'It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical[1]

Michael Rundell

Keywords: *lexicographical theory*, *function theory*, *metalexicography*, *prototype theory*, *regular polysemy*, *lexical functions*, *user-generated content*, *collaborative lexicography*, *adaptive hypermedia*.

Abstract

This paper considers how the practical business of producing dictionaries may be informed by and facilitated by theoretical considerations. What kinds of theory have the potential to make dictionaries better? And is there such a thing as 'theoretical lexicography'? Several theoretical paradigms are discussed. In the case of the metalexicographic contributions of L.V. Shcherba and H.E. Wiegand, it is suggested that their relevance to the practical task of dictionary-creation is limited; and it is argued that the so-called 'theory of lexicographical functions' proposed by Henning Bergenholtz and his colleagues, while helpfully focussing on users and uses, adds little that is new to the debate. Conversely, it is shown that linguistic theory has much to offer lexicographers, and the direct applicability of various linguistic theories is demonstrated in a number of case studies. Finally, the whole discussion regarding appropriate theoretical inputs for lexicography is brought into the radically changed digital world in which lexicography now finds itself.

1. Introduction: Zgusta's dilemma.

Back in 1985, Anna Wierzbicka observed that 'Lexicography has no theoretical foundation, and even the best lexicographers, when pressed, can never explain what they are doing, and why' (Wierzbicka 1985: 5). Is this still true? And if true, does it matter? It depends who you ask; this is a divisive issue. To simplify a little: we have, ranged on one side, a diverse group united by their fervent belief in the idea of 'lexicographic theories' which dictionary-makers ought to follow. On the other side are those who, like Béjoint, 'simply do not believe that there exists a theory of lexicography' (Béjoint 2010: 381). Both positions have a long history: lexicographic theories dating back to the 1940s are discussed in the next section; conversely, Sinclair, speculating 30 years ago on the idea of

lexicography as an academic subject, concluded that 'there is… no heading "Lexicography theory" in my syllabus because I have nothing to put there' (Sinclair 1984: 6), later adding, even more dismissively 'there is no prospect of a theory of lexicography' (ibid. 7). It would not be giving much away to identify myself (and my friend and co-author Sue Atkins) as belonging to the second camp.

There is a line of argument based on what one might call 'the proof of the pudding'. The justification goes that, in the last three decades English-speaking (specifically British) lexicography – despite being seen as light on theory – has been among the most innovative areas of activity in the field, producing dictionaries which have broken new ground, been critically and commercially successful, and been widely imitated. So who needs theory? It is a tempting argument, but not one I propose to make here. The fact that 'the British school' has tended to produce good dictionaries proves nothing: perhaps it would perform even better if supported by an appropriate theoretical paradigm. But in fact, it will be argued, lexicographers (and this certainly includes British lexicographers) are not at all atheoretical, still less anti-theoretical, in the way they approach their work, and indeed most would see such attitudes as perverse.

In a moving account of a long and complex dictionary project, Ladislav Zgusta described a predicament with which many of us will empathize: 'Every lexicographer knows that rosary of agonizing choices that must be taken every time when a decision has to be made and there is no time to do what would be the only reasonable thing to do, namely to research the problem for a year or two' (Zgusta 1992: 91). Making decisions is a big part of the lexicographer's job, and making good decisions depends, as Zgusta suggests, on 'trying to find the underlying regularity, or rule, in a (sometimes only seeming) chaos or randomness' (ibid. 92). Clearly, then, lexicography needs theory – but what kind of theory? Zgusta's quest for 'the underlying regularity' and the need to perceive order and system in the apparent randomness of language is a good indicator of the kind of guidance lexicographers seek. To do the job well, we need to feel confident that our reliance on intuition and subjective judgments is kept to a minimum; that our approach is systematic, internally-consistent, and driven by what the language data is telling us; and that the description we end up with is compatible with our observation of usage, and is complete, with nothing of importance omitted (either by accident or design). Given these needs, it would be foolish not to embrace any theories that may help us do our jobs without

the kind of agonizing Zgusta refers to. In this paper, I hope to demonstrate that lexicographers are indeed open to theoretical concepts and that good dictionaries draw on a wide range of theoretical inputs. And this readiness to engage with theory will be all the more important as we negotiate the next big lexicographic revolution.

2. Metalexicography and its relevance to dictionary-making

In a recent discussion of this topic, Paul Bogaards posed the question: 'If there is such a thing as a theory of lexicography, one is entitled to ask such questions as: Who did conceive or launch it? What is its content? Are there any recent developments? Are there any competing theories or is there just one?' (Bogaards 2010: 313–314). The next two sections will attempt to answer these questions.

Metalexicography is a broad discipline, taking in subjects such as dictionary criticism, dictionary typology, the history of lexicography, and the description of dictionary microstructures. These are all valid areas of study. The question for practising lexicographers is how far any of this impacts on how they do their jobs or helps them produce better dictionaries. There is a vast literature and no space here for a comprehensive survey. Instead, I will look briefly at the work of two of the best-known names in this field: L.V. Shcherba and H.E. Wiegand. But first, a brief aside on a related subject.

2.1. *A short diversion: dictionary user research*

In the case of dictionary user research – often seen as a branch of metalexicography – the practical applications are obvious. Not surprisingly, Euralex has supported user-research programmes since its inception. In his Preface to the Proceedings of the first Euralex congress, Hartmann notes 'the emphasis in all sections [of the conference] on the needs of the dictionary user' (Hartmann 1984); a Euralex-sponsored seminar organized by Tony Cowie in Leeds in 1985 focussed mainly on the needs and abilities of language learners as dictionary users (Cowie 1987); and every Euralex congress since has included papers reporting on a range of user-research projects. Atkins and Rundell (2008: 30–45) briefly summarize aspects of dictionary content and presentation which may benefit from (and have benefited from) the insights provided by user

research; while Tarp (2009: 283–289) and Lew (2011) discuss the various research methods (questionnaires, interviews, observation, experiments, use of log files and so on) which can yield information about users' needs, capabilities, and behaviour that may influence decisions on editorial policy and dictionary design.

Among so much varied research activity, there is inevitably some unevenness in quality. But this hardly justifies the view of Bergenholtz and Bergenholtz (2011: 190) that 'most of the studies of dictionary usage [have been] carried out in the most unscientific way imaginable, as they were conducted without any knowledge and without use of the methods of the social sciences'.[2] This does not chime with my experience. In creating a user-profile – the first prerequisite for a good dictionary – lexicographers have much to learn from studies of dictionary use, and people like Yukio Tono, Robert Lew and his colleagues, Hilary Nesi, and Paul Bogaards (to name just a few) have produced work of great value and clear relevance for anyone involved in making dictionaries. But the research methods they use have their own theoretical underpinnings, drawing on disciplines such as statistics and social science, so there is a case for seeing dictionary user research as an independent field – and one whose applicability to practical lexicography is not in doubt. Consequently, it will not be considered here as an aspect of 'lexicographic theory', and no more will be said about it.

2.2. *L. V. Shcherba*

In 1995, the IJL published an English translation of a monograph by the Russian lexicographer Lev Vladimirovich Shcherba (originally written in 1940) with the encouraging title 'Towards a general theory of lexicography'. With the metalexicographer's characteristic fondness for typologies, Shcherba constructs his article around a series of what he calls 'oppositions'. In six binary choices, he distinguishes pairs such as the 'academic' and 'informative dictionary', the 'defining' and 'translating' dictionary, and the 'ordinary' and 'ideological' dictionary (what we would call dictionary and thesaurus). It has to be said that much of this is either self-evident or oversimplified: no-one has any trouble seeing the difference between a dictionary and a thesaurus, whereas the neat distinction between an 'encyclopedic dictionary' and a general one (Shcherba's second 'opposition') is a lot easier to make in principle than to apply in practice.

It becomes clear, though, as one reads his observations on the different categories of dictionary, that Shcherba has thought hard about the big lexicographic questions. He knows about the difficulties of 'seeking out all the separate meanings of a word' (Shcherba 1940/1995: 326). He addresses questions such as when does a figurative usage acquire the status of a dictionary sense, how fine-grained should one's analysis of a polysemous word be, and what is the proper role of example sentences. In discussing technical terms, he draws a distinction between definitions appropriate for a specialized dictionary, and explanations suitable for a general dictionary aimed at non-specialists (325). One only has to read Shcherba's analysis (327–9) of the Russian word *igla* ('needle') to recognize someone who is keenly aware of the issues and has personally grappled with them in real dictionary projects. For this reason alone, working lexicographers will empathize with him.

But in the end, Shcherba provides little in the way of guidance, beyond telling us how he himself resolved specific questions. He describes, for example, a decision he made on one particular issue when compiling the *Dictionary of Russian* of the Academy of Sciences: 'I decided', he says, that a meaning discussed earlier 'does not exist in standard Russian'. But the effect is spoiled when he continues: 'However, even this may be a debatable point' (328). Inclusion criteria are discussed in similarly inconclusive terms. What he refers to as a 'concordance' – a dictionary such as the *Thesaurus Linguae Latinae* – should aim, in his view, to include 'absolutely every word encountered in a language', including hapaxes (325). That is the easy part. For other types of dictionary, 'infinite variations are possible' but a standard general dictionary should include 'all words with an indisputable place in the language' (331). How we might recognize such words is not explained.

According to Donna Farina, who translated 'Towards a general theory', Shcherba's work has been extremely influential: 'Many ideas that are standard in both lexicographic theory and practice today can be found in [Shcherba's writings]' (Farina 1995: 300). More recently, Tarp has described Shcherba's ideas as 'revolutionary' (Tarp 2008: 21). This surely overstates the case. Shcherba's typologies are hardly groundbreaking, and there is not much that is novel in his discussion of the kinds of challenge which lexicographers face. We have known what the problems are for a long time: people like Johnson and Murray confronted all these issues before Shcherba was even born (and made valiant efforts to overcome them). What lexicographers are looking for is

solutions (not endless restatements of the problems), and they won't find them here.

There is one final point of interest. In a footnote to 'Towards a general theory', Shcherba reveals his intention to devote a future study 'to the nature of the word, its meaning and use, and its relations with other words' (ibid. 344, footnote 2). Now that sounds like a real theory, and one that might have genuinely helped working lexicographers. Sadly, Farina informs us, 'the additional studies that Shcherba planned were never carried out'.

2.3. *H. E. Wiegand*

Even if I were more proficient in reading German, it would be impossible to do justice to Wiegand's voluminous and erudite *oeuvre*. Fortunately for non-German speakers, Wiegand occasionally writes papers in English. Better still, a selection of his papers was translated into English, in response to requests from 'scientists whose native language is not German' (Wiegand 1999, Editors' Preface: 1)

The chapters in this collection range from the highly specific ('On the Meaning Explanation of Sentence Adverbs in Monolingual Dictionaries', 113–138) to more abstract philosophical speculation ('Thinking about Dictionaries: Current Problems', 55–94). On the basis of my limited experience of reading Wiegand, I suspect he is at his most interesting in articles like the latter. In the expansive chapter 'Thinking about dictionaries', Wiegand ponders questions such as what native-speakers really know about their own language, how much real-world knowledge one needs in order to use a dictionary successfully, and the influence of culture and experience on the way words are organized and connected in our mental lexicons. His penchant for granularity is well illustrated in an excursus on dictionary consultations which are *not* motivated by the need to resolve an immediate communicative problem. Even these anomalous look-up situations are meticulously dissected. There are for example '*didactic look-up situations*, which can be divided into the following subtypes…' Four distinct subtypes are then explained, and the discussion is rounded off with an elaborate diagram (Wiegand 1999: 76–7).

