: Blussé & Van Braam.
- Adelung, Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart 1793-1801.** Wien: Anton Pichler.
- I.M. Calisch & N.S. Calisch, Nieuw woordenboek der Nederlandsche taal 1864.** Tiel: H.C.A. Campagne.
- Van Dale's Groot Woordenboek van de Nederlandse Taal 14th ed. 2005.** Utrecht: Van Dale Lexicografie.
- WNT = Woordenboek der Nederlandsche Taal 1882 – 1998.** 's-Gravenhage/Leiden: Martinus Nijhoff/Brill. *Supplement I* 1956. 's-Gravenhage/Leiden etc., Martinus Nijhoff/A.W. Sijthoff etc. *Aanvullingen* 2001 's-Gravenhage: Sdu Uitgevers. <http://wnt.inl.nl/>
- ANW = Algemeen Nederlands Woordenboek 2009 - .** Leiden: Instituut voor Nederlandse lexicologie. <http://anw.inl.nl/>



articles on lexicography in the Netherlands that have been published elsewhere. Thus *Trefwoord* is developing into a repository of lexicographical articles.

Every year the MdV organizes a meeting at the premises of the Institute of Dutch Lexicology (INL, Instituut voor Nederlandse Lexicologie) in Leiden, where members can meet and enjoy two papers on lexicography in its widest sense. Shortly before last year's meeting Macmillan decided to stop publishing printed dictionaries. Ilan Kernerman, already invited to read a paper at the 2012 meeting, has been interested in electronic dictionaries for a long time and his paper happened to discuss these trends and show how his company, K Dictionaries, copes with them. Since Ilan's paper perfectly fitted into the Macmillan discussion, the organizers of the MdV meeting, Tanneke Schoonheim and Anne Dykstra, invited two Dutch dictionary publishers, namely Sander Bekkers (Uitgeverij Prisma) and Ton den Boon (Van Dale), to present their viewpoints on the matter and to take part in a forum discussion on the printed/online controversy. The forum led to an animated and focused debate with conflicting viewpoints. We are happy that Sander Bekkers submitted his paper for publication in KDN, along with the contribution by Jaap Parqui of Van Dale. I hope that it will give you an idea of what was being discussed at the MdV meeting last November 28.

Anne Dykstra
dykstraanne@gmail.com

The new *Lighthouse* dictionary

Kaoru Akasu



Kaoru Akasu is Professor of English Linguistics at Toyo University in Tokyo, and has been President of the JACET (Japan Association of College English Teachers) Society of English Lexicography for the past six years. His most current publications include 'The first dictionary of English collocations in Japan' in *Research on Phraseology in Europe and Asia: Focal Issues of Phraseological Studies* (2012), and the chapter 'Methods in dictionary criticism' in *The Bloomsbury Companion to Lexicography* (forthcoming). He is co-editor of the *Lighthouse English-Japanese Dictionary* 6e and *Luminous English-Japanese Dictionary* 2e.
kakasu@toyo.jp

The *Lighthouse English-Japanese Dictionary* (LHEJD) is targeted primarily at Japanese high-school students who are learning English in order to enter colleges and universities. The sixth edition of the dictionary came out, fully revised, in late October 2012. The first edition was published back in 1984, and each revision was made at regular intervals of six years until the fourth edition (2002). The publication dates of the fifth edition (2007) and this latest edition have a shortened interval of five years apiece. In fact, LHEJD's predecessor, Kenkyusha's *Union English-Japanese Dictionary*, had its very first edition in 1972 and the second in 1978: similar six-year intervals. *Union* was renamed *Lighthouse* in 1984, so this current edition constitutes virtually the eighth edition of the dictionary, which is thus more than forty years old. At the time, the *Union* was epoch-making in that it incorporated a grammar book into a learner's dictionary.¹ I remember the late Professor Emeritus Shigeru Takebayashi, one of the two original editors, telling me the name *Union* had a double meaning. On the one hand, it symbolized the concerted effort on the part of the editors and lexicographers involved or the union of the lexicographical work force, and on the other, it represented the union of grammar and the dictionary.

Nakao (1998: 42) observed that there are three different types of English-Japanese dictionaries: "full-sized comprehensive dictionaries having more than 100,000 entries and more than 2,000 pages"; "medium-sized ... editions of about 1,500 pages, containing from 50,000 to 100,000 entries"; and "E-J dictionaries for beginners". LHEJD belongs to the second category, and the market for this particular group has been, and still is, a fiercely competitive one in Japan. This situation is fairly comparable to that seen among British publishers concerning monolingual dictionaries for advanced learners of English. For this reason, editors and lexicographers of English-Japanese dictionaries at each and every publishing house have been making every effort to invent and innovate new features and new designs with each revision. As

illustrated by Nakao (1989), the grading of headwords, of meanings, and of grammatical information, a dual system of presenting meanings combining historical changes with the logical links of semantic flow, and modifications of verb patterns, among others, are just a few of the features employed.

As all these features are already included in LHEJD6, I will touch here only upon two new ones that we have introduced in this edition: 'Keys to Communication: Politeness' and 'Collocation Plus One'. As regards the former, we made an attempt to integrate into the dictionary a certain type of pragmatic information, which we believe will help Japanese learners use English in a more appropriate manner. 'Keys to Communication: Politeness' columns are provided to entries such as **actually, can, could, great, let's, maybe, sorry, suggest, Thank you, want, and why**, dealing with matters of complimenting, asking for permission, advising, apologizing, and the like.

Take a look at the entry for **possible** in the excerpt, where you will find a box with a [coloured] key symbol at the beginning. This is a case of a downtoner. What it says, translated into English, is basically as follows: 'Would it be possible for you to *do*? is a polite expression of request, less obtrusive by asking the interlocutor about their convenience or availability. This expression has an indirect ring to it and is most often used when you are not sure that they will accept (ie, when you are asking some big favour or when you don't know the interlocutor very well) ((For expressions of asking someone to do something → **could B 1 (4)**): "I need to go to ...").

Another 'Keys to Communication: Politeness' box appears in the entry **possibly**. Japanese natives find it very difficult to use English in terms of politeness because English has its own different system of polite language. We do have the so-called *keigo*, or honorific language in Japanese, which works very differently from English. Information of this kind is, therefore, of great practical use as well as of special significance to Japanese learners who wish to use the English language appropriately in actual contexts.

The other newly added feature,

¹ See Nakao (1989: 303-309) for further details.

‘Collocation Plus One’, is a supplementary gadget, since collocation information is already provided primarily by way of collocation boxes that were introduced in LHEJD3 (1996). Examples of ‘Collocation Plus One’ appear in the entries for **possible** and **possibly** beginning with the [coloured] symbol ‘C+1’. What it says in the former and in the latter is: ‘The following adverbs are often used to modify *possible*: technically *possible* / humanly *possible* / theoretically *possible*’ and ‘The form *A* and [or] *possibly B* expresses the stronger possibility of *B* than of *A*: We could predict the problem and *possibly* prevent it’, respectively.

In this revision we reviewed all illustrative sentences and phrases in the dictionary to make them more readable and easily accessible to the users, and expanded the number and coverage of ‘Remember Words’, a feature introduced in LHEJD3 to help learners to better remember certain sets of associated words by drawing attention to the original meaning of root forms. For example, **arrive**, **derive**, **river**, and **rival** are given in a box at the entry for **arrival**, with an explanation of the meaning of the root RIV.

It goes without saying that print dictionaries like LHEJD6 are now facing an increasingly difficult time.² Our firm belief, however, is that there is still not just a need, but a definite benefit, for those learners who choose to use print dictionaries when studying English because they have advantages over *denshi-jisho*, or hand-held electronic dictionaries, which are the most popular form of dictionary and reference books in Japan. These devices are now so sophisticated and packed with an astonishing variety of information, which is good in itself, that they have reached a point where a user might need, so to speak, more than two hands to use it. They have become too complex a tool to handle, rather than just an electronic version of a dictionary, while a print dictionary is fairly simple to use, with no restriction of small screen size.

