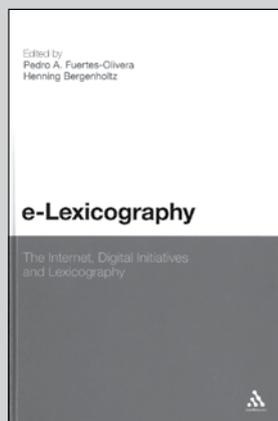
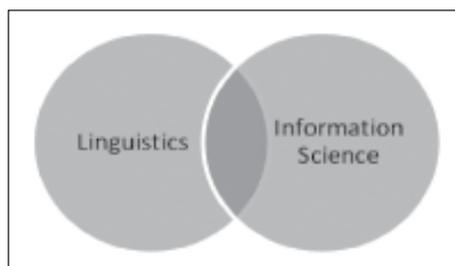


Pedro A. Fuertes-Olivera and Henning Bergenholtz (eds.). *e-Lexicography: The Internet, Digital Initiatives and Lexicography*



e-Lexicography.
The Internet, Digital Initiatives and Lexicography
Pedro A. Fuertes-Olivera and Henning Bergenholtz (eds.)
London/New York: Continuum.
2011. xiv, 341 pp.
ISBN 978 1 4411 2806 5

A lexicographer is a divided soul, part scientist, part tool-builder. The scientist is a linguist, wanting to describe the language. The tool-builder wants to help the user find the information they want, the territory of information science. Lexicography is in the intersection.



One might divide the lexicographic process into two parts: analysis, in which we aim to determine the facts, and synthesis, in which entries are prepared. Analysis is linguistics; synthesis is information science.

In this book information science reigns.

Description

The book is the outcome of a symposium in Valladolid, Spain, in 2010, and comprises fifteen chapters by different authors, most but not all associated with the Aarhus School and its function theory of lexicography. Many of the chapters had their seeds in presentations at the symposium.

The book opens with an introduction by the editors Fuerte-Olivera and Bergenholtz. It starts with a conversation with Wiegand, about whether there should be different theories for print and electronic dictionaries, answering that we need one theory that covers all. It then borrows from Gouws's article four agenda items for the book:

- Using databanks from which different types of entry can be extracted
- The mistake of including too much information
- The broadening of lexicographic theory beyond dictionaries
- E-dictionary users are familiar with the internet and the potential it offers: what implications does that have?

They then summarise the remaining articles.

Gouws calls his chapter 'Learning, Unlearning and Innovation' and addresses a colleague's question, "does all the research in theoretical lexicography lead to an improvement in the quality of dictionaries?" He answers yes, for several reasons: because it will now be based on a sound theory;

because lexicography will no longer be a sub-discipline of linguistics; and because, with bold planning, they will embrace the potential of user-generated content.

Henning Bergenholtz, in his chapter, makes the case for research into user needs and describes some results from logfile analysis. His presentation of the time it took users to find information, and its relation to whether they had found the information they wanted, is thought-provoking: in particular the paper dictionaries gave faster access than the electronic ones! While electronic dictionaries potentially allow fast searching, whether they actually do depends on their design, and electronic media introduce many new ways to get distracted, confused and lost. He then describes a set of four monofunctional dictionaries derived from a single database: the pluri-monofunctional model. This was clearly successful, pointing the way ahead for user experiences tailored to information needs. He looks forward to the time when these dictionaries have been more extensively used, so the logfiles will be a large enough body of data to support extensive user research.

In this chapter I did find the review of the literature partial: he says

"... lexicographic interest in the needs of the users ... has been astonishingly scant" (p 31)

not acknowledging the substantial volume of work on the theme, and only picking out one article, by Bogaards, from 1990, to criticise it. One might have expected him to view it as an early, if modest, attempt to start an enterprise that he and others were continuing.

Tarp, in his chapter, provides an appealing vocabulary, already widely adopted, for talking about e-dictionaries: *copycats* (paper dictionaries copied onto digital media), *faster horses* (as copycats, but faster searching), *model T Fords* (first attempts at using what digital media offer) and *Rolls Royces*. The subtitle is 'Towards the Individualisation of Needs Satisfaction' and this is the key to moving Rolls-Royce-wards. He makes a useful distinction between interactive, active and passive methods for individualization, according to whether the user (or the system) takes the driving seat.