There is plenty of thought-provoking material here, but little that qualifies as a 'theory' on which lexicographic practice could be based. Another paper in this collection, 'Elements of a theory towards a so-

called lexicographic definition' (203–282), looks altogether more promising. But within a few pages, we are bogged down in what looks (to this lexicographer) like pointless disputation. For example, we are presented with this subentry from *Chambers Universal Learner's Dictionary*:

> **out on a limb**: having ideas or opinions not shared by others; in a dangerous or disadvantageous position

Wiegand immediately senses a problem:

> To complicate things, there is no agreement in the pertinent literature as to whether or not the text segments
> TS_5: having ideas or opinions not shared by others
> and
> TS_6: in a dangerous or disadvantageous position
> which give a semantic description of the colloquial expression *out on a limb*, should also be considered lexicographic definitions (205).

I am not sure what 'pertinent literature' Wiegand is referring to, but this feels like a fruitless discussion. There may indeed be problems with this entry: how, for example, is the semicolon to be interpreted (is this one definition or two?). But there would surely be no question in the mind of any dictionary user that the words which follow the canonical form **out on a limb** are intended as providing a definition of the phrase. There are important debates to be had about dictionary definitions: what they are for, what information they should optimally include, whether and when full-sentence definitions may outperform more conventional approaches, and (now) what the status and value of user-generated definitions might be (on which, see section 5). But devoting several pages to the question of whether the string 'having ideas or opinions not shared by others' is or is not a lexicographic definition does not seem especially useful.

It is true that Wiegand does address salient questions about definitional conventions, such as the idiosyncratic use of punctuation and parentheses (206–207) or the desirability or otherwise of 'substitutability' in definitions (230f.). But in the real world, the more problematical aspects of 'lexicographese' have largely been abandoned in modern dictionaries – precisely because lexicographers and publishers have recognized the ambiguities they give rise to and the unnecessary problems

they cause for users. It should be added, however, that these changes have been driven by dictionary-makers themselves, not by theoreticians.

Again and again, one is struck by a tendency to 'problematize': to see difficulty and complexity where little exists. The emphasis is less on 'how should we approach the task of defining?' (a subject on which lexicographers would welcome useful guidance) and more on 'what name should we give to this or that element in the entry?'. The chapter under discussion includes a series of diagrams which at first sight appear dauntingly complex, but when one looks more closely it is hard to escape the impression that the complexity has been manufactured. A good example is shown in Figure 1.

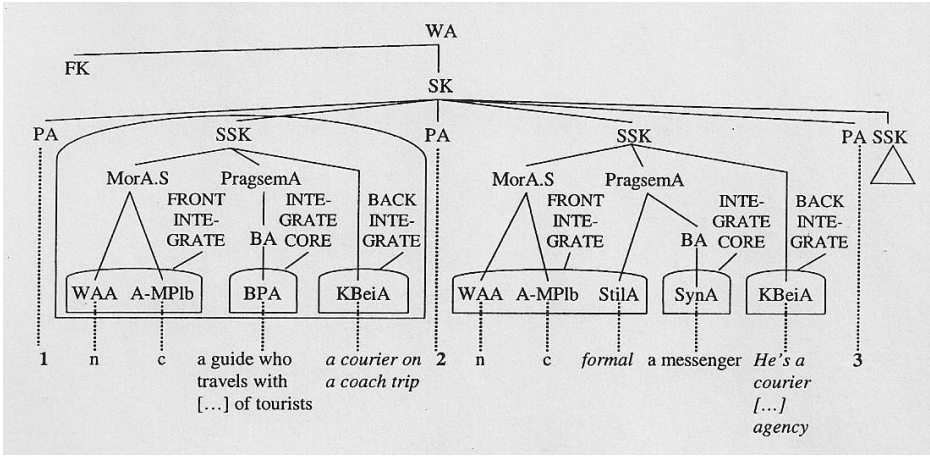


Figure 1. Diagram from Wiegand 1999: 210.

This is a graphic representation of the components of a simple entry for the word **courier** from a learner's dictionary: the entry has three senses, each with its own grammar code, definition and example; one of the senses also has a style label (*formal*). The diagram is alarmingly captioned: 'Partially visualized and annotated structural graph of the simple integrated hierarchical microstructure of DA_6 [the 'dictionary article' in question]'. The abbreviations reflect the names given to each element in the original German, but there is a good deal of scope for simplifying things. An explanatory sentence on the following page is worth quoting in full: 'The partial string WAA < A-MPlb forms the front integrate (cf. fig 8), i.e. the partial string which always immediately follows the item giving polysemy (PA) and which is always positioned immediately before the integrate core' (Wiegand 1999: 211). A simple

translation: the grammar code follows the sense number and precedes the core of the entry (definition, example etc). There is much more in the same vein, and the reader repeatedly goes through a laborious process of decoding an arcane explanation – only to find that it describes something familiar and straightforward.

The entry for **courier** anatomized here contains just 10 datatypes (the DANTE database worked with 94), so it is not especially complex. What happens in a real dictionary project is something like this: those responsible for editorial and database design identify all the possible components of entries in the planned dictionary; they then describe these components and supply information on each of them (when to use them, what kinds of information they should contain) in the Style Guide; the entry components are then incorporated into a DTD ('Document Type Definition') which defines the internal syntax of the entry: which components are compulsory, what order they can or must come in, and so on. A DTD of Wiegand's diagram would look something like this:[3]

```
<!ELEMENT DictEntry          (HWDGroup|LUCont)>

<!ELEMENT HWDGroup           (HWD | FORM)

<!ELEMENT LUCont             (POS | GRAM |REG | SYN |STYLE|
MeaningGp)
<!ELEMENT MeaningGp          (DEF | EX)
```

Figure 2. DTD of the entry described in Figure 1.

For the average lexicographer, this is pretty routine. In a more recent piece on 'hybrid textual structures', described as a 'contribution to the theory of dictionary form', Wiegand illustrates his arguments with no fewer than 25 diagrams. Diagram 3, for example, dissects the word *specialist* and is captioned 'A simple commentated structural graph of the abstract (and isomorphically concrete) hierarchical simple integrated pure entry microstructure' (507). Things get progressively more complicated, and I confess to finding the later diagrams indecipherable. This is an impressive piece of work, but one is bound to ask: who are these elaborate descriptions aimed at? D.A. Cruse, reviewing the 1999 Wiegand collection, was equally puzzled. He notes Wiegand's penchant for fine-grained categorization: 'Dictionary entries are meticulously partitioned and the varieties of their meaning-imparting components exhaustively catalogued' (Cruse 2001: 142). There is no question that Wiegand's

contribution to the description of dictionary structures is unrivalled. But, like me, Cruse finds himself speculating about the intended audience: 'As a lexical semanticist (and a general reader) I confess I found a lot of this heavy going … and it was not obvious to me as an outsider that it would have much practical import. I suspect the real intended audience is other metalexicographers' (142).

2.4. *Metalexicography: some conclusions*

Metalexicographers have a strong interest in 'naming of parts', and there may be advantages in proposing a stable nomenclature for the full range of possible dictionary entry components. But one is reminded of a comment by the late Larry Urdang. Reviewing Hartmann and James' *Dictionary of Lexicography*, Urdang found himself 'confused by a string of entries in the Ds, namely, **diaconnotative information, diaconnotative markedness, diaevaluative information, diaevaluative markedness, diafrequential information, diafrequential markedness, …**through **diaintegrative . . . , diamedial . . . , dianormative . . ., diaphasic .. . , diastratic .. . , diasystematic . . . , diatechnical . . . , diatextual . . . , diatopic(al)**.' He concludes that 'It seemed odd that such terms of art in a field in which I have lived and worked for many years would have eluded me as well as other professionals I have queried' (Urdang 2000: 38–39).[4]

Metalexicographic accounts of dictionary structure are as legitimate an intellectual exercise as any other. But would metalexicographers argue that, if practitioners (like Urdang and myself) would only familiarize themselves with this material, and apply it in their work, the result would be better dictionaries? For many of us at the sharp end of lexicography, there is a sense that Wiegand and his followers occupy a parallel universe, in which people not directly involved in dictionary-making construct theoretical models, which the dictionary-makers largely ignore – not through any antipathy towards theory *per se*, but simply because they can see no practical use in them. This feeling is only exacerbated when one reads that the ideas of Shcherba, Wiegand and their ilk 'have dominated the lexicographic debate' over the last few decades (Bergenholtz and Tarp 2003: 172). This must be a debate from which I have been absent: for me, and I suspect for most lexicographers, the salient issues of recent decades (which are well represented in the Euralex archive) include questions like:

- corpus design, and the appropriate use of corpus data
- the relationship between lexicography and natural-language processing
- the nature of word senses, and their relationship with syntactic and other contextual features
- the effectiveness of different approaches to defining
- the lexicographic treatment of multiword expressions
- the automatic extraction of lexical data from corpora

…and much else. It is only fair to say that Wiegand's interests are wide-ranging, and he continues to engage with current debates as new paradigms emerge. But, next to the big issues confronting lexicographers, many aspects of metalexicography (in particular, the passion for complex typologies, minute categorization, and exhaustive nomenclature) can often seem quite irrelevant.

3. The Aarhus School

3.1. *The theory of lexicographical functions*

Henning Bergenholtz, his colleagues, and his many adherents have written extensively about what they call the 'theory of lexicographical functions'. The Aarhus School's output in recent years has been prodigious, and I hope this (necessarily brief) account does them justice. Sven Tarp has traced the genesis of these ideas over more than 20 years (Tarp 2008: 33–39), while Yukio Tono provides a helpful summary (Tono 2010: 2–5).

I will start by quoting from the horse's mouth. In planning a dictionary, lexicographers need to create 'a profile of the intended user group and a typology of the user situations where problems or needs may pop up that can be solved by providing lexicographic data' (Bergenholtz and Tarp 2003: 173). Here (and in numerous other publications) they discuss the range of user needs and situations of dictionary use in great detail. Armed with this knowledge, 'lexicographers can determine which kind of data should be prepared and incorporated in the dictionary in order to assist each specific type of users in each type of user situation' (ibid.175). A given dictionary's provision of data which is designed to match the specific needs of its target users is defined as that dictionary's

'lexicographic function' (ibid. 176, cf. Tono 2010: 3) – hence the name of the theory.

An especially fine-grained exposition of these ideas appears in a long chapter entitled 'General Theory for Learner's Dictionaries' (Tarp 2008: 125–171). Tarp meticulously catalogues the characteristics of different types of learner, listing the many variables which combine to describe a specific dictionary user. These include factors such as the learner's mother tongue, the extent of their pre-existing cultural knowledge, their motivation for learning another language, and the degree to which they are exposed to that language in their daily lives. This section (Tarp 2008: 136–146) is followed by an equally exhaustive analysis (ibid.: 146–166) of situations in which a learner may need to consult a dictionary. For each learner type and each situation of use, specific categories of lexicographic data are required in order to meet specific needs. Tarp concludes: 'It is a dictionary's functions that …determine which data it should contain and how this data should be structured and made accessible' (168).

Who could argue with any of this? The Aarhus School's focus on users and their needs, and its detailed review of the variety of situations in which people might find themselves consulting a dictionary, is to be welcomed.

3.2. *Function Theory in historical context*

Whether these typologies – useful though they are – could be said to constitute a 'theory' is a matter of opinion. But I am puzzled by the claim that 'it was not until the appearance of the "modern theory of lexicographic functions" that a theory was developed that takes the users, the user needs and the user situations as the starting point for all lexicographic theory and practice' (Bergenholtz and Tarp 2003.172).[5] This will come as a surprise to anyone with even a casual acquaintance with the literature. As Tono politely points out: 'The idea that dictionaries should be based on their users is actually not new' (Tono 2010: 3). Atkins and Rundell (2008) consistently assert the primacy of users and their needs, and devote a whole chapter to user-profiling. This principle is invoked in the book's introduction (and repeatedly thereafter): 'The most important single piece of advice we can give to anyone embarking on a dictionary project is: know your user. … the content and design of every aspect of a dictionary must, centrally, take account of who the users will

be and what they will use the dictionary for' (Atkins and Rundell 2008: 5).