We seek to make the most of the clarity and lucid organisation of print dictionaries by providing plenty of visual information or visually-orientated features, including charts, pictorial illustrations and

² The *Luminous English-Japanese Dictionary*, another learner’s dictionary by Kenkyusha, which may be deemed an advanced version of LHEJD, can be accessed online for free at the following site: http://www.kenkyusha.co.jp/modules/08_luminous/index.php?content_id=1/.

the like, in producing our dictionaries.

It seems to me that, in the UK in particular, bilingual lexicography is not such a major issue and does not attract due attention among learners as well as lexicographers, at least not to the extent that it does here in Japan. I have my own speculations that there are some plausible reasons behind this, mainly related to the marked difference in the English proficiency attained by average Japanese learners and their European counterparts, which promote the creation of bilingual dictionaries for English learners in Japan rather than the use of monolingual dictionaries.

Nakao (1998: 36) states that “[t]he development of dictionaries in Japan is heavily indebted to lexicographical scholarship in China, and subsequently in Europe and America”, while Cowie (1998) writes that “the monolingual EFL dictionary is in various respects indebted to the Japanese bilingual tradition”. I am of the opinion that both statements are true. We have had so much to learn from each other in the past, and, in view of the fact that we have much more dialogue and exchange between Japan and the rest of the world today, we will be more interdependent and our ties will be closer than ever in the future, too. I believe that, as far as bilingual lexicography is concerned, we have so much to offer.

References

- Cowie, Anthony P. 1998.** ‘Editor’s note’. *International Journal of Lexicography* 11.1.
- Nakao, Keisuke. 1989.** ‘English-Japanese learners’ dictionaries’. *International Journal of Lexicography* 2.4, 295-314.
- Nakao, Keisuke. 1996.** ‘A new trend in EFL dictionary: An interview with Mr. Jonathan Crowther’. *The Rising Generation* 141.11, 618-622 & 642.
- Nakao, Keisuke. 1998.** ‘The state of bilingual lexicography in Japan: Learners’ English-Japanese/Japanese-English dictionaries’. *International Journal of Lexicography* 11.1, 35-50.
- Takebayashi, Shigeru. 1974.** ‘Eigo-jisho o kangaeru’ [Some thoughts on dictionaries of English]. *The Rising Generation* 119.11, 758-759.
- Takebayashi, Shigeru. 1978.** ‘Problems in editing a learner’s English-Japanese dictionary’. *The Teaching of English in Japan*, 441-449. Tokyo: Eichosha.
- Takebayashi, Shigeru. 1988.** ‘Nikakokugo-jiten henshu ni okeru mondaiten’ [Problems in editing bilingual dictionaries]. *Philologia Anglica*, 467-475. Tokyo: Kenkyusha.



Lighthouse English-Japanese Dictionary, Sixth Edition

Shigeru Takebayashi, Nobuyuki Higashi, and Kaoru Akasu, eds.
Tokyo: Kenkyusha, 2012
XXIV+1796 pp.
202 x 140 x 52 mm
ISBN 978-4-7674-1506-2

possessive

in possession of ... [前] (格式) ...を所有して: He is in possession of confidential information. 彼は機密情報を所有している.

in the possession of ... = in ...'s possession [前] (格式) ...に所有されて: The island is in the possession of a rich businessman. その島はある金持ちの実業家が所有している. (動) possess

pos-ses-sive /pəzészv/ 所有の; 所有欲[独占欲, 支配欲]の強い; (自分の物を)人に使わせたがらない(of): She's possessive about her car. 彼女は車を他の人に使わせない. ② [文法] 所有の, 所有格の(⇒巻末文法3(1)). (動) possess

— ① [the ~] [文法] 所有格. ② [C] [文法] = possessive pronoun.

possessive pronoun [C] [文法] 所有代名詞 (⇒巻末文法3(1)).

pos-ses-sor /pəzész- /-sə/ [C] (格式) または [こっけいに] 持ち主, 所有者 (of).

pos-si-bil-i-ty /pə:sisəbələtɪ | pəs-/

— ① (-ties /-z/) ① [C,U] 可能性, 起こり[あり]うること; 実現性 [⇒ impossibility]: a strong [definite] possibility 大きな[はっきりとした]可能性 / the possibility of success 成功の可能性 / [置換] There's no [some] possibility of his winning the election. = There's no [some] possibility (that) he'll win the election. [that] 彼が選挙に勝つ見込みは全然ない[少しある]. [用法] to do は後に来ない // There's only a remote possibility that an earthquake will hit this town. この町に地震が起こる可能性はまずない / Agreement is still within the realms [bounds] of possibility. 合意の可能性はまだ残っている.

② [しばしば複数形で] やれること, 考えられること, チャンス (for): explore the possibilities 可能性 (やれることをさぐる) / exhaust all the possibilities やれることはみなやってみる / Take all the possibilities into consideration. 考えられることはすべて考慮に入れない.

③ [複数形で] 発展[改良]の可能性, 見込み: possibilities for growth 成長する可能性 / This invention has wonderful possibilities. この発明にはすばらしい将来性がある. (形) possible

pos-si-ble /pə:sisəbl | pəs-/

— ① 可能; 実行できる [⇒ impossible]: Is it possible for him to get there in time? 彼はそこへ間に合うように着くことができるだろうか / Would it be possible for you to lend me your office keys? 事務所の鍵を貸していただけませんか. (依頼) / The new technology has made it possible to repair the bridge within a week. 新技術のおかげで1週間以内に橋を修理することができた. / There's only one possible solution. 実行可能な解決法はたったひとつしかない. [C+H] 修飾する副詞には次のようなものが多い: technically possible 技術的に可能な / humanly possible 人間の力でできる / theoretically possible 理論的には可能な.

[用法] 「人が...できる」という場合に ×He is possible to get there. のような言い方はできない. He can't get there. や He is able to get there. のように言う.

Would it be possible for you to do? は, 相手の都合や可能性を聞くことにより, 押しつけを弱める丁寧な依頼の表現. 間接的な響きがあり, 相手が応じるとは予測できない場合(負担の大きいことを頼んだり, 親しくない人にお願する場合など)に使われることが多い(依頼の表現については ⇒ could B 1 (4)): "I need to go to L.A. this weekend. Would it be possible for you to lend me your car?" "Again? Why don't you rent a car?" "週末にLAに行かなきゃならないんだけど, 車を貸してもらうことはできるかな" "また? レンタカー借りたら?"

② 起こり[あり]うる, (もしかしたら) ...になりそうな[...かもしれない] (⇒ 類義語): a possible danger 起こりうる危険 / In this region snow is possible even in April. この地域では4月にも雪が降ることがある / It's possible that the roads will be crowded. 道路が混んでいるかもしれない (⇒ that A 2) / This tiny melon costs 5,000 yen. How is it possible? この小さなメロン 5000 円だなんて, こんなことである?

③ [最上級, all, every などとともに用いその意味を強めて] できる限りの, できるだけ...で: the worst possible situation 考えられる最悪の状況 / She spoke with all possible calmness. 彼女はできるだけ落ち着いて話した / The doctors made every possible attempt to save the patient. 医師たちはその患者を救おうとできる限りの手を尽くした.

as ... as possible [副] できるだけ... [置換] Try to exercise as often as possible (= as often as you can). できるだけ多く運動するようにしなさい / Read as many books as possible. できるだけたくさん本を読みなさい // ⇒ as soon as possible (soon 成句).

if possible [副] [文修飾] 可能なら (⇒ 依頼や許可を求めるときに使われる緩和表現; 「可能なら」という条件をつけることにより押しつけを弱め, 相手が断わりやすい配慮を示す): "When should I come to discuss it with you?" "If possible, I'd like to do so today. Could you drop by about 3 o'clock?" "いつ打ち合わせに伺えばいいですか" "できれば今日がいいのですが, 3 時ごろに来ていただけますか"

whenever [wherever] possible [副] 可能な限りいつでも[どこでも]. (名) possibility

類義語 possible は論理的に可能であるか, あるいは起こりうることを意味するのに対して, probable は確実ではないにしても, 統計上, あるいは状況から判断して大いにありそうなことをいう: Rain is possible but not probable before evening. 晩までに雨の可能性はあるが, まず降らないでしょう. likely は possible と probable の中間, 単に可能で起こりうるというだけではなく, ある程度の可能性はあるが, 確実性がそれほど高くはない場合に用いる: His horse is likely to win. 彼の馬が勝ちそうだ.

