

LEXICOGRAPHY AND NATURAL LANGUAGE PROCESSING. A FESTSCHRIFT IN HONOUR OF B.T.S. ATKINS. Edited by MARIE-HÉLÈNE CORRÉARD. Euralex (<http://www.euralex.org/>), August 2002. (24 x 17 cm, viii + 248 pp., ISBN 2-9518583-0-2, € 30)

In this collection of articles an impressive line-up that includes some of the world's most respected lexicographers sings Sue's praises. Here Sue, of course, is B.T.S. ATKINS, who has been dominating the lexicographic landscape for several decades: as professional compiler in the 1960s, as general editor of the innovative COLLINS-ROBERT ENGLISH-FRENCH DICTIONARY in the 1970s, as co-architect of the COBUILD project in the 1980s, as originator of the idea of the British National Corpus (BNC) in the 1990s, as lexicographic advisor to the team that produced the first corpus-based bilingual dictionary, the OXFORD-HACHETTE FRENCH DICTIONARY, a decade ago, and up to this day also as co-driving force behind the FrameNet project. These are just a few of the highlights, but already it is clear from them that she has successfully built bridges over the years, bringing lexicographers, linguists and computationalists together. She further organised and taught at numerous professional and academic training courses and workshops in lexicography, unconditionally imparting her enthusiasm, expertise and knowledge. Over the

past twenty years, metalexigraphers and practical dictionary makers alike have also been able to savour her genius in around forty academic papers. Sue, in short, commands respect.

Sue's brother, JOHN MCH. SINCLAIR, recently remarked that "[t]here is only one book-length guide to making a dictionary from a corpus" (2003: 177), referring to his own edited collection *Looking Up* (SINCLAIR 1987). Aspiring corpus lexicographers have indeed not been well served and, in order to acquaint themselves with more recent practices, have had to make do with the odd article published here and there in the scientific literature. This Festschrift honouring Sue might very well, if not provide 'the' answer to John's call, go a long way in being at least 'the currently hottest collection' in corpus lexicography. Indeed, while the second half of the book, plus the first two articles, may be seen as revolving around the *Natural Language Processing* (NLP) part of the book's title, all but two of these actually belong to the sub-field of corpus lexicography. The remainder of the articles in the first part, conversely, firmly belong to the (more traditional) field of *Lexicography* tout court, the first part of the book's title. The book thus clearly addresses lexicographers, much more than it does computational linguists. Fifteen authors contributed a total of thirteen articles, all of them in English except for one that is couched in French. Both languages are often the objects of lexicographical description throughout the articles, however. The book opens with a reprint of Sue's keynote address pronounced at EURALEX 1996, while most of the other contributions have specifically been written for this Festschrift.¹ The book further also includes a short *Foreword* and *Tabula Gratulatoria* as front matter, and a list of Sue's *Publications* as back matter.

B.T.S. ATKINS's 'Bilingual Dictionaries: Past, Present and Future' remains as timeless as when read (and reread!) when it was first published in 1996 and doubtless belongs to the Canon of Lexicography.² The article consists of three parts, and in a way 'the best of ATKINS' is brought together. The highlight is the middle section, which describes the 'virtual dictionary' – Sue's vision of the electronic dictionary of the future. This section is preceded by an evaluation of the strengths and weaknesses of the traditional (paper) dictionary, and is followed by a brief overview of applying a frame semantics approach to corpus analysis. Although Sue's virtual dictionary is not yet on the market, several of her suggestions to combat the weaknesses of traditional dictionaries have already been implemented in current electronic dictionaries, while her visionary dictionary remains – to this date – one of the only ones that is rooted in a solid theoretical base, *in casu* frame semantics. This linguistic theory continues to attract a lot of attention and was even the topic of a recent special issue of the *International Journal of Lexicography* (FONTENELLE 2003), an issue to which also Sue contributed with brio.

KRISTA VARANTOLA's 'Use and Usability of Dictionaries: Common Sense and Context Sensibility?' is an excellent follow-up to Sue's text and somehow even interacts with it, as it develops some ideas further. Important questions are raised and provocative statements are made – both characteristic of VARANTOLA's oeuvre. The article is open-ended, however, as quite a number of questions remain unanswered while some statements remain undeveloped. The most original concept introduced here is the 'user-controlled search chain' whereby (electronic) dictionaries would stop attempting to give *the* solution but would rather contain keys *to* the solution and leave it to the user to browse through hypertext. That hypertext would in itself be part of an integrated network, one (set of) tool(s) among many others in a toolbox.