But we make no claim to originality. The same point has been made repeatedly over many years. Hartmann (1987), for example, devotes a whole chapter to user studies, and quotes approvingly the conclusions of a 1962 study which states that 'Dictionaries should be designed with a special set of users in mind and for their specific needs' (Householder and Saporta, quoted in Hartmann 1987: 11).

A little further back, Hornby himself insisted on the importance of understanding users' needs. While his insights into phraseology and collocation developed through the research he undertook with Harold Palmer (e.g. Cowie 1998: 7–8), his approach to designing and populating his groundbreaking learner's dictionary was rooted in his experience as a language teacher. In his own words: 'If [a] dictionary is designed for a special class of users, their special needs must be taken into consideration' (Hornby 1965: 104). But why stop here? On the title page of the first recognized English dictionary, the author explains that he will deal with 'hard' English words, and continues (Cawdrey 1604):

> With the interpretation thereof by
> *plaine English words, gathered for the benefit &*
> *helpe of Ladies, Gentlewomen, or any other*
> *vnskilfull persons.*
>
> Whereby they may the more easilie
> and better vnderstand many hard English
> wordes, which they shall heare or read in
> Scriptures, Sermons, or elswhere, and also
> be made able to vse the same aptly
> themselues.

As far back as 1604, we find an explicit recognition of both the target users of the dictionary ('Ladies, Gentlewomen, or any other unskilfull persons') and the receptive and productive contexts in which these users may find themselves consulting it ('which they shall heare or read in Scriptures, Sermons, or elswhere, and also be made able to use the same aptly themselves'). The Aarhus School rightly emphazises the importance of understanding the needs and capabilities of dictionary users. This is, and always has been, at the core of what good lexicographers do. But a

little background reading would tell them that this is by no means a novel insight.

3.3. *What Function Theory doesn't address*

Bergenholtz and his collaborators would not be the first people to overstate the originality of their ideas. A more substantive criticism, however, is suggested by Yukio Tono. In discussing the various situations of dictionary use proposed by Bergenholtz ('cognitive', 'communicative', 'operative', and 'interpretive'), Tono points out that surprisingly little is said about how one would select specific information categories to match these different types of user or use. He concludes that 'function theory does not have the power to produce anything new or different unless deliberate selections and weighting of the information specific to particular functions are specified' (Tono 2010: 15).

More broadly, it becomes clear that the theory's focus is on the ways in which information is selected and presented to the user – but nothing is said about where this information comes from in the first place. Function theory deals only with what Atkins has called the 'synthesis' aspect of lexicography (Atkins 1993: 7–8; Atkins and Rundell 2008: 102–103). Synthesis is the process through which lexicographers select and organize those facts relevant to a specific dictionary and its users: 'Each new synthesis produces a different dictionary…aimed at a different market and designed with a different group of users in mind' (Atkins 1993: 7). But synthesis presupposes an earlier stage (which Atkins calls 'analysis'), in which relevant forms of evidence are mined in order to provide lexicographic raw materials – a lexical database, if you like – which can form the basis for multiple configurations. As Kilgarriff notes in his review of the Aarhus School's most recent collection (Kilgarriff 2012: 28), the book is 'mostly concerned with delivering information to the user …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!'.

Function theory, it transpires, has little to say about what many of us see as the core task for lexicographers: analysing the evidence of language in use in order to identify what is likely to be relevant to dictionary users. In the final chapter – which reports the conclusions reached at a recent Aarhus School symposium – we learn that the participants 'agreed on a rethinking of some of the approaches commonly

used for dictionary-making. One of them is the role of corpora' (Samaniego Fernandez and Pérez Cabello de Alba 2011: 309). Alarm bells start ringing at this point, and ring even louder as we read that we are now 'confronted with the sad truth that lexicographers are required to adapt their work and their data selection …to the results generated by the computer' (ibid.). This terrifying vision, of victimized lexicographers being forced to submit themselves to the tyranny of real language data ('results generated by the computer') will have little resonance with those of us (the majority, I suspect) for whom the corpus revolution has been an overwhelmingly positive development.

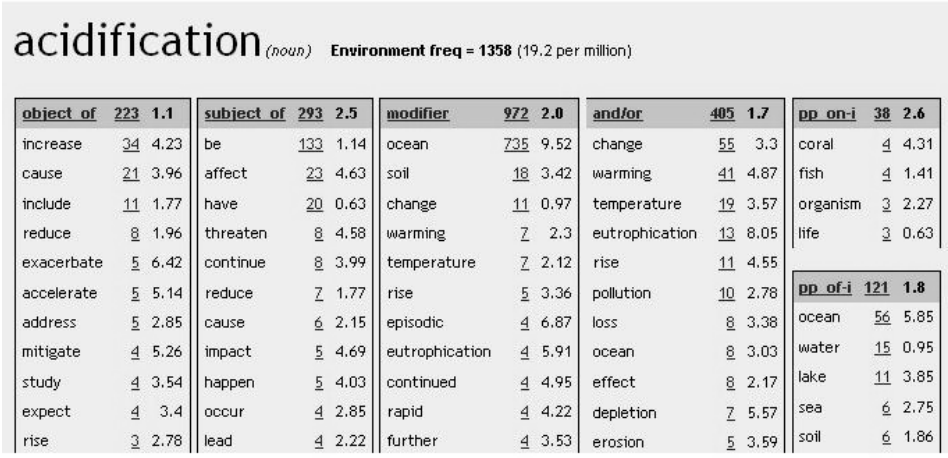
Hostility towards corpora is implicit in another of the symposium's conclusions: 'There are two important criteria when evaluating the … quality of a dictionary: (i) whether the user can find the item that contains the answer to the question that prompted the search, (ii) and how long the research took' (ibid.). I beg to differ. These are indeed important criteria, and a significant challenge for dictionary-producers. But the issue is being actively and intelligently addressed (e.g. Lew 2012 in press), and the challenge will diminish as digital resources benefit from improved search algorithms. Once we have worked out how to do this optimally, the criteria referred to above could become quite marginal indicators of dictionary quality. To put it another way, good 'findability' may remain a critical goal, but achieving it could become a trivial task. Content, on the other hand, will always be important. On this front, the Aarhus School's plans for 'rethinking' the role of corpora (does rethinking mean downgrading?) do not bode well.

3.4. *Hostility towards linguistics*

Bergenholtz and his acolytes reserve their deepest antipathy for linguists, whom they see as attempting to 'colonize' lexicography (see also 4.1 below). One of their arguments against involving linguists in dictionary-making is that a majority of dictionaries are not about general language anyway, so 'the cooperation of expert linguists will be necessary for, at most, 30 per cent of [dictionaries]' (Bergenholtz and Bergenholtz 2011: 189). This is not persuasive. Special-subject dictionaries may outstrip general-purpose dictionaries in purely numerical terms, but this is not what people mean when they talk about dictionaries, and it is not what most lexicographic activity (and metalexicographic discussion) is geared towards.[6] As Kilgarriff says, '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' (Kilgarriff 2012: 27).

Even on its own terms, Bergenholtz's argument fails to convince: the words described in specialized dictionaries often have their own patterning and preferences, and users will benefit from knowing what they are. The following extract (Figure 3) from a Word Sketch for *acidification* (taken from an environmental science corpus in Macmillan's data collection) supplies valuable information about (*inter alia*) the nouns that typically premodify it, the verbs that frequently have it as an object, and other nouns with which it often appears in an 'and/or' relationship. Just as with more everyday vocabulary, contextual data of this type contributes to an understanding of the word's meaning, and is almost essential for successful productive use. In a specialist dictionary, therefore, the information users need is partly scientific (so subject-specialists should be involved), but also partly linguistic (so corpus data should be consulted).[7]



acidification *(noun)* **Environment freq = 1358** (19.2 per million)

object_of	223	1.1
increase	34	4.23
cause	21	3.96
include	11	1.77
reduce	8	1.96
exacerbate	5	6.42
accelerate	5	5.14
address	5	2.85
mitigate	4	5.26
study	4	3.54
expect	4	3.4
rise	3	2.78

subject_of	293	2.5
be	133	1.14
affect	23	4.63
have	20	0.63
threaten	8	4.58
continue	8	3.99
reduce	7	1.77
cause	6	2.15
impact	5	4.69
happen	5	4.03
occur	4	2.85
lead	4	2.22

modifier	972	2.0
ocean	735	9.52
soil	18	3.42
change	11	0.97
warming	7	2.3
temperature	7	2.12
rise	5	3.36
episodic	4	6.87
eutrophication	4	5.91
continued	4	4.95
rapid	4	4.22
further	4	3.53

and/or	405	1.7
change	55	3.3
warming	41	4.87
temperature	19	3.57
eutrophication	13	8.05
rise	11	4.55
pollution	10	2.78
loss	8	3.38
ocean	8	3.03
effect	8	2.17
depletion	7	5.57
erosion	5	3.59

pp_on-i	38	2.6
coral	4	4.31
fish	4	1.41
organism	3	2.27
life	3	0.63

pp_of-i	121	1.8
ocean	56	5.85
water	15	0.95
lake	11	3.85
sea	6	2.75
soil	6	1.86

Figure 3. Part of a Word Sketch for acidification from an environmental science corpus.

3.5. *Function Theory: some conclusions*

The Aarhus School is nothing if not prolific. A common theme in its writings is what seems to me an excessive concern to establish lexicography as 'an independent scientific discipline'. But as Bogaards point out, a 'dependence on other sciences does not deny an independent status to the field of lexicography' (2010: 318). Even Tarp concedes (rightly) that lexicography, for all its specific concerns, draws upon a wide range of other ideas: it is 'a discipline particularly characterized by its big interdisciplinary vocation and its cooperative and integrating nature' (Tarp 2010: 461).

Few lexicographers would take issue with the Aarhus School's commitment to prioritizing the needs of dictionary users. It may be an unoriginal and even banal position, but there is no harm in reasserting it. But for anyone who has spent much time analysing language data in order to write dictionary entries, their antipathy towards corpora and linguistics is puzzling. Bogaards (2010: 316) takes the view that Function Theory 'is not a theory in any sense given to that notion in modern methodology'. I am agnostic on this point, but a more serious criticism is that it 'lacks any form of empirically verifiable or falsifiable hypotheses' (ibid.). The Aarhus School claim that 'the subject field of lexicography is dictionaries, a human-made product, whereas the subject field of linguistics is language' (Bergenholtz and Tarp 2003: 172) – ergo, linguistics has little value as an 'input' to lexicography. This is simply wrong. The study of dictionaries, as artefacts, is the subject-field of metalexicography. The goal of lexicography itself is the creation of dictionaries – and dictionaries attempt to describe the way language works. As Kilgarriff points out, 'if we had a database containing all the facts and generalizations 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' (Kilgarriff 2012: 29). The fundamental weakness of Function Theory is its failure to engage with the question of where dictionary content comes from. Implicitly, the 'analysis' part of lexicography is dismissed as if it were a mere technicality. But for most of us, this is an essential part of what making dictionaries is about.

4. Linguistic theory and lexicography

4.1. *Why consult linguists?*

Before I knew better, I thought dictionaries were books about words. In fact, of course, they are books about language – so it is logical that those of us who make dictionaries should seek the advice of people who specialize in thinking about how language works. Not everyone agrees, however. A recurrent complaint of the Aarhus School is that linguistics has attempted to 'take over' lexicography. They refer in tones of outrage to 'linguistic colonialism' (Fuertes-Olivera and Bergenholtz 2011: 3), and lament the fact that 'linguistics has laid claim to the dictionary arena and its theory' (Bergenholz 2011: 2). This is a perverse line of thinking: colonialism is when you invade without being asked. McCawley's 1986 paper on 'What linguists might contribute to dictionary making if they could get their act together', could be seen as an example of unsolicited intervention, but he is an exception. In the main, the traffic has gone in the opposite direction, with lexicographers inviting the colonizers in. An early and notable example is the case of Sue Atkins and Charles Fillmore. At the 1988 Summer School in Computational Linguistics in Pisa, Atkins approached Fillmore and, having persuaded him of the benefits of using corpora for language analysis, proceeded to seek his advice in making sense of the data: a fruitful and enduring collaboration ensued, which led among other things to the idea of building a frame-based lexicon.