This theme is taken further in the following chapter by Bothma. His discipline is information science. He surveys the methods and techniques that modern information technology offers to

lexicography, for filtering and adapting data as held in a database according to user needs. The chapter is full of examples of how particular online dictionaries use particular methods, so provides lots of examples of good (and not-so-good) practice. Online dictionaries are shown as belonging in the same sphere as Google, Booking.com and Amazon, with methods pioneered in those places available for dictionary database and interface designers.

A finding referred to twice in the book is from Leroyer's chapter, that only one quarter of lexicographic works published in 2008-2009 are general language dictionaries, whereas three quarters are

“made up of wordlists and language data organized in dictionary articles, but which nonetheless have nothing to do with language as a scientific object of study” (Leroyer p124)

They are special language dictionaries of one kind and another. This gave me pause for thought. Most lexicography, it suggests, is not linguistic at heart.

But then:

- When ordinary people refer to dictionaries, they mean general language dictionaries like the *Oxford English Dictionary*, *Le Grand/Petit Robert*, *Duden*, *Webster*, etc.
- Purchases and sales of general language dictionaries dwarf those of special language dictionaries.
- Almost all substantial dictionaries (more than 4cm thick, if we take print as a reference point) are general language dictionaries.
- Almost all large lexicographic projects (comprising, say, more than ten people over more than three years) are for general language dictionaries.

The comparison is like noting that there are more local airstrips than international airports in the world, so basing an account of aviation on local airstrips. Numbers of publications alone do not give a good overall picture, and I remain convinced that general language dictionaries are central to the lexicographical firmament.

For lack of space I'll take the chapters by Spohr, Nielsen and Almind, Fuertes-Olivera and Niño-Amo, Bergenholtz and Bergenholtz, and Anderson and Almind together. They present technical challenges, and present examples, of the approach to dictionary-making where a single database meets a range of user needs by selecting only the appropriate information to show in a particular case. The dictionaries referred to are monolingual and multilingual, general language (English phrasal verbs) and special language: music, in the *Danish Music Dictionary* (Bergenholtz and Bergenholtz),

and the accountancy dictionaries, for English, Danish and Spanish (with English as the hub), which Fuertes-Olivera and Niño-Amo's and Nielsen and Almind's chapters describe. The project looks good, though it was disappointing to find no references to the extensive discussion of issues relating to hub-and-spokes models and to translation mismatches, for example in Janssen (2004). Also their example displays the hazards of lexicography moving into encyclopedic territory, with Fig 7.4 showing a definition for the accountancy term *deemed cost*, which begins

“an amount used instead of cost or depreciated costs at a specific date. Any following amortisation or depreciation is made on the assumption that ...”

I googled the term (in inverted commas, to get an exact match) and the second hit stated

‘**Deemed cost**’ is a surrogate for cost at a given date. For example if a building is purchased at \$100000 this is cost and also the **deemed cost** at that given date. ...

Accurate, extensive encyclopedic entries are very often already available, and very easily accessible via google, as here. A case has to be made for what value lexicographers are adding.

Lew, in his chapter, is good to his title, ‘Online Dictionaries of English.’ He introduces some useful criteria for thinking about online dictionaries – for example individual (standalone) dictionaries, vs clusters of dictionaries (eg, from the same publisher and on the same website) vs portals (websites that give links to lots of dictionaries) vs aggregators (which offer entries for a word from unrelated dictionaries, e.g., Dictionary.com). He then reviews a large number of online resources, mainly classifying and describing what is out there, sometimes evaluating.

In contrast Sanchez and Pascal review the case of online monolingual Spanish dictionaries and find just four, all closer to copycats than Rolls Royces. They then develop an account of what could be done, making use of the potential of the electronic medium to give a very rich account of a word within their Lexical Constellation Model.