In the well written and well structured article 'La métalangue, un mal nécessaire du dictionnaire actif' ALAIN DUVAL revisits two lexicographic classics, viz. the dichotomy passive vs. active, and the (related) dichotomy unidirectionality vs. bidirectionality. Just a handful of examples chosen from three French-English dictionaries, published between the end of the nineteenth century and the 1930s, suffice to show how the (progressive) inclusion of metalanguage in the bilingual dictionary facilitated the arrival of

¹ At least two articles, viz. the ones by KILGARRIFF and TUGWELL, as well as FONTENELLE, are near-verbatim reprints, here of contributions to the 'Collocation' workshop proceedings at the 39th ACL meeting.

² Rather ironically, however, it was not included in the first Anthology of the field, being the three-volume set edited by HARTMANN (2003). But then, that Anthology seems to have overlooked the field of *Computational Lexicography* altogether (cf. DE SCHRYVER 2005).

the ‘active dictionary’. Whereas early bilingual dictionaries may be characterised by a ‘forced bidirectionality’, implying that each half of the dictionary is actually only useful to the mother-tongue speaker of either language, the modern move is indeed towards the ‘bidirectional bifunctional dictionary’, which, and linking it to Sue’s keynote, is analogously “highly redundant for any particular user” (p. 25). Unfortunately, DUVAL does not link his contribution to Sue’s virtual dictionary, as the latter indeed offers a way out of the impasse.

RICHARD WAKELY and HENRI BÉJOINT’s ‘Word Groups in Bilingual Dictionaries: OHFD and After’ is a straightforward account of how and why certain word groups (called ‘lexical sets’ in the article) were treated by means of usage notes in the OXFORD-HACHETTE FRENCH DICTIONARY (OHFD). The text provides an enjoyable glimpse of practical dictionary making, including layout issues that somehow ‘bend reality’, such as “[i]n some cases, we had to change the title of the note (choosing between, for example *disease*, *illness*, *health*, etc.) in order to make sure that the notes would be distributed evenly along the pages of the dictionary” (p. 64). The authors also indicate how work on the dictionary eventually led to *French Usage*, a new type of book on French language usage specifically aimed at speakers of English.

In ‘Examples and Collocations in the French “Dictionnaire de langue”’ A.P. COWIE begins by setting up a framework for types of illustrative examples. Although one is actually dealing with a continuum, the main types he recognises are (A) minimally adapted quotations, (B) decontextualised sentences (being either adapted quotations or invented examples), and (C) skeleton examples (known as ‘minimal syntagmatic units’, or collocations, in the French tradition, and the result of simplification and abstraction, with substitutables often listed). Instances of these various types are then discussed in four French *dictionnaires de langues* [lexical dictionaries], viz. in two one-volume and in two multi-volume dictionaries. The reasoning is unfortunately at times rather fuzzy, with a multitude of ‘little facts’ being added as if in concentric layers. Even though one will not dispute that, “[o]n the whole, they order this matter better in France” (p. 73), this contribution is, paradoxically, not well ordered.

JURI D. APRESJAN’s ‘Principles of Systematic Lexicography’ is a highly entertaining piece in which five principles of ‘systematic lexicography’ are defined and illustrated. The author maintains that these are of paramount importance to practical dictionary making. According to the first principle each language ‘forces’ its speakers to express specific meanings, such as in certain Russian sentences where one is forced to specify the manner of locomotion through verbs like *walking*, *flying* or *crawling*, as the use of the (more) general *leaving* results in doubtful constructions. The second principle insists on a perfect ‘coordination’ of dictionaries and grammars. The current discrepancy is convincingly exemplified with a discussion of the labelling of numerals as either ‘nouns’ or ‘adjectives’ in English dictionaries vs. their characterisation as ‘numerals’ in their own right in grammars. In the third principle it is advocated that ‘lexical classes’ ought to be treated in full and described uniformly in dictionaries. Under the heading ‘lexicographic types’ factive and putative predicates are looked into as an example, leading to highly interesting oppositions (in English) such as ‘knowledge has a source, but not a reason’ vs. ‘opinions have a reason, but never a source’. The converse verbs *buy*, *sell*, *pay* and *cost* are discussed as an instance of a ‘lexical-semantic paradigm’ under a second heading. Moving from the macro- to the microcosm, the fourth principle stresses the importance of an exhaustive linguistic description of lexemes, while the fifth points to the need to pay attention to meaning interaction across language units.

A good example of APRESJAN’s fourth principle can be found in CHARLES J. FILLMORE’s contribution, ‘Lexical Isolates’, these being “words that appear to be of unique semantic or syntactic type” (p. 108). The outliers FILLMORE presents are *let alone*, *mention*, *else*, *wrong* and *ilks*. The analysis of existing dictionary treatments of these items is rather critical indeed, yet FILLMORE’s own relation with corpus data isn’t clear-cut either. For instance: “The value of using a corpus is that it can show that one’s intuitions about word use can be incorrect” (p. 119) vs. “The BNC contains in very small numbers both ‘whose else’ and ‘whose else’s’. My inclination is to declare them simply wrong, in the sense that their creators would have rejected them on a moment’s reflection. I should probably know better by now, but old habits tend to hang on” (p. 117–118). One would also have wished to see how the various analyses impact on (interact with?, supplement?, modify?) the FrameNet data, or, in simple terms, how are exceptions and idiosyncrasies *framed*?