Collaborations like this are now common. One thinks, for example, of the DELIS project in the early 1990s – 'a cooperation between computational and theoretical linguists, lexicographers, and software builders' – coordinated by Ulrich Heid.[8] A quick look at the Euralex archive shows that Heid has been involved in (and continues to be involved in) numerous projects of this type, with an emphasis on automating the acquisition of lexical data from corpora. One example (among many) is reported in Docherty and Heid (1998), which describes a collaboration where linguistic theory and language-engineering techniques are applied to the practical task of revising and updating a commercial dictionary. Indeed, McCawley, as he catalogues the shortcomings of dictionaries in general, makes an exception of the *Longman Dictionary of Contemporary English* (then fairly new), whose approach to describing syntactic behaviour he finds 'heartening'. He continues: 'I am especially heartened by the not coincidental circumstance that LDCE has made more extensive use of the services of

linguists than has any of its predecessors' (McCawley 1986: 15–16). Though this is an oversimplification (Hornby's original learner's dictionary drew on linguistic expertise), the point is well made: lexicographers have much to learn from linguists and it is increasingly common for dictionary publishers to consult linguists on points of editorial policy.[9]

Why should lexicographers enlist the help of linguists? When lexicographers analyse language data in order to gather the raw materials for a dictionary entry, there are two things they need to know: which bits of the data – which linguistic facts – are relevant to the task in hand? And how can we be sure that our account is complete, and that nothing important has been missed? When we scale this up from individual entry to complete dictionary, a third question arises: have we included all the relevant lexical items, and have we given the same treatment to items which share common features; or in other words, how do we ensure that our description of a language is systematic? For lexicographers, linguistic theory offers the promise of finding answers to these questions – finding what Zgusta called 'the underlying regularity'. This section will not only assert the potential value of linguistics for lexicography, but will demonstrate its actual value by reference to specific features in published dictionaries

4.2. *Sinclair and the COBUILD project*

Discovering 'underlying regularity' is a central theme in the work of John Sinclair. An outstanding example of linguist/lexicographer collaboration is the COBUILD project of the 1980s, led by Sinclair (a linguist) and based in a university linguistics department. Sinclair's much-quoted distinction (Sinclair 1991: 109ff) between the 'open-choice principle' (in which any word can occur in a 'slot' where it satisfies grammatical and semantic constraints) and the 'idiom principle' (where our lexical choices are, in practice, far more limited, and recurrence is more salient than randomness) embodies ideas latent in his earlier work but crystallized through the experience of studying corpus data. Linguists like Palmer, Firth and Hornby had already observed the formulaic character of language, but Sinclair's approach was 'marked by his insistence on analysis of corpus evidence in order to establish details of normal phraseology' (Hanks 2008: 228). A related insight was the interdependence of phraseology and meaning. Early attempts to describe

the way contextual features support word-sense disambiguation include Stock (1984) and Atkins (1987), both associated with the COBUILD project. More recently, the ambitious 'Corpus Pattern Analysis' (CPA) project of Patrick Hanks (editor of the first COBUILD dictionary) aims to establish – at the level of individual headwords – the precise ways in which patterns of usage are associated with word meaning (e.g. Hanks 2002). [10] The ideas of Sinclair and his co-workers have profoundly influenced lexicographic practice and products in the last quarter century, and will no doubt continue to do so as their implications are worked through in the digital media now engaging our attention.

4.3. *When theory replaces intuitions*

Lexicography involves an endless series of judgment calls, as one scans language data and tries to extract what is important. Good lexicographers instinctively make the 'right' calls most of the time – but that leaves too much to chance. Confronted by these two (superficially similar) sentences:

(1) *She shot him in the leg*
(2) *She shot him in the kitchen*

most of us intuitively sense that the PP in (1) contains information relevant to an account of the verb *shoot*, while the PP in (2) does not. But why? FrameNet's 'Hit_target' frame clarifies the distinction: example (1) includes an instantiation of the frame element 'Subregion' (realized by the name of a body-part), while *in the kitchen* in (2) provides peripheral information which does not tell us anything about the valency of *shoot*. Lexicographers follow Style Guides, and these support decision-making at each point in the compilation process, providing detailed guidance for handling every individual component of a dictionary entry.[11] The specific example here relates to describing syntactic preferences, and it is clear that a theoretical perspective helps us to see what is lexicographically relevant. It also helps us to be systematic, because a frame-based analysis tells us that many other verbs (like *hit, bash, punch* and *whack*) behave somewhat like *shoot*. This process, where the editorial guidelines are informed by relevant linguistic theory, is repeated across the piece, enabling lexicographers to see underlying systems rather than rely on their intuitions.

4.4. *Linguistic theory: general benefits*

On the day I arrived in Birmingham as a new recruit to the COBUILD team, I was handed a copy of John Searle's *Speech Acts* (1969). This wasn't because Searle's ideas would be of direct operational usefulness when writing entries for the COBUILD dictionary. Rather, it is an example of the kind of book lexicographers ought to read 'more for their consciousness-raising discussion than for immediate applicability' (Atkins 1993: 19). Among many other examples, Cruse's work on lexical semantics (Cruse 1986, 2004) stands out, and is worth recommending to anyone embarking on a career in lexicography. In a different way, the bracingly contrarian Anna Wierzbicka is well worth engaging with too. Reading her sometimes feels like having an argument with someone who makes a persuasive case for a political position which you profoundly disagree with. She makes tough demands of lexicographers and has no time for intellectual laziness. A typical quote: 'Concepts encoded in natural language are, in a sense, vague…but this does not mean that their semantic description should be vague, too. The challenge consists in portraying the vagueness inherent in natural language with precision' (Wierzbicka 1990: 365). Anyone who writes things like this has to be worth reading.

4.5. *Prototype theory: a preference for 'preferences'*

Prototype theory is a good example of a body of theoretical work which has both general 'consciousness-raising' value *and* direct applicability to many areas of lexicography. Geeraerts explores the relevance of prototype semantics to practical dictionary-making, concluding that it is 'well suited as a theoretical basis for a lexicographical metatheory, since it accurately models the kind of semantic phenomena that lexicographers have to face up to' (Geeraerts 1990: 210). Atkins and Rundell discuss its relevance to the tasks of word-sense disambiguation (2008: 276–280) and defining (417–419). A prototype approach can be seen at work, too, in the entry structure of the *Oxford Dictionary of English*, with its use of 'core senses' and associated 'subsenses'. The dictionary's Introduction does not explicitly mention prototype theory, but its influence is clear: one only has to look, for example, at ODE's entry for **climb** to see how its structure reflects Hanks' discussion of this verb's core (or prototypical)

meaning and the ways in which this is developed in subsenses which approximate in varying degrees to the prototype (Hanks 1994).

In explaining the genesis of a more recent model of language, Hanks describes his 'theory of norms and exploitations' (TNE) as 'a bottom-up theory, created in response to the general question, how can we account for the ways in which people use words to make meanings?' (Hanks 2009: 4). At its heart is the contention, driven by observing language in use, that the rules governing the 'normal, conventional' use of words are 'intertwined with a second-order set of rules governing the ways in which those norms are exploited' (ibid.: 5). One consequence of this, which has relevance for any lexicographic account of meaning, syntax, or collocation, is that it makes more sense to think in terms of 'preferences' than 'restrictions'. For lexicographers, this is an important distinction: it not only helps us separate signal from noise in corpus data, but also shows us why we should not even attempt a description of a word's behaviour that would account for every possible instantiation in text. TNE provides support for decisions about what to include in dictionaries, and what to say about the words we do include – and in a sense it authorizes lexicographers to ignore, with confidence, anything in the data which doesn't conform to the norms that a given dictionary sets out to describe. The influence of prototype theory is explicitly acknowledged: 'First and foremost, TNE is a theory of prototypes and preferences' (ibid.: 5).

4.6. *Some specific cases*

Theoretical ideas from linguistics have an important role in contributing to policy formulation (at the design and planning stages of a dictionary project) and in informing individual editorial decisions (during the compilation phase). In some cases, the application of a theory is quite explicit, and a few examples will be given here.

4.6.1. *Apresjan and 'regular polysemy'*. Dictionary editors have long known that words belonging to distinct lexical sets tend to behave in similar ways. Given the way work is parcelled out on big projects, it is likely that a different editor will be faced with producing the entry for *Pisces* or *Wednesday* from the one tasked with compiling *Scorpio* or *Friday*. So it makes sense to agree standard entry formats at the outset. This approach has generally been adopted for a small number of closed

sets, but Apresjan's paper on regular polysemy (Apresjan 1973) suggested that it might have more extensive applications. Apresjan demonstrated that words belonging to many *open* sets exhibit common features, typically allowing a core meaning to generate one or more related meanings, which will often be signalled by a change in grammatical behaviour.[12] Applying these research findings, what have been called 'template entries' (also known as 'proformas') were used in the *Oxford-Hachette English-French Dictionary* and the *Macmillan English Dictionary for Advanced Learners*, and most comprehensively in the DANTE project. Sixty-eight proformas were developed during the project's planning phase, and the dictionary database was populated in advance with proforma features (Rundell 2012 in press: 23–24). Here, the application of a theoretical model to a practical operation delivers multiple benefits. Firstly, it helps us to achieve the goals identified earlier: identifying what is relevant in the data, ensuring nothing important is missed, providing a systematic account across the dictionary. Secondly, it makes life easier for lexicographers and speeds up the compilation process. Thirdly, because dictionaries deal with the whole of the lexicon, we often uncover fresh instances of the theory at work.[13]

In a thought-provoking critique of this approach, Swanepoel – while approving its goals – subjects it to a theoretical analysis and identifies a number of unresolved problems (noting for example that the notion of 'lexical set' is ill-defined). He proposes, as another route to definitional consistency and coherence, a broader category of 'lexical conceptual models' (formal representations of what users know about words, including their taxonomic and ontological features), which could be hyperlinked from conventional definitions in an electronic dictionary (Swanepoel 2010). There is much to think about here, but this is a positive example of linguist/lexicographer interaction: linguist-1 (Apresjan) develops a theory, independently of lexicography; lexicographers exploit this, in a practical and bottom-up way; linguist-2 (Swanepoel) applies theoretical insights to what the lexicographers have reported, and proposes a further development; now the ball is back in the lexicographers' court. A virtuous circle.

4.6.2. *Mel'čuk's lexical functions*. As its author has always made clear, the 'Explanatory Combinatorial Dictionary' (ECD) is not intended to be used as a practical dictionary. It is a formal lexicon, exhaustive and productively-oriented, based explicitly on a theory of language (the so-called 'Meaning-Text Model'), and with only 'a limited practical

purpose' (Mel'čuk 1988: 167). Yet, one specific feature of the ECD has proved of great value for practical lexicography. The 'Lexical Relations Zone' of an ECD provides a detailed account of collocation, based on a set of 'lexical functions' which collectively describe every conceivable category of lexical cooccurrence. On several dictionary projects I have been involved in (including the *Macmillan English Dictionary* and the *Macmillan Collocations Dictionary*), key collocation types have been extracted from Mel'čuk's huge inventory of functions to create useful checklists for lexicographers (see Atkins and Rundell 2008: 151–152 for details). Mel'čuk's ideas have informed numerous dictionary projects, a recent example being the 'DiCoInfo', a trilingual electronic dictionary of terms in computer science and the Internet. DiCoInfo uses lexical functions to provide a systematic account of collocation, translating the formalisms in Mel'čuk's theoretical version into accessible explanations for the end-user (L'Homme, Robichaud and Leroyer 2012 in press). Here again, a set of linguistic ideas has been adapted to impart theoretical rigour to a practical lexicographic task.