Verlinde's chapter describes his *Base Lexicale de Français*: here, as also discussed at the e-lexicography conference in Bled, Slovenia, in November 2011, is already something we might call a Rolls Royce. As the chapter describes, and the Bled presentation demo'd, this is a dictionary – extensive, and widely-used – which applies many lessons of what can be done, online, with current technology, to customise according to the user's characteristics and information need.

eLex 2013 Electronic lexicography in the 21st century

The third international conference on electronic lexicography, eLex 2013, will take place in Tallinn, Estonia, on 17-19 October 2013. It is organized by the Institute of the Estonian Language, Estonia, and Trojina, Institute for Applied Slovene Studies, Slovenia.

The previous two eLex conferences (at Louvain-la-Neuve, 2009, and Bled, 2011) have shown that there is huge interest for such events from both academia and industry. This conference will be yet another great opportunity for presenting state-of-the-art projects and research in the fields concerned with e-lexicography, as well as discussing topical issues and networking.

The conference will be preceded by SKEW4, the 4th annual Sketch Engine workshop, which will take place on 16 October 2013 at the same venue, and is organized by Lexical Computing Ltd.

More information is available on the conference website, <http://eki.ee/elex2013>.

Iztok Kosem, Jelena Kallas
Co-chairs of the eLex 2013
organizing committee

Heid takes a well-established method from information science – the usability study – and applies it to dictionaries. He provides an overview of usability studies and shows how they can be applied in lexicography – and indeed how they can often provide an answer to the questions that motivate much of the book: what queries, forms, data and structure – give the user (with a particular information need) most help. The study in this chapter compares three ‘live’ online dictionaries, and suffers from the methodological problem that there are many differences between the three, so it is not clear which differences resulted in a more or less successful user experience. Heid notes the problem and discusses, as ‘further work,’ a model where the ‘dictionaries’ to be compared are closely controlled so there is just one parameter according to which they differ. He and his co-workers have since lived up to this promise, with a study presented at the Bled conference, in which they do just that.

The book concludes with ‘Ten Key Issues,’ a chapter which summarises the discussions from the Valladolid workshop, edited by Samaniego Fernández and Cabello de Alba.

The Aarhus School and linguistics

The Aarhus School denigrates linguistics with vigour:

“Linguists were the princes of meta-lexicographic discussions, and meta-lexicography and practical lexicography were subsections of the work done by these linguists. This era in the history of lexicography can rightfully be regarded as representing a form of *linguistic colonialism*.” (Gouws p22)

“... linguists who also masquerade as lexicographers...” (Bergenholtz and Bergenholtz p188)

Let me state my vantage point. My friends and colleagues Sue Atkins and Michael Rundell chose to include “practical” in the title of their *Oxford Guide to Practical Lexicography* and state:

“This is not a book about ‘theoretical lexicography’ – for the good reason that we do not believe that such a thing exists.” (p4)

Authors of most chapters in this book, however, state the need for theoretical lexicography. Unsurprising then, that my friends are its enemies, already identified as such (along with Henri Béjoint) in the introduction (p8), and explicitly denounced in Bergenholtz and Bergenholtz’s chapter (p189). I won’t pursue the question “is there such a thing as theoretical lexicography” as I fear it would be an arid discussion on

the meaning of *theory*.¹ I hope it is not contentious to say, theory or no theory, the Aarhus School is concerned to place the user’s information needs centre stage, and the ways and means and implications of doing that are the central theme of the book. This is somewhat in spite of its title, ‘e-Lexicography’. In the 21st century, pretty much all lexicography is e-lexicography, in the senses that the writing is based on digital evidence, takes place on a computer and employs dictionary writing systems, and most users will be accessing the data through a computer or other electronic device, and there is little more needing saying. The subtitle, ‘The Internet, Digital Initiatives and Lexicography’ does not add much. A more informative title would have been ‘Putting user needs at the centre of lexicography.’ The relation between user needs, and having the dictionary data in electronic form, is that we can show different users different things, according to their information needs. This is the link between the ‘e’ and the real topic of the book.