In ‘Sketching Words’ ADAM KILGARRIFF and DAVID TUGWELL introduce the word sketch – an automatically produced summary (sketch) of a word’s grammatical and collocational behaviour. The word sketch is arguably one of the most revolutionary e-tools that have been added to the lexicographer’s workbench over the past few years. Computationally, one ‘only’ needs a large corpus of a particular language, as well as a part-of-speech tagger, a lemmatiser and a grammar. At the heart of the word-sketch engine lies a database of all grammatical relations in the corpus, stored as quintuples of the form {Rel, Word1, Word2, Prep, Pos}. The authors indicate how, with a grammar of just twenty-six relations (Rel), around 70 million quintuples may be derived from the 100 million BNC. Word1 is never null, as it is the lemma of the word for which Rel holds. Word2 (the lemma of the other open-class word involved) and/or Prep (the preposition or particle involved) can be null. The last slot in the quintuple is the position number of Word1 in the corpus, which allows for the display of the actual occurrences in context for the lexicographers. This ability to be able to more easily access *relevant* examples turned out to be but one of the boons during the production of the recent MACMILLAN ENGLISH DICTIONARY, the first dictionary project that effectively used word sketches during compilation. While the professional lexicographers found the word sketches useful (obviously), the sketches also sped up compilation time, and now strengthen the resulting dictionary’s claims to completeness. Last but not least, apart from its potential in dictionary making, word sketch analyses have proven to be useful for the development of NLP systems as well, in particular for word sense disambiguation.

It is fitting to have MICHAEL RUNDELL, the editor-in-chief of the MACMILLAN ENGLISH DICTIONARY, have the next word with ‘Good Old-fashioned Lexicography: Human Judgment and the Limits of Automation’, in which he confirms that “for most editors the Word Sketches came to be the preferred starting point for looking at a word” (p. 141). In a way this leads to the main question around which this article is built: “can we now foresee a time when human beings will play only a subordinate, organizing role in the process of producing descriptions of languages?” (p. 138). As it turns out, RUNDELL is of the opinion that this idea is “unlikely and possibly misguided” (p. 152). The descriptions of two exciting innovations in the MACMILLAN ENGLISH DICTIONARY are of particular interest, as they suggest that linguistic findings, corpus analyses and dictionary making can successfully go hand in hand. The first of these is the inclusion of usage notes showing how metaphors are typically encoded – a first in lexicography,³ the second a novel article structure reflecting the view that words have *some* clearly distinct meanings only, and then much fuzzier meaning-clusters.

Meanings, too, and how these relate to use, are the topic of PATRICK HANKS’s contribution ‘Mapping Meaning onto Use’. As an ode to Sue, and following a review of FILLMORE and ATKINS’s discussion of the semantic frame of *risk*, HANKS presents his own corpus analyses of *lean* and *tank* for lexicographical purposes. Rare are such detailed accounts in which the reader is led by the hand and allowed to see how the master cuts his way through the corpus vines. The latter, including their analyses, are displayed in full as addenda, hereby allowing the reader to appreciate the hesitations – about which HANKS is quite open – even more. Once the path has been cut, once HANKS unspun the hanks, the reader is offered the view that syntagmatics in tandem with ‘perceived meaning’ ought to be the organising principle of dictionary entries for verbs and adjectives. The organisation for nouns is similar, but slightly more complicated.

HANKS’s view that lexicographers should primarily focus on a word’s *usage* takes on a new dimension in GREGORY GREFENSTETTE’s ‘The WWW as a Resource for Lexicography’, in which it is optimistically claimed that, starting from Web material, “we will be able to automatically build large, organized collections of word usages for specific languages and for specific domains” (p. 214). The author indeed seems to place more trust in computed analyses than in (good old-fashioned) language artisans. Time will have to tell which of the two will become most prominent, but it is indeed already the case that the Web is used both *for* corpus creation as well as simply *as* a corpus, and not only for English at that (cf. e.g. DE SCHRYVER 2002). Given that GREFENSTETTE convincingly demonstrates how one can

³ A full-blown discussion of the ‘series of forty special features on metaphor’ in the MACMILLAN ENGLISH DICTIONARY can be found in MOON (2004), who developed RUNDELL’s idea and constructed the entries.

obtain impressive results even with simple, approximate linguistic tools, the contents of this contribution, then, are of particular interest to lexicographers (and linguists at large) working in so-called exotic languages.