4.6.3. *Lakoff and Johnson's idea about metaphor*. Lakoff and Johnson's classic *Metaphors We Live By* (1980) is another of those 'consciousness-raising' texts that is worth reading to gain a better understanding of how language works. But its ideas have also been directly applied in dictionaries. *MEDAL* includes – as a 'language awareness' feature – around 60 'metaphor boxes' which 'provide contextualized examples of sets of metaphors relating to particular concepts, and explanations of the metaphorical mappings that link them' (Moon 2004: 196). A box at the entry for **conversation**, for example, begins by explaining the metaphorical concept ('A conversation or discussion is like a **journey**, with the speakers going from one place to another'), then illustrates how the mapping works in terms of specific lexical items, with example sentences like 'Let's **go back** to what you were saying earlier' and 'We **wandered** off the topic' (see Moon 2004 for details). In Macmillan's *Phrasal Verbs Plus Dictionary* (2005), an attempt was made to identify the semantic characteristics of the 12 most common particles used in phrasal verbs. Again, this was directly inspired by a Lakoffian view of metaphor and, though only partially successful, it represents an effort to replace apparent randomness (English phrasal verbs are notoriously difficult for learners) with something approaching a learnable system.

4.7. *Conclusions: lexicographers' autonomy*

There is still much to be done. On the one hand there are ideas from linguistic theory which have not yet been well exploited in dictionaries. Fillmore's concept of 'null instantiation', for example – where the omission of an expected frame element is authorized for some words but not for others – looks like a promising basis for an information-type in dictionaries (Atkins, Rundell and Sato 2003: 351–354). And in fact this was tried during the DANTE project – but abandoned when it became clear that the editorial policy was causing confusion because it had not been sufficiently well worked out. More broadly, few of the theoretical ideas under the general heading 'pragmatics' have yet found their way into dictionaries. Their importance is recognized and some attempts have been made, but with little success so far. On the other hand, there are aspects of dictionary practice where more theoretical guidelines would be welcomed, a prime example being register. Thanks to research efforts in computational linguistics, good progress has been made towards automating the application of labels indicating domain (subject-field) or regional preferences (e.g. Rundell and Kilgarriff 2011: 275–276). But we are still some way off achieving anything like this for labels like 'formal' and 'informal'. Applying these labels remains a more subjective operation than we would like. Perhaps they are just too problematic to be useful – or perhaps what we lack is a robust theoretical model of register.

Lexicography has benefited enormously from its engagement with theoretical linguistics. But lexicographers and linguists have different agendas, so there is generally a process where linguistic theories need to be adapted in order to be of use in the specific environment of a dictionary. Geeraerts – well placed to comment as both a cognitive linguist and editor of the *Woordenboek der Nederlandsche Taal* – recognizes the value of Wierzbicka's exhaustive theoretical definitions, but insists that 'a definitional technique that is optimally justified from a theoretical point of view need not be so from the pragmatic point of view of practical lexicography' (Geeraerts 1990: 197). This is a common position: lexicographers have much to learn from theorists, but the particular goals of a dictionary (and the practical constraints within which it is created) preclude the uncritical application of theoretical ideas. The lexicographic convention of showing numbered senses is a good example of the disjunction between theoretical knowledge (we recognize that 'word sense' is an unstable category) and what we do in dictionaries (and what users expect us to do) – though new media, as we shall see, offer

opportunities for bringing theory and practice into closer alignment. The polysemy/homonymy distinction is another interesting case. Its theoretical validity is not in doubt, and its relevance in historical dictionaries is obvious. But many contemporary dictionaries (especially pedagogical ones) have abandoned homonymy as an organizing principle, on the grounds that it presupposes knowledge about word history which is not available to most users (Atkins and Rundell 2008: 280–282). A final example: Sinclair's model of language led to the introduction of full-sentence definitions (FSDs), and in a sense his theory requires them. Some dictionaries have opted to follow COBUILD in using FSDs systematically (e.g. the Spanish learner's dictionary DAELE being developed in Barcelona: Mahecha and DeCesaris 2011: 183). But others, for good pragmatic reasons, have adopted FSDs as a strategy to be used in some types of entry but not wholesale (Rundell 2006). This should dispel any notion that lexicographers are slavishly in thrall to the dictates of 'colonizing' linguists.

5. e-lexicography and the relevance of theory

5.1. *The world we live in now*

Dictionaries are going the same way as encyclopedias. In just a few years most activity has moved from paper to electronic platforms. For pedagogical dictionaries, whose users are mainly young (and therefore digital natives), the switch from old to new media is even more marked. Though 'electronic lexicography' – the use of digital media for delivering dictionary data – dates back at least as far as 1990, the pace of change has picked up dramatically in the last five years, after a leisurely start.

The migration from print to digital is the second big upheaval for lexicography in the last 30 years. The corpus revolution forced a major rethink of lexicographic practice in both 'analysis' and 'synthesis' modes (as well as changing our perceptions of how language works). Yet the changes it led to have been mainly 'internal', affecting the way lexicographers work and improving the reliability of their output. The end-product is still recognizably a dictionary, and for the average user the changes going on behind the scenes may be barely perceptible. But the new and ongoing digital revolution will be more disruptive. Its effects are 'external', in that it impacts directly on dictionary users, and is in a sense driven by their changing behaviour. Against this background, reappraisal

is unavoidable. The new (or rather, emerging) paradigm raises fundamental questions about what dictionaries are for (we will come to this later), and means that most of the old 'rules' need to be revisited.

The most obvious difference between old and new models is that dictionaries are no longer limited by space. Many of the familiar dictionary conventions – the codes and abbreviations, the compressed defining styles, the economical deployment of example sentences, the use of undefined run-ons, and so on – developed in response to the goal of providing as much information as possible in a limited space (Atkins and Rundell 2008: 20–23). Without these constraints, what should publishers do? Some, like the providers of the handheld dictionaries popular in east Asia, have responded by simply piling in more data: some of these devices hold over 100 separate dictionaries. But this approach, where a random collection of (originally printed) resources are stuck together with minimal integration, looks more like a knee-jerk reaction to the falling cost of storage than a properly thought-through policy. As many have observed (e.g. de Schryver 2003: 163f.), endless space shouldn't be a licence for 'swamping' the user with data just because we can. Robert Lew makes the useful distinction between 'storage space' (effectively infinite) and the space visible 'above the fold' on a computer screen, which he calls 'presentation space' This 'refers to how much can be presented (displayed, visualized) at a given time to the dictionary user' (Lew 2012 in press), and of course how much a user can be expected to process. In addition to what appears in this 'presentation space', there are also opportunities for exploiting 'the dynamic potential of electronic displays in other ways' (ibid.), notably through hyperlinking to other layers of information, or providing instant assistance by means of data which appears when the mouse hovers over part of an entry.

The implications are still being worked through. It may be possible, for example, to bring the divergent needs of regular and computational users into closer alignment. Swanepoel (1994: 20) notes 'the lack of systematicity in lexical semantic description, i.e. the lack of equal treatment of individual members of a specific syntactic and/or semantic class with regard to the kind of semantic information provided'. Paper dictionaries tend to favour usefulness over consistency, and so – to the despair of computationalists – they do not generally record the regular polysemy of *every* member of a set, leaving the more marginal items to fend for themselves (Kilgarriff 1994: 101–103). This is a sensible policy in the print medium, but if space is unlimited its main rationale disappears.

5.2. *A period of transition*

As the science-fiction writer William Gibson memorably observed: 'The future is already here – it's just not evenly distributed'. His point is well illustrated in the widely differing ways that dictionary-makers have responded to new conditions. There is a spectrum here. At one end, the online dictionary of the Real Academia Española preserves all the space-saving devices used in its print edition (abbreviations, tildes etc), along with its concise defining language and its old-style cross-references that you can't click on. At the other end are resources like Serge Verlinde's consistently innovative site for learners of French (the Base lexicale du français, or BLF), which makes imaginative use of the new technology and seems to be in a state of continuous improvement. Most online dictionaries fall somewhere in the middle: sensibly adjusting their metalanguage (so 'adj' becomes 'adjective'), providing full morphological information (see e.g. the 'conjugar' button in DAELE), making most content clickable (so you can jump to the entry for **lava** by clicking this word in the definition at **volcano**), and providing additional information through hyperlinks (such as the thesaurus in MEDAL).

If publishers sometimes look slow off the mark in adapting to the new situation, this isn't because they are unaware of the opportunities. More likely it is because of their 'legacy' data: reference materials originally developed for the medium of print. In an ideal world, we would pulp most of this and start from scratch, producing new resources optimally adapted to digital media. But this would be commercially insane. (Publishers faced a similar dilemma when corpora first became available: those starting with a clean slate, like COBUILD, had a clear advantage over competitors like Longman and Oxford, who had to adapt dictionaries produced in the pre-corpus era to take account of vastly improved language data – a painful process with often unsatisfactory results.) Even where there is no print legacy, problems can still arise. Wordnik never existed in print form, but its cutting-edge technology (which can automatically populate a new entry by grabbing images from flickr.com and sourcing up-to-the-minute examples from social media) is married with content from antiquated dictionaries – the type that define *pedantic* as 'of, pertaining to, or characteristic of a pedant'.

5.3. *Some specific features, and how they are affected*

There is much to be done, but the direction of travel is for dictionaries to exploit the electronic medium more fully. What is harder to predict is where this process will end (or even *if* it will end). There is no particular reason why dictionaries should survive in anything like their present form, but before we address that issue, it is worth looking at some specific areas where old and settled ideas are already coming under pressure.

5.3.1 *Inclusion policy: what goes in the dictionary*. On the Oxford Dictionaries website, a flowchart entitled 'How a new word enters our dictionaries' provides a useful account of conventional inclusion criteria.[14] Aspiring entrants to the dictionary have to overcome various hurdles: 'does it have a decent history of use?' requires a 'yes', but if 'its use [is] limited to one group of users', the next move is 'reject for now, monitor its use for possible future inclusion'. And so on. All of which makes sense when your dictionary appears in print form: if space is limited, robust criteria are needed to exclude what is not deserving. But what if space constraints no longer apply? The answer isn't straightforward because there are other factors at play. The dictionary's 'exclusiveness' taps into another traditional feature: its status as 'gatekeeper', as an 'authority' on language. Lexicographers may be uncomfortable with this role, but it tends to be foisted on them both by publishers (who claim authority) and by users (who ascribe it to the dictionary). Until now, that is. Where older users, discovering that something is not 'in the dictionary' might see this as a judgment on the word ('it's not a proper word, then'), digital natives – with their general (and well-founded) expectation to find what they are looking for on the Web – are more likely to see omission as an indictment of the dictionary. The extended timescale described in the Oxford flowchart reflects the old publishing cycle, where dictionaries would be updated every five years or so. But this no longer works. The last printed edition of MEDAL, for example, was published in 2007, just before the global financial meltdown which spawned (or popularized) so much new vocabulary (*credit crunch, subprime, quantitative easing* etc). With the main edition of the dictionary now digital, terms like this can be added as they become salient in public discourse. But this means that, with regular updates now the norm, traditional inclusion principles have to be re-assessed.[15] For many users up-to-dateness trumps 'authority' – a point made vigorously by Wordnik's Erin McKean at the 2011 e-lexicography conference. There

is no simple answer to the question of how we replace or update old criteria, but this is one of the areas where theoretical input would be useful.

5.3.2. *Controlled defining vocabularies (DV)*. Most English pedagogical dictionaries use a DV as a way of ensuring the accessibility of their definitions. There is a theoretical basis for this practice, as well as a body of (mainly supportive) user research (Atkins and Rundell 2008: 449–450) – though the issue is by no means uncontroversial. But when any word in a definition is clickable, everything changes. Compared with the old cross-referencing operation (from a non-DV word in a definition to its own dictionary entry), 'what used to be laborious page-turning and letter hunting, can now become a simple click of the mouse or even as little as hovering your mouse over the target' (Lew 2010: 293). The general principle – that definitions should always be easy for a dictionary's target users to understand – remains intact, but its implementation in the digital medium is clearly up for discussion. Lew (2010) considers the issues, and proposes a loosening of the old constraints, while advising against a free for all. But so far, we have no settled policy to replace the previous one.