Another perspective on the role of linguistics in lexicography is this. The chapters of this book are mostly concerned with delivering information to the user. This is of course fully legitimate, and the questions “how much information” and “which information, when” are good ones – but none of the chapters discuss the risk of delivering false or misleading information. They proceed as if the truth were known and the database contained all and only correct material. Would that it were so! A careful review of any dictionary – see for example Hanks on *Merriam-Webster’s Advanced Learners English Dictionary* (*International Journal of Lexicography* 22.3, 2009) – will uncover points at which it is likely to mislead and confuse. Even in this book, where presumably the authors have chosen examples with care, I noticed a lexicographical bloomer. On pp 211-213 we have an analysis of the English phrasal verb *call back*. It is given six meanings of which the sixth is given the example “I cannot call his face back.” As an English native speaker, I go *eeeeeeugh*. This is blazingly wrong. (We might say “I cannot recall his face.”) A little research revealed that this ‘example sentence’ exists in a number of dictionaries and translation tools: a dictionary error that has been copied and recopied from dictionary to dictionary.

A simple and central case, in both general-language and special-language, concerns variability of set phrases and idioms. Somehow, if the user is trying to decode “quaking in one’s boots” – or even “quivering on one’s Doc Marten’s” we would like to direct them to the idiom that

(debatably) has as its core form “shaking in one’s shoes” (example from Moon 1994). This is a hard problem: it is both hard to work out how to represent the facts in a usable way, and then it is hard to work out, for each individual expression, what the facts are: different phrases allow variability to different degrees, in different ways. In Bergenholtz’s article on his dictionary of fixed expressions, he considers the Danish *på vulkaner*, *være på vulkaner*, *danse med vulkaner* and several other variants but does not discuss the challenge of how the lexicographer might discover the range of variation of the phrase, or of how this might be represented in the database or shown to the user.

To come back to the Venn diagram with which I opened the review, the linguistics part is for analysis. If we had a database containing all the facts and generalisations about the behaviour of all the words and phrases of the language, optimally structured, then we wouldn’t need linguistics. But we don’t. That is what linguistics aims to do, and what the lexicographer, when working on a particular word, aims to do for that word.

One surprising and disappointing aspect of the book is the poor standard of production. Given the topic, one would have expected a book where figures are readable, and where thought had been given to the best typography for presenting complex dictionary entries. But it took a magnifying glass to read the text on screenshots in Chapter 10, and throughout the book complex lexical entries are presented in plain text, with no use of font, font size or weight, indentation or other formatting to make them digestible. Running headers relate to chapter names rather than author names.

Conclusion

This is largely an Aarhus School book, with a number of interesting and useful chapters exploring and developing the model of the pluri-monofunctional dictionary.

While I find the Aarhus School’s attention to the information-science side of lexicography often useful and enlightening, I find its attacks on the linguistic side puzzling. The Aarhus School doubts the relevance of corpora for lexicography (explicitly, in the concluding chapter, p309). But you need corpora to get the facts right.

Notes

¹ We Anglo-Saxons are often dubious about grand statements of theory. When Wittgenstein pronounced “Whereof one cannot speak, thereof one must be silent” his English friend and colleague Frank Ramsey (no intellectual slouch, a founding figure in mathematical economics and decision theory) responded “What we can’t say we can’t say, and we can’t whistle it either.”

References

- Atkins, S. and Rundell, M. 2008.** *Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Janssen, M. 2004.** Multilingual Lexical Databases, Lexical Gaps, and SIMuLLDA. *International Journal of Lexicography*, 17.2: 137-154
- Moon, R. 1994.** *Fixed Expressions and Text*. PhD thesis, School of English, University of Birmingham.

Adam Kilgarriff

Lexical Computing Ltd
adam@lexmasterclass.com

Olga Karpova. *English Author Dictionaries (the XVIth – the XXIst cc.)*

Olga Karpova’s *English Author Dictionaries* begins with her wonderment about why the English author dictionary has been neglected in dictionary research in spite of the fact that it “has at its disposal about 300 titles of linguistic and encyclopedic reference works to single and complete works of more than eighty writers” (p. ix).

I am much in agreement with this sentiment. We are well aware of the fact that, over the centuries, the “author’s dictionary,” or the reference work “which provides information on the vocabulary of a specific author” (Hartmann and James

2001, 10), and the “dictionary of authors,” or the reference work “providing literary, biographical and sometimes encyclopedic information about a selection of authors” (Hartmann and James 2001, 43) concerning English writers have been published in great number. We also know that a copious number of reviews have been written for such works.

However, when it comes to the point of how often they have been researched from a holistic perspective, it is quite another story. For instance, even in the voluminous *The Oxford History of English Lexicography*