— ① [C] ふさわしい[もの]; 適任者, 有力候補.

pos-si-bly /pə:sisəblɪ | pəs-/

— ① [文修飾] ことによると, もしかすると, あるいは (⇒ likely 表): "What caused the accident?" "Possibly, the driver fell asleep." "事故の原因は何ですか" "もしかすると運転手が眠っていたのかもかもしれません" / "Do you think Ken will marry her?" "Quite possibly [Possibly not]." "ケンとは彼女と結婚するのかな" "十分ありうるね[しないかもね]" (⇒ not (5) (ii)). [C+H] A and [or] possibly B の形で, A より程度の高い可能性 B を表わすことができる: We could predict the problem and possibly prevent it. 我々はその問題を予測できるだろうし, もしかすると防ぐこともできるのではないかな.

could や can を使った依頼や許可を求める疑問文で, 押しつけを弱めるために possibly が挿入されることがある. 「ひょっとして...できませんか」という意味になり, 内容・状況的に難しいことをお願する場合には, 「相手が依頼に応じるのは当然」という態度を避け, 遠慮がちな姿勢を示す (⇒ could B 1 (4), (5), can 2 (1), (2)): "Could you possibly spare me a few minutes before the meeting?" "Sure, No problem." 「会議の前にできれば数分お時間をいただくとはいけませんか」「いいですよ」/ Could I possibly ask you to give me a ride? もしできたら車で送っていただけませんか.

[肯定文で can とともに] できる限り, 何とかして; [否定文で can とともに] どうあっても...できない, とても...でない; [疑問文で can や could とともに] いったい...できるだろうか: I'll come as soon as I possibly can. できるだけ早く参りましょう / I can't possibly come tomorrow. 私はどうしてもあすは来られません / How could you possibly speak to me like that? ⑤ 私に対してよくもそんな口がきけますね.

pos-sum /pə:zəm | pəs-/ [C] (略式, 主に米) = opossum. play possum [動] ⑤ (略式) 死んだ[寝た]ふりをする.

post /pəʊst/ [発音]

— ① (posts /pəʊsts/)

語義の展開 イタリア語で「替え馬の置いてある宿場」の意 → (早飛脚, 早馬) → 「郵便物」の意となった; post 2 と同語源.

① [U] または the ~ (英) 郵便, 郵便制度 [(米) mail]: put ... in the post ... を郵送する / Your parcel possibly got lost in the post. あなたの小包は郵送中に紛失したのかもしれない.

② [普通は the ~] (英) 郵便物 (全体); (1 回に配達される) 郵便物; 郵便の集配[配達](時間) [(米) mail]: "Has the post come yet this morning?" "No, not yet." 「けさはもう郵便が来ましたか」 「いえ, まだです」 / There's an unusual amount of post today. 今日は郵便物がやけに多い / I caught [missed] the last post. 最終便に間に合った[間に合わなかった]. ③ [P-; 新聞名で] ...ポスト: ⇒ Washington Post.

by post [副] (英) 郵便で [= by mail]: I'll send the book by post. その本を郵送します. (形) postal

— ① (posts /pəʊsts/; post-ed /-ɪd/; post-ing) ② (英) (...) を郵便で出す, 郵送する (off); (人に)(...) を郵送する; (手紙)を出す, 投函[送る] [(米) mail]: Post this letter, please. この手紙を出してください / [置換] I posted John that book this morning. [V+O+to+名] その本はけさ[今]に郵送した. keep ... posted [動] ③ ⑤ (人)に最新の事情[情報]を絶えず知らせる.

post /pəʊst/ [発音]

— ① (posts /pəʊsts/) ① [C] (格式) 地位; 職 (⇒ job 類義語): the post of personnel manager 人事部長の地位 / a teaching post 教職の口 / The authorities relieved her of her post as professor at the university. 当局は彼女をその大学の教授の地位から

Why do we need pattern dictionaries (and what *is* a pattern dictionary, anyway)?

Patrick Hanks and Jane Bradbury

Abstract

After a lifetime in lexicography, the first author reached the alarming conclusion that words don't have meaning. Does that mean that dictionaries are useless? No, far from it. We argue that, strictly speaking, the neat numbered definitions listed in dictionaries can be regarded as presenting **meaning potentials** rather than meanings as such. Meanings, we say, are events—events activated in a process in which context acts on the meaning potential of each word or phrase that is used. Ordinary dictionary users find dictionaries useful because they can use common sense to supply contextual information that the dictionary does not give explicitly. Computer software for NLP (natural language processing), on the other hand, has little or no common sense to draw on, and so is often baffled by problems of word meaning. Language learners are somewhere in between: some aspects of “common sense” are language-specific; others are universal. Work in recent decades on pattern grammar (e.g. Francis, Hunston and Manning 1996, 1997; Hunston and Francis 2000) and on construction grammar (e.g. Goldberg 1995, 2006) has shown that contributions to the meaning of utterances come from grammatical constructions as well as from individual words. Construction grammarians point out that the meaning of a sentence such as “she slept her way to the top” is quite clear; it is something like: she got a senior job by having sex with powerful men. However, this meaning cannot be deduced from a concatenation of the meanings of the individual words in the sentence. Instead, it is associated, at least in part, with the whole sentence, i.e. the construction as a whole. They argue, with interesting consequences for lexicography, that meaningful constructions such as this are pervasive in ordinary language. But just as a reductionist approach to words (treating words as if they were building blocks in a child's Lego set) is insufficient for an understanding of meaning in language, so also syntactic analysis of grammatical constructions tells only part of the story. Somehow, ways and means need to be found for expressing the conditions under which

different meaning potentials of a word are activated. In this article we shall suggest some of the ways in which this can be done. In short, we shall present a case for including much more information about phraseology as well as meaning in dictionaries.

1. Do words have meanings?

What's the meaning of **blow**? It could refer to what the wind does, or a bitter disappointment. Or it could be something you do with your fist, your nose, a whistle, or even a lot of money.

The *Macmillan English Dictionary for Advanced Learners* (Macmillan, 2002) offers more than fifty potential meanings of the verb **blow**, including phrasal verbs and idioms. There are another eleven potential meanings for the noun. Out of context, it is impossible to know which of these meanings is being activated; but given some context, things start to become clearer. Here are some sentences from the British National Corpus:

1. Use a fan to **blow** air through a screened doorway from the egg room or other work area into the main poultry house.
2. Arbroath has been dealt another jobs **blow**. The engineering firm of Giddings and Lewis is to make 50 workers redundant.
3. Officials said unidentified saboteurs also used a dynamite-packed petrol tanker to **blow up** a bridge near the town of Mostar.

In example 1, the infinitive marker **to** designates a verb, while use of the noun **fan** before the verb and **air** after it suggests that in this instance **blow** is being used to denote the process of air being moved around by a machine.

The determiner **another** in example 2 shows that **blow** in this sentence is a singular noun, so **jobs** (being a plural noun) must be being used as a modifier, while **blow** is the head of the nominal group. The only possible interpretation here is that **blow** is being used to mean some kind of disappointment (an interpretation that is elaborated in the sentence that follows). In example 3, the occurrence of **up** after **blow** narrows down the possible meaning, and the object that follows the verb, **a bridge**,



Patrick Hanks is Professor in Lexicography at the University of Wolverhampton, where he is directing a corpus-driven project in computational linguistics, the output of which is the on-line *Pattern Dictionary of English Verbs* (PDEV). He has developed a lexicographical technique called Corpus Pattern Analysis, used for PDEV and other projects.

He is also a Visiting Professor at the University of the West of England in Bristol, where he is the lead researcher on a project investigating family names in the UK (FaNUK).

From 1990 to 2000 he was Chief Editor of Current English Dictionaries at Oxford University Press. Before that (from 1971 to 1990) he was chief editor of English dictionaries for Collins publishers, and worked closely with John Sinclair on the first edition of the *Collins Cobuild English Language Dictionary* (1987).