In ‘Lexical Knowledge and Natural Language Processing’ THIERRY FONTENELLE succinctly revisits various aspects of earlier work on Mel’čukian Lexical Functions and goes on to reach out to Frame Semantics. Lexical Functions, indeed, seem to have their limitations when the sole populator of ‘truly intelligent’ NLP systems. The type of lexical database that FONTENELLE advocates for this purpose combines traditional dictionary data and thesauri, as well as collocational and lexical-semantic networks. This article is the only one in the collection that focuses primarily on NLP, rather than on lexicography.

In ANNIE ZAENEN’s monumental capstone, ‘Musings about the Impossible Electronic Dictionary’, the major lexical theories of the past ten to fifteen years are reviewed. PUSTEJOVSKY’s Generative Lexicon, FILLMORE and ATKINS’s Frame Semantics, MILLER and FELLBAUM’s WordNet, MEL’ČUK’s Meaning-Text Theory, as well as statistical models, are all briefly evaluated – subjectively, yes, but that is exactly the strong point of this article. The polytheoretical and multifunctional lexicon for both human consumption and NLP envisaged over fifteen years ago, has not been realised – and (probably) never will. Instead, ZAENEN predicts three types of enterprises in lexicography and NLP: small lexical databases structured along the lines of this or that lexical theory, bigger lexica based on statistical methods, and, yes, the traditional dictionary after all.

By way of general assessment one can thus confidently state that the impressive line-up of scholars alluded to at the start of this review did not disappoint. The editor furthermore also managed to order the contributions in a most logical way – the text flows gently from cover to cover. What the editor should also have done, however, was to enlist a proof-reader – the typographical errors are just too numerous, and Sue deserved better. This is especially so for the reprint of Sue’s keynote, as more new errors were introduced than old ones corrected compared to the original, with one section (under §1.1) even missing, and especially Figure 2 badly redrawn. Sue also deserved to be praised by many more of her numerous admirers from around the world (e.g. ULRICH HEID, CARLA MARELLO, D.J. PRINSLOO, etc.) and they should have been invited. Lastly, at the first LEXICOM training workshop in lexicography and lexical computing, 16–20 July 2001, Sue contended that there are only three linguists to whom God speaks directly, viz. JURI D. APRESJAN, IGOR A. MEL’ČUK and ANNA WIERZBICKA. Reading through this Festschrift, too, and no wonder, is celestial.

Literature

Dictionaries

- COLLINS-ROBERT ENGLISH-FRENCH DICTIONARY = ATKINS, B.T.S. and PAUL ROBERT eds. : 1978. Collins-Robert French-English, English-French Dictionary / Robert-Collins dictionnaire français-anglais, anglais français. Glasgow: Collins and Paris: Le Robert.
- MACMILLAN ENGLISH DICTIONARY = RUNDELL, MICHAEL ed. : 2003. Macmillan English Dictionary for Advanced Learners. Oxford: Bloomsbury Publishing Plc.
- OXFORD-HACHETTE FRENCH DICTIONARY = CORRÉARD, MARIE-HÉLÈNE and VALERIE GRUNDY eds. : 1994. The Oxford-Hachette French Dictionary: French-English, English-French. Oxford: Oxford University Press and Paris: Hachette Livre.

Secondary Literature

- DE SCHRYVER 2002 = DE SCHRYVER, GILLES-MAURICE : 2002. “Web for/as Corpus: A Perspective for the African Languages.” In: *Nordic Journal of African Studies* 11/2, 266–282. Online: <<http://tshwanedje.com/publications/>>

- DE SCHRYVER 2005 = DE SCHRYVER, GILLES-MAURICE : 2005. "Book Review: R.R.K. Hartmann, ed. 2003. *Lexicography: Critical Concepts*." In: *International Journal of Lexicography* 18/1, 93–102. Online: <<http://tshwanedje.com/publications/>>
- FONTENELLE 2003 = FONTENELLE, THIERRY ed. : 2003. *FrameNet and Frame Semantics*. In: *International Journal of Lexicography* 16/3, 231–366.
- HARTMANN 2003 = HARTMANN, REINHARD R.K. ed. : 2003. *Lexicography: Critical Concepts*. London: Routledge.
- MOON 2004 = MOON, ROSAMUND : 2004. "On Specifying Metaphor: An Idea and its Implementation." In: *International Journal of Lexicography* 17/2, 195–222.
- SINCLAIR 1987 = SINCLAIR, JOHN MCH. ed. : 1987. *Looking Up. An account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary*. London: Collins ELT.
- SINCLAIR 2003 = SINCLAIR, JOHN MCH. : 2003. "Corpora for lexicography." In: PIET VAN STERKENBURG (ed.) : *A Practical Guide to Lexicography. Terminology and Lexicography Research and Practice 6*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 167–178.

Gilles-Maurice de Schryver