5.3.3. *Dictionary examples*. When space was limited, even pedagogical dictionaries had to be selective about where examples sentences appeared. MEDAL, for example, has an 'asymmetric' policy, based on a distinction between receptive and productive vocabulary: core items (the 7500 high-frequency headwords shown in red) typically have a lot of examples, so that syntactic and collocational behaviour can be fully illustrated; whereas many low-frequency words have none at all. This is far from ideal, but was seen as an optimal way of using scarce available space. Now anything is possible. The idea of creating a direct link between dictionary entries and available corpus resources has been around for almost 20 years (de Schryver 2003: 167–172), and a number of examples of this approach are already in place. To mention just a few: in the BLF, users can 'research' a word by clicking the 'exemples de corpus' link, which provides a user-specified number of examples (of the headword or a longer string containing it) in a range of French corpus resources; users of the ANW can specify search criteria to find examples in the institution's 'integrated wordbank'; in a paper on the goals of the Danske Ordbog, Trap-Jensen refers to the aim of providing 'closer integration between a dictionary component and a corpus component in order to enable the users to make their own research on the spot and to provide a given

reference with additional example material on request' (Trap-Jensen 2006: 349).

As always, there are teething troubles. In Wordnik, for example, there is often a mismatch between its web-sourced (or crowd-sourced) examples, and the dictionaries that supply its definitions: thus the examples for **traction** all foreground its use in political discourse (*Senate conservatives mostly repeated the same tired attacks that failed to* ***gain traction*** *earlier in the week*), while the definitions come from dictionaries too old even to acknowledge the existence of this (relatively recent) meaning. More broadly, there is the serious computational challenge of matching corpus instances to specific dictionary *senses*. But this is not intractable.

5.3.4. *Using multimedia.* Lew (2010: 291) questions 'the traditional verbal orientation of lexicography', and discusses the use of multimedia features (see also de Schryver 2003: 165–167). One question that arises is whether we still need IPA: if we can hear what a word or phrase sounds like, do we need a graphic representation? This probably depends on the user, but Wordnik is one dictionary that has already dispensed with IPA or respelling systems. Sound effects are often the most effective way of 'defining' words which describe particular sounds. MEDAL has hundreds of these, for words such as *oboe, reggae, arpeggio, cough* and *ricochet*. For static images, the Web offers endless possibilities and these are beginning to be exploited (with varying degrees of success). User research has an important role as we experiment with multimedia options. Lew (2010: 297–299) cites recent work which suggests that static pictorial explanations often helpfully complement (or even replace) verbal explanations, and may also promote retention; whereas, surprisingly, animations seem to perform less well. As he concedes, though, 'more research is needed before we are able to identify the optimal combination of ways of presenting meaning in dictionaries' (ibid. 303).

5.3.5. *Entry structures*. Just like the hard-copy concordances used in the 1980s at COBUILD, the structure of an entry in a paper dictionary is unavoidably static. But electronic media offer the possibility of configuring entries to match the needs of different users. Thus Kosem and Krishnamurthy (2007: 3) propose a Dictionary of Academic English aimed at non-native users, in which word senses of particular relevance to academic discourse are 'promoted' to appear at the top of the entry. In

this proposal, the sense-ordering would be done by the publisher, but this could equally be achieved through adaptive technologies (on which, more below).

5.4. *Beyond the dictionary: some current trends*

The discussion so far assumes the long-term survival of dictionaries: enhanced and expanded, for sure, but still recognizable as dictionaries. But is this realistic? Dictionaries evolved to meet specific communicative needs: what does this word or phrase mean? how do I say it? what's its equivalent in my language? how can I use it correctly and idiomatically? 'The dictionary' is a well-embedded cultural artefact, but my hunch is that it won't prove very resilient in the long term. It is equally plausible to imagine that its heterogeneous functions might be better performed by separate, more specialized resources: among others, automatic translation tools, text-remediation software, or the kinds of tool described by Prinsloo et al. (2011), which are designed to guide users' lexical and grammatical choices in text-production mode. In this section we will look at a number of themes which point the way to future developments.

5.4.1. *No more binary distinctions*. The typologies proposed by people such as Shcherba, based on binary oppositions like 'dictionary vs thesaurus' and 'dictionary vs encyclopedia' are rapidly breaking down. Many online 'dictionaries' include translation and thesaurus features. Among the so-called 'aggregators' (sites like dictionary.com and thefreedictionary.com) design and functionality are often clunky, and content outdated and poorly integrated. But at the more serious end, there are positive developments: the Macmillan and Cambridge learners' dictionaries both include integrated thesaurus functions attached to every sense of every headword; the BLF has both a 'reverse dictionary' function and a translation tool; and the ANW's onomasiological (meaning to word) search option 'allows users to look for a word that they have forgotten or… can be used to find out whether there is a word for a certain concept' (Tiberius and Niestadt 2010: 748).

Meanwhile the boundary between dictionary and encyclopedia (always problematic) is increasingly irrelevant. Someone searching for *oligarch* may just want to know its generic meaning, but they might equally want to find out about oligarchies in ancient Greece, or be looking for information about powerful people in contemporary Russia. It is

unlikely they will identify their needs as specifically lexical or encyclopedic. What they are engaged in is 'search', for which the Web offers endless possibilities. And increasingly, the starting point for most people will be a search engine like Google – and they may or may not end up at a 'dictionary'.

5.4.2. *Adaptable and adaptive models*. The idea that an electronic dictionary should be a flexible object, allowing for customization to the needs of particular groups of user or of one specific user, has been around for some time (de Schryver 2003: 183–185). There are two main approaches. In an 'adaptable' model, users set their own parameters. Thus if accessing a varied library of resources, an individual user will find it more efficient to be able to view just those parts that meet his or her needs. For their planned database of idioms and multiwords, Bergenholtz and colleagues will 'allow every individual user to specify such settings and even to self-define a series of search combinations and display options' (Bergenholtz 2011: 8). This is an appealing idea, but the experience of the Danske Ordbog suggests that caution is in order. Having offered this kind of functionality, Trap-Jensen found that users 'were generally unable to analyze their needs ('I don't care if it is reception or production, I just want to know what the word means')' and concluded that, for this approach to be successful, users must be 'able to analyze their own needs in every look-up situation and pick the right button, [and] there is not much evidence to support such a rational user behaviour' (Trap-Jensen 2010: 1139).

Kwary (2012) shows the potential benefits of an 'adaptive' model. 'Adaptive hypermedia' is a major research area with relevance to Web applications of all kinds. In essence, it learns from user behaviour and responds accordingly (think of how an Amazon account works). In lexicographic terms, this implies that 'an online dictionary can be adapted to the needs of each dictionary user' (Kwary 2012: 35). To do this, 'the systems can adaptively select and prioritize the items which are most relevant to their users' (ibid). The key word is 'adaptively' because, as a user's needs and knowledge change, the dictionary continually alters and updates the way its content is configured. Something on these lines was foreseen by de Schryver and Joffe (2004), who analyzed users' log files in order to make improvements to an online dictionary. This was a manual operation, but they observed that 'Ultimately, the idea is that an automated analysis of the log files will enable the dictionary to tailor itself to each and every particular user' (188). The adaptive technology is

developing rapidly, so the prospects for personalized dictionaries are good.

5.4.3. *Crowd-sourcing and user-generated-content*. User-generated content (UGC) is a salient feature of activity on the Web. It ranges from the trivial (people sending in photos of snow to a TV weather show) to the serious ('citizen journalists' providing documentary evidence of human rights abuses). A similar range is evident in the lexicographic sphere. The Urban Dictionary, for example, has high entertainment value and its coverage of colloquial American English is unrivalled. But when a single term like *Republican* includes 256 (sic) subjective and often scatalogical 'definitions', we know we are not dealing with an entirely serious dictionary. Macmillan's experience with its crowd-sourced Open Dictionary suggests that the most fruitful areas where users can make a contribution are neologisms, regional varieties, and technical terminology. The latter is also a major feature of Wiktionary. Describing Wiktionary's strengths, Meyer and Gurevych (2012 in press) note that 'Each contributor has a certain field of expertise. This broad diversity of authors fosters the encoding of a vast amount of domain-specific knowledge'. In this sense it resembles Wikipedia. While lexicographers can fairly argue that general vocabulary is best left to them (you can be an expert on *homeopathy*, *permafrost* or the *nitrogen cycle*, but not on *decide*, *limitation*, or *dull*), a collaborative approach to describing and translating terminology has much to recommend it. A similar strategy has yielded impressive results for Eijiro Pro on the Web, a bidirectional Japanese-English dictionary whose exceptional coverage of technical vocabulary owes much to its crowd-sourced beginnings.

UGC is still a fairly new phenomenon and raises a number of questions. The involvement of subject-specialists, linked to a collaborative paradigm where users can improve or correct entries, should mean that factual errors are rare. On the other hand, the nature of the exercise may lead to widely varying approaches to entry-writing. This variability is something that professionally-constructed dictionaries take a lot of trouble to iron out (though whether this matters to the end-user is another question). In any case, Wiktionary provides entry templates to ensure a degree of internal consistency, and an optimistic view would be that the community of contributors will self-regulate to ensure that best practice generally prevails. Lexicography, especially multilingual lexicography for specialized domains, is a highly labour-intensive

business, and crowd-sourcing offers the possibility of quickly accumulating large amounts of lexical data at low cost.

There are downsides of course, most obviously the randomness of what contributors produce. Comparing Wiktionary's entries for **chloroplast** and **acidification**, we find that the first includes translation equivalents for Azeri, Icelandic, Italian, Turkish, and Portuguese, while the latter has ten translations (including Chinese and Finnish). This would not happen on a conventional dictionary project – but then again, would a project on such a scale ever get off the ground? A more substantial criticism would be that even entries for technical terms benefit from analysis of corpus data for the domain (see the point about *acidification*, 3.4 above), but perhaps this indicates an area where (expert) lexicographic input might complement the knowledge of subject-specialists. Interestingly, Meyer and Gurevych see Wiktionary as a project in which 'large communities, backed up by the phenomenon of collective intelligence, compete with expert lexicographers'. But one could equally imagine a more collaborative relationship. ABBYY's LingvoPro resource includes a function allowing users (if registered and logged in) to add their own translations, and a user forum for discussion and questions (which can be answered by other users). The Dewan Bahasa dan Pustaka (the government body responsible for the Malay language) also has a discussion forum, and one of its functions is to allow users to express a preference when two or three Malay equivalents are proposed for a new technical term.[16] So, given the different skills of lexicographers, translators, and subject-specialists, there are opportunities for collaboration. The old idea of the dictionary as an 'authority' on language may not survive these upheavals, but not all lexicographers will regret this.