After taking early retirement from OUP in 2000, he held teaching and research posts in the USA, Germany, and the Czech Republic. His monograph *Lexical Analysis: Norms and Exploitations* (MIT Press, 2013) presents a corpus-driven, lexically-based theory of meaning in language. patrick.w.hanks@gmail.com



Jane Bradbury is Research Associate in Lexicography at the University of Wolverhampton, where she is working on a corpus-driven project in computational linguistics, the *Pattern Dictionary of English Verbs*. She is also a consultant for Oxford University Press, working on the BBC/Oxford Children's Writing Corpus, as well as writing and editing educational resources. Jane began her career as a lexicographer in 1989, working for Cobuild at the University of Birmingham on a range of projects including the *Collins Cobuild Student's Dictionary*, the *BBC English Dictionary*, the *Collins Cobuild English Grammar*, and the *Collins Cobuild English Guides*, of which she was series editor. From 1997, she worked as a freelance lexicographer whilst training as a teacher of English in order to explore the possibilities of using corpora to develop resources for L1 students and teachers of English, and to explore children's use of language. She is now keen to apply Corpus Pattern Analysis to these areas of interest.
jane.bradbury@mail.com

confirms that here we are talking about a physical object being destroyed rather than a person losing their temper or inflating a balloon.

The contrasting meanings of *blow* in these examples illustrate that many words do not have meaning in isolation: rather, we are forced to say that they have *meaning potential*. We need to examine the context, and in particular the *collocations* of a word, to realize this potential and identify a unique meaning.

2. How do collocations shape meaning?

Collocations are co-occurrences of words near each other in any given text or (at a more general level) they are pairs or sets of words that *typically* co-occur in many texts. One of the most important findings of corpus linguistics has been that (while the number of *possible* co-occurrences of words is in principle infinite) the *actual* number of frequently recurring collocations associated with any given content word in any language is comparatively small. Collocations can be measured and processed lexicographically. Very often, they yield unique interpretations of words that, in isolation, have more than one potential meaning.

The idea that collocation is key to meaning is not new. The first edition of the Cobuild dictionary was accompanied by a book of essays by the lexicographers (Sinclair, 1987). In a chapter entitled 'The analysis of meaning', Rosamund Moon draws attention to the relationship between collocations and meanings:

Collocation ... frequently reinforces meaning distinctions The noun *gap* has four main meanings: a physical space, an interval of time, a deficiency, and a discrepancy. Each of these has a distinctive set of collocates. The physical space sense collocates with *mountain, teeth, in, and between*. The interval of time sense collocates particularly with *year* and *of* ...; the deficiency sense collocates with *fill, record, and in* ...; the discrepancy sense collocates with *close, poor, rich, widen, bridge, trade, generation, narrow, reduce, and between*. ...

Arguably, the only way to make distinctions in meaning or use within the major delexical verbs such as *have, give, and take*, is to split according to the type of object collocate. A further area where collocation supports – or enforces – meaning distinctions is that of verbs and the animate/inanimate identity of subject and object, or valency patterning.

In another chapter, 'The Nature of the Evidence', Sinclair observes:

Our initial assumption, that the words are distributed at random, is false.

He goes on to illustrate this with a discussion of corpus evidence for the distribution of collocations of the verb *set*, which has since been much quoted. Church and Hanks (1989 [1990]) used it as a basis for their work on statistical analysis of corpus data.

By 1998, after a further ten years of corpus analysis and growth of the Birmingham Corpus into what was to become 'the Bank of English', Sinclair had moved on to declare that "many, if not most meanings, require the presence of more than one word for their normal realization", and to argue that "patterns of co-selection among words ... have a direct connection with meaning". Nowadays, data from large corpora, extending to billions of words of text, confirm that word use is highly patterned. It is these phraseological patterns that give readers and listeners the contexts they need to activate the meaning of words. However, despite the initiative of Cobuild, patterns of word use in English and other languages have still not yet been satisfactorily identified or explained. In particular, more information about collocations needs to be given. Foreign learners in particular need to be given much more information than is customary in standard dictionaries about the normal phraseology with which each sense of each word is associated. Thanks to the technology of corpus linguistics, it is now possible to represent such phraseology systematically, although some variations may be expected, depending on the actual corpus and statistical measure(s) used to identify salient collocations.

2.1 Valency and collocation

Valency in language defines the number of syntagmatic arguments that go with a word. For example, the verb *shower* in *he showered* has a valency of one; in *he showered the dog* it has a valency of two; in *he showered her with gifts* it has a valency of three. It is sometimes difficult to distinguish between an optional adjunct and an adverbial argument. For example, few people would claim that *he showered her with gifts every day* has a valency of four. 'Every day' is a time adverbial which does not attach itself specifically to the verb *shower*. Instead, systemic grammarians prefer to say that time adverbials normally attach themselves to the general concept 'event verb', rather than affecting the meaning of any one specific verb.

For effective sense disambiguation, information on both collocations and valency is needed. More often than not, the relevant collocations are in a particular syntagmatic relationship with the target word. Hanks (2012) discusses the example of the verb *shower* in more detail: one sense

of this verb (broadly, ‘wash the body under flowing water’) can be clearly distinguished from other senses because it is intransitive and has a valency of one; however, other senses are less easy to separate on the sole grounds that they have the same number of arguments. For example, it is insufficient simply to report that *shower someone with praise* is transitive and has a valency of three. *Shower someone with rocks*, *shower someone with praise*, and *shower someone with gifts* all have a valency of three, however they have different meanings. To disambiguate these meanings effectively, we must look to both the syntagmatic patterns and the collocations (*rocks*, *praise*, or *gifts*). The point is that all three of these nouns are regular collocates of the verb *shower*: the different collocates activate different senses of the verb, which need to be explained specifically in dictionaries. Moreover, the different arguments correlate with one another: thus, an *explosion* can *shower* people or locations with *debris*, but no sentences have been found in which an *explosion* showers them with gifts or praise. This general approach to correlating arguments in order to get at the meaning is called triangulation.

Hanks (2013, chapter 5) shows that most meanings of most verbs and other words denoting events work in this way.

- **Firing** a person from a job has a different meaning from **firing** a bullet from a gun.
- **Filing** a lawsuit in a law court denotes activation of a process, whereas **filing** papers in a filing cabinet denotes cessation of active use of those papers.

In this paper, we propose that corpus evidence should be analysed by triangulation to group all normal uses according to their valency and syntax, for only then can a well-founded attempt be made to explain the meaning.

3. Why has no one made a pattern dictionary before?

The need for a dictionary that identifies and reports on patterns of syntax and collocation was established by Sinclair *et al.* in the 1980s (in the Cobuild project), and yet still no satisfactory pattern dictionary has been completed. This is because until very recently there was insufficient corpus data to provide an empirical basis for a reliable pattern dictionary. Let us look a little more deeply at the example of *shower*.

We have established that to disambiguate senses effectively, it is not enough to separate by valency alone. The next step is to look at patterns of adverbials and complementation, followed by patterns of collocation. Here, some delicate decisions

must be made by the lexicographer. For example, it is clear that *showering someone with presents* is different from *showering someone with praise*. This is because *presents* are (normally) physical objects, whereas *praise* is an eventuality activated by a person’s speech or actions. But *rocks* are physical objects too, so should *showering with rocks* and *showering with presents* be lumped together in the same pattern, or be split and dealt with separately?

A similar problem arises with *shower with praise* and *shower with abuse*. Both *praise* and *abuse* are eventualities activated by a person’s speech or actions; do they belong in the same pattern?

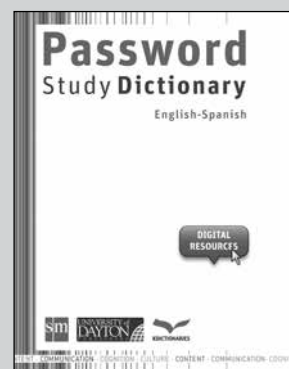
When you start to throw lexical items into the mix along with valencies, the decision as to whether to lump or split becomes difficult, but this is precisely the task that we are ambitiously undertaking.

4. Disambiguation of Verbs by Collocation

The Disambiguation of Verbs by Collocation (DVC) is an AHRC-funded project based in the Research Institute for Information and Language Processing at the University of Wolverhampton. The project aims, by doing Corpus Pattern Analysis, to establish an inventory of normal phraseological conventions, or *patterns*, for English verbs. For each of these verbs, an initial sample of 250 corpus lines is analysed and tagged to show which pattern they are typical of; this sample size is doubled where a verb is identified as having 10 patterns or more, and doubled again if the total number of patterns reaches 20. Each pattern is linked to a set of tagged corpus lines.

The key objective of DVC is to identify normal usage, or phraseological *norms*. A useful by-product of identifying these norms is that it draws attention to authentic uses of verbs which are not norms but which are one-off deliberate *exploitations* of established patterns, for example for literary or humorous effect. Exploitations (Hanks 2013) are deliberate irregularities in language use, which do not form part of a pattern and must be ruled out as lexicographical evidence.

DVC also allows the calculation of the relative frequency of each norm for each verb, shown as a percentage. An account is given of the meaning (semantic and pragmatic – we do not distinguish) associated with each phraseological norm, using a shallow ontology of semantic types. The DVC ontology is based on lexicographical need rather than received scientific theory. For example, there is no place in the ontology for a semantic type ‘mammal’, because there are no verbs in



PASSWORD Study Dictionary English-Spanish

Ediciones SM

Madrid, Spain

April 2013

Editorial coordination: Yolanda

Lozano and Concepción

Maldonado

Flexicover, 1088 pages

ISBN: 978-84-154-7867-6

From the series: **Kernerman**

Semi-Bilingual Dictionaries

Fortunes of National Cultures in Globalisation Context

The international conference on Fortunes of National Cultures in Globalisation Context was held by the Faculty of Eurasian and Oriental Studies at Chelyabinsk State University (CSU, Russia) on 4-5 April 2013. It embraced a wide range of topics concerned with cultural and globalization phenomena, including preservation of languages and national identities, linguistics, cognitive linguistics, international cooperation, etc. Considering that any culture represents an interdisciplinary object of research, the issues of ethnology, political studies, history, psychology, linguistic and cultural studies, philosophy, semiotics and lexicography were all discussed.

Plenary speakers included Nikolay Alefirenko (Belgorod State University), Konstantin Averbukh (Moscow City Pedagogical University), Vera Budykina (CSU, Conference Organizer), Olga Felde (Siberian State University), Zan Hun Ge (Capital University of Education, Beijing), Igor Golovanov (Chelyabinsk Pedagogical State University), Ilan Kernerman (K Dictionaries, Tel Aviv), Valery Kuznetsov (Moscow State Linguistic University), Elvira Sinetskaya (Institute for Oriental Studies of the Russian Academy of Sciences, Moscow), and Svetlana Ter-Minasova (Lomonosov Moscow State University and President of the National Association of Teachers of English in Russia).



English that select all and only mammals as arguments. On the other hand, as we shall see below, there are plenty of verbs that select 'horse' as an argument.

Case study: *harness*, verb

The DVC accounts for the normal patterns for *harness*, verb, as follows:

- 5% **[[Human]] harness [[Horse]]**
[[Human]] puts harness on [[Horse]] in preparation for riding or driving it, or getting it to pull a cart, carriage, or plough
- 95% **[[Human | Institution]] harness [[{Eventuality 1 | Entity 1} = Resource]] (to [[Eventuality 2 | Entity 2]])**
[[Human | Institution]] makes use of [[{Eventuality 1 | Entity 1} = Resource]] (in conjunction with [[Eventuality 2 | Entity 2]]) for some purpose

In Pattern 1, the lexicographer faces a dilemma that is a typical issue in DVC research. Prototypically, it is horses that get harnessed, but (as it happens) only 50% of the BNC citations for this pattern involve horses. The remaining 50% involve harnessing other animals: the British National Corpus (BNC) gives us the following examples of animals other than horses that get harnessed:

- dogs (huskies, for pulling sledges)
- oxen
- bullocks
- deer
- donkeys
- reindeer
- camels
- mules

When a speaker or writer talks about *harnessing* a bullock, reindeer, or mule, this is not a linguistic exploitation for effect; they are literally talking about the act of putting one of these animals into a harness in order to ride it, drive it, or get it to pull a cart etc. DVC must account for this regular alternation for the benefit of both language users and NLP applications. Therefore, it might be better to state Pattern 1 as **[[Human]] harness [[Horse | Animal]]**.

However, if **[[Animal]]** is given as an argument alternation of this pattern, the scope is too broad, as it could be taken as implying that it is normal to harness cats, primates, and cows, which is not correct. On the other hand, as we have seen, stating **[[Horse]]** alone is over-restrictive, appearing to rule out dogs, bullocks, oxen, etc. The answer to this apparently irresolvable dilemma is that, whatever semantic type (or set of types) is chosen, it is really only a form of shorthand, encapsulating a set of lexical items that are prototypical in this slot. Semantic typing is helpful as far as it goes, but it is possible to put too much weight on the type, as opposed to the actual lexical items that 'populate' the semantic type.

The DVC Ontology places the semantic type 'Animal' in a hierarchy, as follows:

```

Animate
  Human
  Animal
    Horse
    Dog
    Cat
    Primate
    Cow
  Bird
  Insect
  Fish
  Snake
  Spider
  Cetacean
  
```

Given this ontological set, by choosing the type **[[Animal]]** as an alternate for **[[Horse]]**, the lexicographer can signal that it is normal for other types of living creatures to be put into a harness (though not birds, insects, fish, or cetaceans, which are separate semantic types, associated with distinctive sets of verbs).

Pattern 2, which refers to the non-literal *harnessing* of abstract resources in order to use them, would once have been considered an exploitation:

[[Human | Institution]] harness [[{Eventuality 1 | Entity 1} = Resource]] (to [[Eventuality 2 | Entity 2]])

However, DVC has discovered that this pattern now accounts for 95% of uses of *harness*, verb, in this corpus: a clear example of an exploitation becoming a norm. It will be interesting to compare the relative frequencies of these two patterns in other corpora.

The example below shows a one-off exploitation of *harness*:

Perot wants to take us all back in time and *harness* us behind mules!

The writer is not suggesting that people will literally be forced to wear harnesses and pull carts behind mules: most readers will work out that this is a metaphorical extension of Pattern 1, with the intended meaning that Perot would treat people as no better than beasts of burden, valued for their physical strength only. However, in

other corpora, we may find that

[[Human 1]] harness [[Human 2]]
has become established as a pattern in its
own right in certain domains or in a more
recent time-frame than that of BNC.

5. Conclusion

The *Pattern Dictionary of English Verbs* (PDEV) represents a new development in lexical analysis, based on careful empirical analysis of a corpus. We hope that it will take its place alongside other innovative approaches such as FrameNet in accounting for words and meanings. It represents only one of many possible approaches to identifying and explaining patterns of word use and the connection between such patterns and their meanings. If it is successful, PDEV can function as a set of ‘seed’ patterns for semi-automatic expansion over much larger sets of data, including domain-specific corpora, corpora of children’s language, historical corpora, etc. We do not claim that it is possible that any pattern dictionary could account for all and only the meanings of words in any natural language. “All and only” represents a theoretical goal that was exploded as unrealistic and distorting for natural-language research (including lexicography) during the second half of the 20th century. Instead, the aim now is to represent prototypical usage and associate it with prototypical meaning.

PDEV is work in progress and is in the public domain. It can be accessed at <http://deb.fi.muni.cz/pdev/>.

Although it is still only work in progress, we urge you to explore it. Comments and feedback are invited.