5.4.4. *The self-updating dictionary*. Rundell and Kilgarriff (2011) charted progress in the automation of the various stages in creating a dictionary. Inevitably, things have progressed even in the short time since this was written. There is a good deal of computational research geared to the automatic identification of new vocabulary items as they emerge. Some of this is straightforward: it is easy enough for computers to spot a completely new word like *omnishambles* or *nanodrone*, and to assess its currency. Some aspects are more challenging, notably the task of identifying new senses of existing words, such as *cougar* (predatory older woman) or *toxic* (describing debts and assets). But once this technology is functioning well, it can combine with tools for automatic entry population

to enable dictionaries to update themselves with minimal human intervention. This is an explicit goal of the Wordnik project, which foresees a model where 'dictionary droids' define words automatically on the basis of contextual features.[17] The full package is still some way in the future, but many of its components already work well. And research efforts in automatic definition writing relate quite closely to Web technologies in areas such as the automated production of written texts like financial reports or summaries of sports events.[18]

5.5. *Theoretical implications*

In a detailed paper in 2003, de Schryver described the state-of-the-art in electronic lexicography, and outlined a series of 'lexicographers' dreams' – ideas then circulating about the features, content, and functionality that electronic dictionaries might incorporate in the future. Technology has moved on: few would now predict a future for CD-ROM dictionaries, for example, and even handhelds may not last much longer. Nevertheless, de Schryver's analysis provides a valuable benchmark. Some 'dreams' are already reality, others are in development, while others may still be some way off. But as de Schryver foresaw, the risk is that dictionary publishers might – in random fashion – do whatever the technology allows, so that the electronic dictionary 'will simply be a jamboree of all these dreams' (de Schryver 2003: 188).[19] The risk is real because we are in new territory, and the current situation is unstable: many older lexicographic norms have been ditched, but without – as yet – anything robust replacing them. This raises the question of whether the new situation requires new 'theories'. My provisional answer would be no. The basic principle of identifying target users, and starting from an assessment of their needs and capabilities, is a powerful guide to dictionary development, and as relevant now as it was in Cawdrey's day. At the same time, there are linguistic theories which can help us develop policies appropriate to the new medium. Thus if inclusion criteria become less stringent (but at the same time we don't want to abandon them altogether), Hanks' concept of 'norms and exploitations' provides a theoretical perspective which could underpin a new approach. Similarly, we now have the possibility of handling word-sense disambiguation in a way that more faithfully reflects what corpus data tells us. Linguistic theory, especially prototype theory, may help us exploit these opportunities. Lexicographers and linguists have long recognized the shortcomings of conventional (mostly linear)

presentations of the meanings of polysemous words. As far back as 1990, Geeraerts foresaw that 'computerized dictionaries open up new perspectives for dealing with prototypically clustered polysemy' (Geeraerts 1990: 200).

The current situation is messy, with a great deal of interesting but uncoordinated activity, and plenty of trial and error. For the moment, there are a lot more questions than answers. And the situation continues to change rapidly, as technologies from the wider field of Internet search increasingly impact on what we do. Thus there are new areas whose theories and applications we need to pay attention to. But the model which has served us so far still looks serviceable: the basic principles of focussing on the user and being faithful to the language data; seeking guidance from relevant linguistic and computational theory; and drawing on good-quality user research to identify what works. It is not clear that there is a role for 'lexicographic theory' as such – especially perhaps because such theories presuppose a world in which lexicographers and publishers are in control. But we no longer have control over the behaviour of users seeking information on the Web or wishing to make their own contribution to it.

6. Conclusions

This is a big subject, and limitations of space, time and above all my own knowledge preclude a more comprehensive survey of all the theoretical ideas (linguistic, computational, or metalexicographic) which might have applications in practical lexicography. There are huge areas which have not been discussed at all, such as theoretical work on corpus design, research in second-language acquisition and learner corpora, or developments in translation technology. For this reason alone, any conclusions can only be tentative.

In the end, what is the goal of lexicography? We want to produce better reference resources (whether or not these are accessed through what we would recognize as a dictionary), and so our requirement is for 'a framework for analysis and description that will do least distortion to evidence and be most helpful to the target audience' (Hanks 2008: 221–2). If we can discover the 'underlying regularities' which Zgusta referred to, we will be better placed to achieve this, and a further advantage of discovering 'systems' is that it facilitates the automation of lexicographic tasks. This in turn has multiple benefits: saving

lexicographers' time and reducing editorial costs, of course, but also enabling us to provide more complete and systematic lexicons.

Given these goals, lexicographers recognize that they need all the help they can get. Consequently, most people working in dictionaries are very receptive to theoretical ideas – from whatever quarter – which might help them to make their decisions and judgment calls with greater confidence. Discussing his proposed syllabus for an academic programme in lexicography, Sinclair (1984: 6–7) observed that 'there is substantial input from IT and linguistics, because I believe that the relevant theory is to be found in these areas'. Much of the discussion above bears this out. Admittedly, as Swanepoel points out, there is an element of eclecticism in the way that practitioners, like magpies, pick and choose from linguistic theory and adapt ideas for their own purposes. However, although the principles which lexicographers choose to guide them 'may be eclectically constituted, i.e. they may not form a systematic or strictly coherent body of hypotheses on lexical semantic matters … that does not diminish their status as guiding principles' (Swanepoel 1994: 13). Conversely, the widespread lack of enthusiasm among lexicographers for 'lexicographic theory' is not due to ignorance, laziness, or an antitheoretical mindset. It is simply a utilitarian impulse to not waste time on what looks unlikely to be useful. As Béjoint notes, most of the lexicographic theories which have been proposed 'have not been found convincing by the [lexicographic] community, and for good reasons' (2010: 381).

A.S. Hornby's first job, as a young graduate in Japan in 1924, was teaching English literature. He quickly concluded that what his students really needed was better language teaching, and within two years he had joined Harold Palmer's Institute for Research in English Teaching (IRET). IRET's remit went beyond the improvement of teaching methods and teacher training, to include 'research and experiment in linguistics' (Cowie 1999: 5), and its impressive theoretical output included work on syntax, 'vocabulary control', and most notably collocation. Hornby's teaching experience gave him a keen appreciation of what language-learners needed in a dictionary, and why the dictionaries of the time were not fit for purpose. Thus when he became a lexicographer in the early 1930s, his work was informed both by his teaching background and by his research in linguistics. Even without access to corpus data, Hornby (like Palmer) grasped the central importance of phraseology and recurrence in language, as is evident in the design (and title) of his ground-breaking dictionary. So there is every reason to believe that, if Hornby were

working in our field today, he would be a linguistically-aware corpus lexicographer

Notes

[1] Thanks to all those with whom I have discussed the issues addressed in this paper, especially Adam Kilgarriff, Piet Swanepoel, Robert Lew, Gilles-Maurice de Schryver, and above all Sue Atkins.

[2] cf. Tarp 2009.293: 'it is not difficult to reach the conclusion that the majority of the previous user research is in fact "a waste of time and money".'
[3] Thanks to Valerie Grundy for supplying a DTD.
[4] Like Urdang, I found myself baffled by most of the keywords listed in a recent Wiegand paper (Wiegand 2010), which include: HIERARCHICAL ARCHITECTONICALLY ENRICHED ARTICLE MICROSTRUCTURE, HIERARCHICAL HYBRID DEEP DOUBLE GLOSS-CONDITIONED ITEM STRUCTURE, and ELEMENT-HETEROGENEOUS STRUCTURE-CARRYING SET.
[5] Elsewhere, they trace the history of theoretical lexicography, and see its third (and by implication, culminating) stage as one guided by 'Bergenholtz and Tarp's functional approach, which is centred on … their interest in putting the dictionary user and the situation of use at the centre of the discussion' (Fuertes-Olivera and Bergenholtz 2011: 3).
[6] cf. van Sterkenberg (2003: 3) 'The prototypical dictionary is the alphabetical … general-purpose dictionary'.
[7] Similarly, the specialized DiCoInfo (a trilingual computer science dictionary) includes information about collocation, drawing explicitly 'on the Meaning-Text Theory approach to lexicography, specifically the *Dictionnaire explicatif et combinatoire* [of Igor Mel'čuk]' (L'Homme 2009: 5). For more on Mel'čuk, see next section: 4.6.2.
[8] On DELIS, see www.ims.uni-stuttgart.de/projekte/delis/.
[9] For example: Longman dictionaries used an advisory panel chaired by Randolph Quirk and including linguists such as John Lyons and Geoffrey Leech. Macmillan dictionaries has an advisory board whose members include Michael Hoey (chair), Hilary Nesi, Robert Lew and Adam Kilgarriff.
[10] Introducing CPA, Hanks explicitly acknowledges his debt to the work of three linguists: 'Pustejovsky on the Generative Lexicon …Sinclair's work on corpus analysis and collocations…and [Fillmore's] frame semantics' (nlp.fi.muni.cz/projects/cpa).
[11] The Style Guide for the DANTE project, for example, runs to well over 100 pages: www.webdante.com.
[12] Examples include the alternation between a container and its contents, a tree and its wood, and mass or unit instances of a drink. See Atkins and Rundell (2008: 139–141) for other examples.
[13] On a similar note: the DANTE project drew on data in Levin (1993) to improve its coverage of certain classes of verb, such as the 'spray/load' verbs like *dab*, *smear*, and *squirt* (ibid. 117–119). In the process, 'new' members of this class were identified: 'This is a good example of linguistic theory being applied to practical lexicography – with

benefits accruing, in return, to the theory, as new verbs belonging to this category are uncovered' (Rundell 2012: 27).

[14] http://oxforddictionaries.com/page/newwordinfographic/how-a-new-word-enters-an-oxford-dictionary.

[15] For a series of blog posts on 'new' inclusion principles, see http://www.macmillandictionaryblog.com/how-words-get-into-the-dictionary-part-1-the-past

[16] Thanks to Fadilah Jasmani (personal communication).

[17] http://illinois.edu/blog/view/25/65807?count=1&ACTION=DIA
LOG: this article (written in January 2012) describes these tools as if it they were already in place. This is not yet the case, but the goal is a realistic one.

[18] The *Atlantic* magazine (April 2012) reports on a company called Narrative Science, which has developed tools for automatic text production: http://www.theatlantic.com/entertainment/archive/2012/04/can-the-computers-at-narrative-science-replace-paid-writers/255631/

[19] cf. Atkins and Rundell 2008: 23: 'We need to be clear about the difference between doing things just because we *can*, and doing them because they will be of real value to the user'.

References

A. Dictionaries

Algemeen Nederlands Woordenboek. Leiden: INL. (ANW). http://anw.inl.nl/

Base lexicale du français. Leuven: Katholieke Universiteit Leuven. (BLF). http://ilt.kuleuven.be/blf/

Cawdrey, R. 1604. *A Table Alphabeticall of Hard Usual English Words*.

Corréard, M.-H. and V. Grundy 1994. *Oxford-Hachette English-French Dictionary*. Oxford: Oxford University Press.

DANTE: Database of Analysed Texts of English. http://www.webdante.com/

Den Danske Ordbog. Copenhagen: Det Danske Sprog- og Litteraturselskab. http://ordnet.dk/ddo/

Diccionario de Aprendizaje de Español como Lengua Extranjera. Barcelona: Pompeu Fabra University. (DAELE). http://www.iula.upf.edu/rec/daele/

Diccionario de la lengua española. Real Academia Española. http://buscon.rae.es/

Eijiro Pro on the Web. http://eowp.alc.co.jp/info2/

Rundell, M. (ed.) 2005. *Macmillan Phrasal Verbs Plus Dictionary*. Oxford: Macmillan.

Rundell, M. (ed.) 2007. *Macmillan English Dictionary for Advanced Learners*. (First edition 2002). Oxford: Macmillan. (MEDAL)

Rundell, M. (ed.) 2010. *Macmillan Collocations Dictionary*. Oxford: Macmillan.

Soanes, C. and A. Stevenson 2003. *Oxford Dictionary of English* (First edition 1998.) Oxford: Oxford University Press. (ODE)

Urban Dictionary. http://www.urbandictionary.com/

Wiktionary. http://en.wiktionary.org/

***Wordnik*.** http://www.wordnik.com/

B. Other literature

Apresjan, J. D. 1973. 'Regular Polysemy'. *Linguistics* 142: 5–39.

Atkins, B. T. S. 1993. 'Theoretical Lexicography and its Relation to Dictionary-Making.' *Dictionaries: Journal of The Dictionary Society of North America* 14: 4–43.

Atkins, B. T. S. 1987. 'Semantic-ID tags: corpus evidence for dictionary senses.' In *The Uses of Large Text Databases: Proceedings of the Third Annual Conference of the New OED Centre*. University of Waterloo, Canada, 17–36.

Atkins, B. T. S. and M. Rundell 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Atkins, B. T. S., M. Rundell and H. Sato 2003. 'The Contribution of FrameNet to Practical Lexicography.' *International Journal of Lexicography* 16.3: 333–358.