References

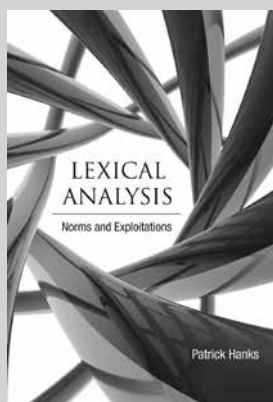
- Church, Kenneth W., and Patrick Hanks. 1989.** ‘Word association norms, mutual information, and lexicography.’ In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, 26–29 June 1989. University of British Columbia. Revised in *Computational Linguistics* 16 (1), 1990.
- Francis, Gill; Hunston, Susan; and Manning, Elizabeth. 1996.** *Collins COBUILD Grammar Patterns 1: Verbs*. London & Glasgow: HarperCollins.
- Francis, Gill; Hunston, Susan; and Manning, Elizabeth. 1997.** *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London & Glasgow: HarperCollins.
- Goldberg, Adele. 1995.** *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goldberg, Adele. 2006.** *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Hanks, Patrick. 2012.** ‘How People Use Words to Make Meanings: Semantic types meet Valencies’. In Alex Boulton and James Thomas (eds.), *Input, Process and Product: Developments in Teaching and Language Corpora*. Brno: Masaryk University Press.
- Hanks, Patrick. 2013.** *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: MIT Press.
- Hunston, Susan, and Francis, Gill. 2000.** *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam & Philadelphia: John Benjamins.
- Moon, Rosamund. 1987.** ‘The analysis of meaning’ in Sinclair, John (ed.).
- Rundell, Michael (ed.). 2002.** *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.
- Sinclair, John (ed.). 1987.** *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD Language Dictionary*. London & Glasgow: Collins ELT.
- Sinclair, John. 1998.** ‘The Lexical Item’ in E. Weigand (ed.) *Contrastive Lexical Semantics*. Amsterdam & Philadelphia: Benjamins.
- Sinclair, John, Patrick Hanks, and others (eds.). 1987.** *Collins COBUILD English Language Dictionary*. Glasgow: Collins.



There were 7 parallel sessions, namely: Theoretical and Methodological Aspects of National Culture Studies in Various Paradigms of Knowledge; National Spiritual Culture: Traditions and Innovations; Cross-Cultural Communication, Cross-Cultural Competence and Globalisation; The Dialogue Between Cultures: West, East and Russia; National Mentality Representation in the Modern Information Globalisation and Preservation of National Cultures in Literary Perception; Lexicography, Terminology Banks and National Identity. The latter session concerned national and cultural aspects of lexicography and problems associated with the formation of corpora and databases for dictionaries. New tendencies in lexicographic practice were discussed and compared to the Russian tradition of dictionary compilation. These included papers on ‘Professional communication in terms of globalization’ (Averbukh), ‘Terminology system of higher education of Russia: National identity or harmonization?’ (Budykina), and ‘Lingua franca, mother tongue, and pedagogical lexicography: Developing a global dictionary series for learners’ and a masterclass on ‘The current status, changes and prospects in the dictionary world’ (Kernerman). The conference proceedings, comprising 800 pages, are the issue of these discussions.

Vera Budykina
vbudykina@gmail.com

Patrick Hanks. *Lexical Analysis: Norms and Exploitations*



Lexical Analysis
Norms and Exploitations
 Patrick Hanks
 Cambridge MA: MIT Press
 2013. 478 pages
 ISBN 9780262018579

This book is a long-awaited milestone of lexicology, lexicography, and empirical theories of language. As the title suggests, its main focus is on the analytic task that is central to lexicography, but the scope of the work inevitably reaches into other aspects of the broader lexicographical undertaking. Ultimately, the book and its theory set a course toward uncharted territory, where lexical analysts will create language resources that may bear little resemblance to dictionaries as we know them. But the book is not only about lexicography: it aims to establish a complete evidence-driven theory of language.

Overview

Lexical Analysis: Norms and Exploitations (henceforth LA:N&E) unifies Patrick Hanks's work on the Theory of Norms and Exploitations (henceforth TNE) in thirteen chapters, of which four have been published elsewhere over the past two decades. Readers who have read the previous publications should avoid merely skimming them in this edition. Much as words have different meanings in different contexts, the new context of this book brings to the earlier papers meaning and nuance that is not as apparent when they are read in isolation.

Chapters 1–3 establish basic concepts and theoretical foundations: why do we need a new approach to meaning? What are 'words', and what exactly might 'meaning' consist of? Most of us will have read these kinds of discussions many times before, but these chapters are where Hanks defines the assumptions and terminology that drive the rest of the book, so they are an important preamble to the work in its particulars.

Chapters 4–8 are the meat of the theory. Chapters 4–6 describe prototypes, patterns, and norms & exploitations, and firmly assert the central importance of context to understanding lexical items. Chapter 7 introduces Hanks's detailed account of alternations, which are variations in pattern structure that do not change the normal meaning of a clause's components. Chapter 8 discusses exploitations in broad scope and detail, finally presenting the first revision of a typology of exploitations. These two chapters broaden the footprint of the theory enough to make it a very comfortable space for further research.

Chapters 9–13 explore applications of TNE to problems that range far afield of lexicography. Chapter 9 reaches into

literature to exemplify how creative writers are creative exploiters of word meaning, and to consider the ramifications of that creativity on the process of lexical analysis and corpus creation. Chapter 10 discusses the range of complexity that is possible in different words' meanings (or meaning potentials): some can be very complex, but not all types of words have the same level of complexity, and so different levels of analysis are appropriate for the different types of words Hanks accounts for.

Chapters 11 and 12 are surveys of preceding approaches to meaning and the lexicon, evaluated in light of TNE. Chapter 11 traces western philosophies of meaning from Aristotle to the present, while 12 considers treatment of the lexicon by previous linguistic theories. Confronting other theories of meaning and lexicon gives Hanks a wealth of opportunities both to credit previous scholarship and to assert how TNE differs from other theories. These chapters are not dutiful surveys of related research: they are dynamic debates with the theorists who have defined our common understanding of words and meaning.

Given the diverse backgrounds of working lexicographers, it is quite possible that many professionals will not have studied the majority of the scholarship that Hanks confronts. The contextualization that Hanks gives these theories is inevitably prejudicial in Hanks's favor, but these chapters serve to put the daily practitioner on equal theoretical footing with the well-read academic.

Chapter 13 reviews the "broader picture", adds references to late-breaking research, considers implications for other fields such as language teaching, and points to many areas for further development of the theory. The book is remarkably strong on concluding material overall. Most of the chapters conclude with sections of "Implications" and a bullet-pointed "Summary", both of which reiterate key arguments and concepts. These sections are a nice feature that more books should have.

The central theme of the book is the usefulness of "distinguishing systematically between normal patterns of collocations and creative uses of those patterns" (p 6). If analysts make such distinctions, and then set aside the exploitations and focus on the norms, Hanks argues that "some well-known linguistic problems are largely solved" (p 17). Hanks hopes that TNE will resolve enough problems that it may be

applied as an evidence-driven theory of language as a whole, rather than simply a theory of lexis and semantics.

The nature of linguistic evidence drives several other core themes. Hanks argues, with passion and wit, against the use of invented examples in any part of empirical language work. He acknowledges the inescapable fuzziness of linguistic events, and appeals to the central role of context in reducing this fuzziness and creating ‘meaning’. He accounts for ‘meaning’ as a confluence of fuzzy ‘meaning potentials’ rather than as a selection from among delineated, mutually-exclusive options. The tension between empiricism and intuition is a standard theme of works in corpus-driven lexical analysis, and Hanks shows that statistical analysis is not enough to overcome the biases and prejudgments of intuition. To accompany statistical measures, he offers many rules and heuristics for deciding whether a given linguistic event belongs in one fuzzy category or another.

Hurdles for the reader

The notion of “setting aside the exploitations” appears to be a common stumbling block for lexicographers upon their first exposure to TNE (at least, it was for this reviewer upon my first reading of Hanks’s 1994 paper). When Hanks tells you that a given word has five “normal” meanings, your creative brain immediately remembers sixth and seventh meanings that you saw written by ‘reputable’ authors just the other day. A lexicographer who has been trained to account for all uses of a given definiendum may feel justifiably uncomfortable if told to pass over a frequent collocation that has a discrete meaning.

In answer to such concerns, Hanks makes compelling arguments that the distinction between normal and creative usage is essential, and can indeed be made empirically, even if allowing for some fuzziness at the edges. Hanks makes a distinction between cognitive salience (“memorable because it is unusual”) and social salience (“conventional and for that reason unmemorable”) (p 5). This distinction, together with chapter 9’s survey of literary exploitations, build a strong case that the tremendous variety of possible exploitations can be managed only by means of a solid understanding of what happens in normal usage.

Hanks lays out clear and principled criteria to tell the difference between norms, exploitations, alternations, and simple errors. These criteria provide coherent theoretical justification to some analytic conclusions which, heretofore, even careful

analysts of corpus evidence may reach only ad hoc or through intuition.

Of course, even the clearest criteria may not always prevail over unclear data, but a satisfying majority of cases appear to be resolvable with the techniques that Hanks has refined and developed in the course of his career. Furthermore, it is not the case that fuzziness creates usage that is utterly incomprehensible to humans or uncategorizable by machines. Rather, part of the strength of the theory is that it leverages normal uses to make sense—as much sense as can be made—of all the varieties of non-normal usage, in ways that show good signs of being computationally tractable.

Readers not familiar with frame semantics (Fillmore 1982), generative lexicon theory (Pustejovsky 1995), or Hanks’s earlier work around TNE will, in a few places, need to be indulgent of certain formal technical description that is foreshadowed or introduced without being fully explained: I refer here chiefly to the double-square-bracket markup of the corpus patterns, which begins to appear in sprinkles in chapter 2, but is most fully explained only just before it flows in a torrent in chapter 4. A page or so of chapter 2 seemed to me to be more appropriate for a later chapter, but a patient reader will benefit from the warm-up if they take the time to grok it in chapter 2.

Hurdles for the theory

Hanks’s productivity and influence means that the TNE has been well-exercised, and in many places it reflects consensus among the schools of corpus analysis and lexicology of which Hanks is a major figure. Unfortunately, in lexicography it is an acknowledged truth that much of the theoretical consensus is not implemented within the most widely consumed lexical references. This may delay the impact of the theory on lexicography, and in the absence of lexical resources built on these principles, computational linguistics will not be able to reap the theory’s full benefit without first helping to create those resources.

Hanks has been at work to remedy the gap between theory and practice, with the Corpus Pattern Analysis project and subsequent *Pattern Dictionary of English Verbs*. These projects are an exciting, practical testing ground for the theory, and Hanks’s eminence in practical lexicography emphasizes the applicability of the theoretical methods he proposes here. Considering the troubled state of the dictionary business as we know it, it is possible that a complete lexicon on TNE principles may not be made by a “dictionary” group. Still, the book’s tips,

KERNERMAN JAPANESE



Series of bilingual learners’ dictionaries for speakers of Japanese on SD cards for MICRO hand-held e-dictionaries by Seiko Instruments, Tokyo 2013.



Chinese 中中日辞典

Chinese: Saihong Li (chief editor), Huimin Li, Xilan Sun, Yuan Yuan Wang, Chao Zhang, Hao Chen, Chia Chen Liu

Japanese translation: Koji Kawa



French 仏和辞典

French: Nathalie Lanckriet (chief editor) | Valérie Lecœur, Marine Lesprit, Frédérique Melanie-Becquet, Emilienne Morruzzi, Caroline Moulart-Fremaux, Edouard Trouilleux, Colette Alouch

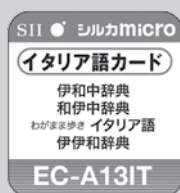
Japanese translation: Ken’ichi Hisamatsu, Fumitoshi Hayakawa, Yuji Obataya



German 独独辞典

German: Susanne Billes (chief editor), Monika Böhme, Sabine Corso, Susanne Dyka, Marion Hahn, Christine Moehrs, Emmanuelle Worbs

Japanese translation: Yoshinori Kaisaki, Naoko Toshimitsu, Tamaki Sakurada-Kaufmann



Italian 伊伊和辞典

Italian: Palma Gallana (chief editor), Davide Levi
Japanese translation: Yu Akao, Castrese Cacciapuorti, Yoko Shibue



Korean 韓韓日辞典

Korean: Youn-Sill Couralet (chief editor), Stéphane Couralet, Migyung Seok, Kyung-Ryang Kim, Kwi-Jeong Yoon
Japanese translation: Shigeru Hashimoto



Russian 露露和辞典

Russian: Margarita Sapozhnikov (chief editor), Lana Philippov, Polina Brukman, Hanna Ioffe
Japanese translation: Nobuyoshi Adegawa, Elena Baibikov, Yuka Ikeda, Tomoko Masuda, Yuko Wakabayashi



Spanish 西西和辞典

Spanish: Miguel Eduardo Montoro (co-chief editor), Irene Renau Araque (co-chief editor), Sergio Mario López, María Verónica Elizondo, Carlos Sebastián Ceró
Japanese translation: Kobayashi Kazuhiro, Shingo Kato, Keisuke Gozawa

http://www.sii.co.jp/cp/products/card/index_micro.html

insights, and simple rules of thumb are as actionable for anyone working with corpora as they are for lexicographers.

Among the obiter dicta that pepper the book, there is mention that “expectation of homogeneity” is a trap for the “unwary [lexical] analyst”—a warning that “there is no reason to believe that the apparatus required for the semantic analysis of verbs will be necessary or sufficient for the semantic analysis of nouns (or vice versa)” (p 16). Hanks describes an apparatus for verbs and also for nouns, but spends much more time on the verbs. Indeed, Hanks’s current project is a pattern dictionary of 3,000 English verbs. It would be interesting to know how the book and the theory might have developed differently if Hanks’s ongoing projects focused on nouns instead of verbs.

It also seems relevant for a complete theory of language to consider the interactions between patterns of nouns and of verbs. LA:N&E is also virtually silent on the other parts of speech, aside from mentions that the TNE is driven by content words. Hanks acknowledges that the theory is incomplete, and in many places points to further work that is necessary. To paraphrase Melville, small theories may be finished by their first architects; great ones ever leave the coperstone to posterity. Recently, Ken Litkowski has undertaken a Corpus Pattern Analysis project for prepositions (Litkowski 2012) that may precipitate some answers to how patterns interact.

Early in the book, Hanks notes that ontologies have yielded disappointing results for word-sense disambiguation, and attributes this to severe underestimation of the fuzziness of boundaries between categories. TNE offers a very different model of meaning than what is traditionally considered in word sense disambiguation (WSD), and does not waste energy trying to avoid fuzziness, instead quarantining boundary cases from normal behavior.

Far larger-scale implementations are necessary to learn whether the theory can reduce the underestimation enough to serve a purpose like sense disambiguation (using whatever empirically-derived substitute for WSD is most applicable, since nobody who seriously works with meaning believes in ‘word sense’ as such (Kilgarriff 1997)).

In addition, LA:N&E does not address the question of how reproducible Hanks’s own analyses might be. In reading the book I occasionally found myself disagreeing not with the broader theory, but with individual judgments that Hanks had made in describing the theory’s practical output. No matter how evidence-driven the theory is, human analysts will not always interpret the

same evidence in the same way. Fuzziness is one thing, but undoubtedly some usages will defy agreement. Will TNE be able to yield reliable inter-annotator agreement on the whole?

The future that the TNE points to

LA:N&E is only the bottom of the bottom-up theory of language that Hanks envisions. The book points to many areas where the theory needs further development, and leaves many questions unanswered. How well does TNE apply to languages other than English? Do other languages have types of exploitations that are not possible in English? How are word senses activated by context? Without clear criteria, “sense activation” will remain merely a hiding-place for intuition.

A very real question is whether this kind of analysis still needs to be done by humans. Hanks does not extensively address the prospect of doing lexical analysis computationally—the theory is expressed for the human analyst. Some have seen this as under-ambitious, supposing that if we can do it with our brains, we should be training machines to do it. I believe that we can train machines to “do” TNE, but we cannot train them without first understanding what exactly the task is. For both human and computational implementations of this theory, Silvie Cinková (and her group at Charles University in Prague) have begun investigating the capacity for humans to be trained in Corpus Pattern Analysis (Cinková 2012a) as well as ways to manage the fuzziness of semantic categories (Cinková 2012b). Hanks also points to Popescu (2012), as a path to automating the processing of corpus patterns in the not-too-distant future. These efforts may ultimately be what closes the gap between theory and practice, both for human and computational lexicography.