Béjoint, H. 2010. *The Lexicography of English*. Oxford: Oxford University Press.

Bergenholtz, H. 2011. 'Do we need another Theory of Lexicography?' *Lexicon* (Journal of the Iwasaki Linguistic Circle) 41: 1–10.

Bergenholtz, H. and I. Bergenholtz 2011. 'A Dictionary is a Tool, a Good Dictionary is a Monofunctional Tool.' In P. Fuertes-Olivera and H. Bergenholtz (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography.* London and New York: Continuum, 187–207.

Bergenholtz, H. and S. Tarp 2003. 'Two Opposing Theories: On H. E.Wiegand's Recent Discovery of Lexicographic Functions.' *Hermes, Journal of Linguistics* 31: 171–196.

Bogaards, P. 2010. 'Lexicography: Science without Theory?' In G.-M. de Schryver (ed.), *A Way with Words* (Festschrift for Patrick Hanks). Kampala, Uganda: Menha Publishers, 313–322.

Cowie, A. P. (ed.) 1987. *The Dictionary and the Language Learner*. Tübingen: Niemeyer.

Cowie, A. P. 1998. 'A. S. Hornby: a Centenary Tribute.' In T. Fontenelle et al. (eds.), *Euralex '98 : proceedings I-II : papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège*, 3–16.

Cowie, A. P. 1999. *English Dictionaries for Foreign Learners*. Oxford: Oxford University Press.

Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

Cruse, D. A. 2001. 'Review of H.E. Wiegand Semantics and Lexicography.' *International Journal of Lexicography* 14.2: 139–143.

Cruse, D. A. 2004. *Meaning in Language: An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.

de Schryver, G.-M. 2003. 'Lexicographers' Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography* 16.2: 143–199.

de Schryver, G.-M. and D. Joffe 2004. 'On How Electronic Dictionaries are Really Used.' In G. Williams and S. Vessier (eds.), *Proceedings of the eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Lorient: Université de Bretagne-Sud, 187–196.

Docherty, V. and U. Heid 1998. 'Computational Metalexicography in Practice: Corpus-based Support for the Revision of a Commercial Dictionary.' In T. Fontenelle et al. (eds.), *Euralex '98 : proceedings I-II : papers submitted to the Eighth EURALEX International Congress on Lexicography in Liège*, 333–346.

Farina, D. 1995. 'L. V. Shcherba's "Opyt": a Contribution to Theoretical Lexicography.' *International Journal of Lexicography* 8.4: 304–313.

Fuertes-Olivera, P. and H. Bergenholtz 2011. 'Introduction: The construction of Internet dictionaries.' In P. Fuertes-Olivera and H. Bergenholtz (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London and New York: Continuum, 1–16.

Geeraerts, D. 1990. 'The Lexicographical Treatment of Prototypical Polysemy'. In S. L. Tsohatzidis (ed.), *Meanings and Prototypes*. London: Routledge, 195–210.

Hanks, P.W. 1994. 'Linguistic Norms and Pragmatic Exploitations, Or Why Lexicographers Need Prototype Theory, and Vice Versa.' In F. Kiefer, G. Kiss, and J. Pajzs (eds.), *Papers in Computational Lexicography: Complex '94.* Hungarian Academy of Sciences, Budapest.

Hanks, P.W. 2002. 'Mapping Meaning onto Use.' In M.-H. Corréard (ed.), *Lexicography and Natural Language Processing*: *A Festschrift in Honour of B. T. S. Atkins.* Huddersfield: Euralex, 156–198.

Hanks, P.W. 2008. 'The Lexicographical Legacy of John Sinclair.' *International Journal of Lexicography* 21.3: 219–229.

Hanks, P.W. 2009. 'The Linguistic Double Helix: Norms and Exploitations.' In *After Half a Century of Slavonic Natural Language Processing* (Festschrift for Karel Pala). Brno, Czech Republic: Masaryk University, 63–80.

Hartmann, R.R.K. (ed.). 1984. *LEXeter '83 Proceedings*. Tübingen: Niemeyer.

Hartmann, R.R.K. 1987. 'Four Perspectives on Dictionary Use: a Critical Review of Research Methods.' In A. P. Cowie (ed.), *The Dictionary and the Language Learner*. Tübingen: Niemeyer, 11–28.

Hornby, A. S. 1965. 'Some problems of lexicography.' *English Language Teaching* XIX, April 1965, 104–110.

Kilgarriff, A. 1994. 'The Myth of Completeness and Some Problems with Consistency.' In W. Martin et al. (eds.), *Proceedings of the Sixth Euralex Congress*. Amsterdam, 101–116.

Kilgarriff, A. 2012. 'Review of Fuertes-Olivera and Bergenholtz, e-Lexicography: The Internet, Digital Initiatives and Lexicography.' *Kerneman Dictionary News* 20: 26–29.

Kosem, I. and R. Krishnamurthy 2007. 'A New Venture in Corpus-Based Lexicography: Towards a Dictionary of Academic English.' In *Proceedings of the 2007 Corpus Linguistics Conference*. Birmingham: University of Birmingham (available online).

Kwary, D. A. 2012. 'Adaptive Hypermedia and User-Oriented Data for Online Dictionaries : A Case Study on an English Dictionary of

Finance for Indonesian Students.' *International Journal of Lexicography* 25.1: 30–49.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Lew, R. 2010. 'Multimodal Lexicography: The Representation of Meaning in Electronic Dictionaries.' *Lexikos* 20: 290–306.

Lew, R. 2011. 'User studies: Opportunities and Limitations.' In K. Akasu and S. Uchida (eds.), *Asialex2011 Proceedings*. Kyoto: Asian Association for Lexicography, 7–16.

Lew, R. 2012 in press. 'How Can We Make Electronic Dictionaries More Effective?' In S. Granger and M. Paquot (eds.), *Electronic Lexicography*. Oxford: Oxford University Press.

L'Homme, M.-C. 2009. 'DiCoInfo Manual' http://olst.ling.umontreal.ca/dicoinfo/DiCoInfo%20manual_Eng.pdf

L'Homme, M.-C., B. Robichaud and P. Leroyer 2012. 'Encoding Collocations in the DiCoInfo: From Formal to User-friendly Representations.' In S. Granger and M. Paquot (eds.), *Electronic Lexicography*. Oxford: Oxford University Press.

Mahecha, V. and J. DeCesaris 2011. 'Representing Nouns in the *Diccionario de aprendizaje del español como lengua extranjera* (DAELE)'. In I. Kosem and K. Kosem (eds.), *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011*. Ljublana: Trojína, 180–186.

McCawley, J. 1986. 'What Linguists might Contribute to Dictionary Making if they could get their act together.' In P. Bjarkman and V. Raskin (eds.), *The Real-World Linguist: Linguistic Applications in the 1980s*. Norwood, New Jersey: Ablex, 3–18.

Mel'čuk, I. 1988. 'Semantic Description of Lexical Units in an Explanatory Combinatorial Dictionary: Basic Principles and Heuristic Criteria.' *International Journal of Lexicography* 1.3: 165–188.

Meyer, C. M. and I. Gurevych 2012 in press. 'Wiktionary: a New Rival for Expert-built Lexicons? Exploring the Possibilities of Collaborative Lexicography.' In S. Granger and M. Paquot (eds.), *Electronic Lexicography*. Oxford: Oxford University Press.

Moon, R. E. 2004. 'On Specifying Metaphor: an Idea and its Implementation.' *International Journal of Lexicography* 17.2: 195–222.

Prinsloo, D. J., U. Heid, T. Bothma and G. Faaß 2011. 'Interactive, Dynamic Electronic Dictionaries for Text Production.' In I. Kosem and K. Kosem (eds.), *Electronic Lexicography in the 21st Century: New Applications for New Users. Proceedings of eLex 2011*. Ljublana: Trojína, 215–220.

Rundell, M. 2006. 'More than One Way to Skin a Cat: Why Full-Sentence Definitions have not been Universally Adopted.' In E. Corino, C. Marello and C. Onesti (eds.), *Atti del XII Congresso Internazionale di Lessicografia : Torino, 6-9 settembre 2006*. Alessandria: Edizioni dell'Orso, 323–337.

Rundell, M. 2012 in press. 'The Road to Automated Lexicography: An Editor's Viewpoint.' In S. Granger and M. Paquot (eds.), *Electronic Lexicography*. Oxford: Oxford University Press, 13–30.

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

Samaniego Fernández, E. and B. Pérez Cabello de Alba 2011. 'Conclusions: Ten Key Issues in e-Lexicography for the Future.' In P. Fuertes-Olivera and H. Bergenholtz (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography.* London and New York: Continuum.

Shcherba, L. V. 1940/1995. 'Towards a general theory of lexicography.' *International Journal of Lexicography* 8.4: 314–350.

Sinclair, J. 1984. 'Lexicography as an academic subject.' In R.R.K. Hartmann (ed.), *LEXeter '83 Proceedings*. Tübingen: Niemeyer, 3–12.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Stock, P. 1984. 'Polysemy.' In R.K.K. Hartmann (ed.), *LEXeter '83 Proceedings*. Tübingen: Niemeyer, 131–140.

Swanepoel, P. 1994. 'Problems, Theories and Methodologies in Current Lexicographic Semantic Research.' In W. Martin et al. (eds.), *Proceedings of the Sixth Euralex Congress*. Amsterdam, 11-26.

Swanepoel, P. 2010. 'Improving the Functionality of Dictionary Definitions for Lexical Sets.' *Lexikos* 20: 425–449.

Tarp, S. 2008. *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Niemeyer.

Tarp, S. 2009. 'Reflections on Lexicographical User Research.' *Lexikos* 19: 275–296.

Tarp, S. 2010. 'Reflections on the Academic Status of Lexicography.' *Lexikos* 20: 450–465.

Tiberius, C. and J. Niestadt 2010. 'The ANW: an online Dutch Dictionary.' In A. Dykstra and T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6-10 July 2010*. Ljouwert: Fryske Akademy / Afuk, 747–753.

Tono, Y. 2010. 'A critical review of the Theory of Lexicographical Functions.' *Lexicon* (Journal of the Iwasaki Linguistic Circle) 40: 1–26.

Trap-Jensen, L. 2006. 'Making Dictionaries for Paper or Screen: Implications for Conceptual Design'. In E. Corino, C. Marello and C. Onesti (eds.), *Atti del XII Congresso Internazionale di Lessicografia : Torino, 6-9 settembre 2006*. Alessandria: Edizioni dell'Orso, 349–355.

Trap-Jensen, L. 2010. 'One, Two, Many: Customization and User Profiles in Internet Dictionaries.' In A. Dykstra and T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6-10 July 2010*. Ljouwert: Fryske Akademy / Afuk, 1133–1143.

Urdang, L. 2000. 'Review of R. R. K. Hartmann and G. James. Dictionary of Lexicography.' *International Journal of Lexicography* 13.1: 35–42.

van Sterkenberg, P. 2003. '"The" Dictionary: Definition and History.' In P. van Sterkenberg (ed.), *A Practical Guide to Lexicography*. Amsterdam: Benjamins, 3–17.

Wiegand, H.E. 1999. *Semantics and Lexicography: Selected Studies (1976-1996)*. Edited by Antje Immken and Werner Wolski. Tübingen: Niemeyer.

Wiegand, H.E. 2010. 'Hybride textuelle Strukturen und hybride textuelle Einheiten. Ein Beitrag zur Theorie der Wörterbuchform.' *Lexikos* 20: 495–539.

Wierzbicka, A. 1985. *Lexicography and Conceptual Analysis.* Ann Arbor: Karoma.

Wierzbicka, A. 1990. 'Prototypes save.' In S. L. Tsohatzidis (ed.), *Meanings and Prototypes*. London: Routledge, 347–367.

Zgusta, L. 1992. 'The Czech-Chinese Dictionary and the theory of lexicography.' *International Journal of Lexicography* 5.2: 85–128.