The value of the TNE

This is not a book of abstract theory that cannot be practically implemented. The methods described here can be used by anyone who works with corpus evidence, on a large or small scale. A work with the aims of the *Pattern Dictionary of English Verbs* is a major undertaking, but even for the daily practice of lexicography on existing projects, the book offers many simple heuristics for the classification and definition of corpus evidence, which can be used as soon as the principles are understood. Indeed, once the principles are understood, any other way of handling corpus evidence may seem unacceptable.

The book and the theory aim to do much more than improve lexicography. Working

through all the details of the empirical basis of TNE yields a theory and a practical framework that shows strong potential to help with many of the other difficult problems in linguistics. The book is overflowing with suggestions for future research, any of which will solidify the undeveloped parts of TNE while also advancing the sciences of language. LA:N&E is an important work, a milestone of several fields, but there is still much work to do.

References

- Cinková, S., Holub, M., Kríž, V. 2012a.** Optimizing semantic granularity for NLP - report on a lexicographic experiment. In *Proceedings of the 15th EURALEX International Congress 2012*, 523-531, European Association for Lexicography, Oslo.
- Cinková, S., Holub, M., Rambousek, A., Smejkalová, L. 2012b.** A database of semantic clusters of verb usages. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 3176-3183,

European Language Resources Association, İstanbul, Turkey.

- Fillmore, C. J. 1982.** Frame Semantics. In *The Linguistic Society of Korea (ed.), Linguistics in the Morning Calm*: 111-137. Seoul: Hanshin Publishing Co.
- Hanks, P. 1994.** Linguistic Norms and Pragmatic Exploitations, or Why Lexicographers need Prototype Theory, and Vice Versa. *Papers in Computational Lexicography: COMPLEX '94*, 89-113.
- Kilgarriff, A. 1999.** I don't believe in word senses. *Computers and the Humanities*, Vol. 31, No. 2, 91-113.
- Litkowski, K. 2012.** Corpus Pattern Analysis of Prepositions, Technical Report 12-02. Damascus, MD: CL Research.
- Popescu, O. 2012.** Building a resource of patterns using semantic types. In *Proceedings of LREC 2012, Istanbul*.
- Pustejovsky, J. 1995.** *The Generative Lexicon*. Cambridge, MA: MIT Press.

Orion Montoya
orion@mdcclv.com



KERNERMAN Semi-Trilingual English-Korean-L3

Afrikaans, Arabic, Bulgarian, Chinese Simplified, Chinese Traditional, Danish, Dutch, French, German, Hindi, Hungarian, Indonesian, Italian, Japanese, Malay, Norwegian, Polish, Portuguese Brazil, Russian, Serbian, Spanish, Swedish, Thai, Turkish, Ukrainian, Urdu, Vietnamese

K-Apps for Android
Daolsoft, Seoul. January 2013

Launching of *LEXICOGRAPHY: Journal of ASIALEX*

The long-awaited, flagship publication of the Asian Association for Lexicography will be launched in 2014. *LEXICOGRAPHY: Journal of ASIALEX* will be published by Springer, the prestigious worldwide scientific and professional publisher. It will have two issues a year, appearing in April and October, in print and online. The chief editor for the initial three-year term is Professor Yukio Tono, of the Graduate School of Tokyo University of Foreign Studies and current President of ASIALEX, and the editorial board consists of Gilles-Maurice de Schryver (Ghent University), Chu-Ren Huang (Hong Kong Polytechnic University), Shin Ishikawa (Kobe University), Ilan Kernerman (K Dictionaries), Deny A. Kwary (Airlangga University), Lan Li (Hong Kong Polytechnic University) and Shigeru Yamada (Waseda University), with the assistance of an advisory board including international scholars.

ASIALEX was established in 1997, with the principal mission of fostering scholarly and professional activities in the fields of lexicography and dictionaries in and for Asia. There are time-honored lexicographic traditions in Asia and great many topics to discuss and problems to solve. However, the new journal aims to serve as a leading-edge crossroad and powerhouse for all global issues of lexicographic interest, with an emphasis on Asian perspectives and concerns. Just like the ASIALEX conferences, *LEXICOGRAPHY* is open for the world's researchers, lexicographers, dictionary makers,

translators, teachers, students, and all language lovers to discuss issues concerning lexicography and dictionary history, typology, use, criticism, structure, components, compilation, application, media, and interaction with other fields such as phraseology, translation, corpus linguistics, language learning, ICT, etc. Subscription will entail membership to ASIALEX and the right of free publication to authors, subject to peer review of papers.

The text above is based on the ASIALEX announcement that was drafted by Shigeru Yamada (with additions by Chu-Ren Huang and Deny A. Kwary) and revised by me. I had the privilege to negotiate on behalf of ASIALEX the framework of this publication and cooperation with Springer for the last year, and would like to thank my colleagues on the board for their support and Springer's representative, Federica Corradi Dell'Acqua, for her dedication. During the recent DSNA meeting at the University of Georgia in May 2013, I had the opportunity to discuss our forthcoming publication with the editors of *Dictionaries* and *International Journal of Lexicography*, respectively Edward Finegan and Anne Dykstra, who heartily welcomed the news. We look forward to establishing *LEXICOGRAPHY* as a prominent international journal and proceeding to collaborate with all our sister publications around the world.

Ilan Kernerman



Paul Bogaards, 1940-2012

Paul Bogaards died at his home in Den Bosch, the Netherlands, on October 3, 2012.

He studied French, general linguistics and Italian at Leiden University, where he was appointed as a teacher and researcher in the Department of French in 1976, and in 1982 he defended his PhD. dissertation on foreign language teaching at the Vrije Universiteit Amsterdam. As a teacher of French he was very interested in dictionaries as a tool in foreign language teaching, which was the subject of many of his articles. He also was general editor of a series of French / Dutch bilingual dictionaries of the Van Dale and Robert publishing houses.

Many people will know Paul as the editor of the *International Journal of Lexicography* (IJL). He held that position from 2003 to 2012 and during that period he was an ex-officio member of the EURALEX Executive Board. During his editorship, IJL made it to the top of linguistic journals.

One of the good things of EURALEX congresses and board meetings was that I would meet Paul Bogaards. It is so sad that this will never happen again. Paul was on everybody's mind at the interim meeting in Bolzano on June 1, where we observed a moment of silence in his memorium.

Paul Bogaards was a gentle and amiable colleague and friend. His simple presence was enough to make people relax. He even managed to make me feel good and at home when I visited him three days before he died. He was so sick and tired, but wanted to make sure that IJL would remain in good hands. We worked together for a couple of hours. I do not know how he managed this, but he did. That was Paul all over.

Paul will be sorely missed by IJL, EURALEX and the international lexicographical community.

Anne Dykstra



Wolfgang Worsch, 1953-2013

On March 17, 2013, Wolfgang Worsch died unexpectedly.

Born in Erlangen, Wolfgang studied German and Literature, and began teaching at the University of Birmingham in 1980. There he met John Sinclair, got bitten by the dictionary bug, and became involved in the Collins Birmingham University International Language Database (COBUILD). While in England, he started to work for Oxford University Press as well as for the American publisher Collier Macmillan.

Returning to Germany in 1986, he became a lexicographer at Langenscheidt in Munich, where his most important project was the new edition of the *Muret-Sanders*, the largest existing bilingual English-German dictionary and the Langenscheidt flagship. In 2007 he took on a new challenge and moved to Berlin to help build the dictionary department at Cornelsen, managing to produce a series of seven school dictionaries in only two years. He also worked for Duden and for Akademie der Wissenschaften (Academy of Sciences) and took part in a number of international projects like the *New English-Irish Dictionary*.

WoWo, as he was known by friends, was an enthusiastic basketball player and coach, as well as a passionate chef. After his son Fabian was born disabled in 1990, he joined the board of directors of *Helfende Hände* (Helping Hands). Everybody was impressed by the way he balanced his family, career, and volunteer work, and he will be thoroughly missed by his family, colleagues and friends.

He is survived by his wife, Christine, his son, Fabian, and his daughter, Laura.

Helga Holtkamp and Benedikt Model