John Benjamins Publishing Company

Schryver, Gilles-Maurice de (ed.). 2010. *A Way with Words: Recent Advances in Lexical Theory and Analysis, A Festschrift for Patrick Hanks*. Kampala (Uganda): Menha Publishers. ISBN 978-9970-101-01-6.

Reviewed by Henri Béjoint

Patrick Hanks has become one of the leading figures of lexical semantics, standing out among his colleagues as someone who draws all his evidence from corpora, who focuses on the more frequent and the more 'normal' features of discourse, and also someone who has compiled dictionaries of all kinds and thus has been confronted with the whole range of problems of the description of lexical items. The list of authors in this Festschrift is impressive. There are twenty articles in all, by twenty-six authors, ranging from seven to thirty-one pages, and arranged — after the introductory essay by Gilles-Maurice de Schryver — in three parts corresponding to Patrick's main three domains of research: 'Theoretical aspects and background' (five articles), 'Computing lexical relations' (seven articles) and 'Lexical analysis and dictionary writing' (seven articles). Unfortunately, the book has no index. The introductory essay (fifteen pages), half biography and half summary of Patrick's contribution to the field, is followed by a sixteen-page list of Patrick's main publications from 1967 to 2010, from a manual of business English to various dictionaries of surnames and given names, from translations from German and Latin to the edition of the *New Oxford Dictionary of English* via the monumental, six-volume *Lexicology: Critical Concepts in Linguistics* published by Routledge in 2008.

The first four contributions come from linguists who share with Patrick a conviction that the description of a lexical item must include a description of the other lexical items that occur with reasonable frequency and demonstrable significance in its more or less immediate environment. The first is by John Sinclair, who worked with Patrick in Birmingham at the compilation of the *Collins COBUILD Dictionary* (1987), and whose views exerted a profound influence on the shaping of Patrick's theory of language. In 'Defining the definiendum' (eleven pages), Sinclair says, *defining* must be understood as 'fix the limits of' rather than as 'give a definition for', as the article is about the identification of the true building blocks of language — at least in performance, if not in competence —, those strings that are found with significant frequency in discourse, ranging in length and complexity from compounds, now relatively well known and documented, to larger, ill-defined 'units' formed by a word with its collocators and colligators, those units — 'configurations

of text', Sinclair says (38) — that should be the object of the metalanguage of lexicography, even though lexicographers have traditionally tended to focus on single words, because it is easier and because the user will also naturally tend to use words to access the information in the dictionary. 'The word, Sinclair writes, is not the principal unit of meaning in a language' (37). He takes the example of the verb *sever* in his corpus and applies the concepts that he has invented or developed to describe some of the larger units that *sever* enters into: not only 'collocation' and 'colligation' but also 'core sense', 'semantic prosody', 'local grammar', etc. This may not be the best of John Sinclair's publications, as the development is too allusive to convince anyone who has not already been convinced, but it is definitely the most moving, because it is the last text on which he worked. It is given to us as he left it, unfinished, with a few minor adjustments by Rosamund Moon, as we are told on page 17, or perhaps with no adjustments at all, as we are told on page 45.

The second paper, 'Very large lexical entries and the boundary between linguistic and knowledge structures' (seventeen pages), is by Yorick Wilks, who was for a few months Patrick's doctoral supervisor at the university of Essex. It was originally written in the 1970s but is given here, with slight modifications, because it has not been easily available. It addresses one of the most fundamental problems of semantics, how the meaning of a lexical item can be extended and what are the limits to these semantic extensions. Wilks sums up his theory of Preference Semantics — probably too briefly for the non-specialist — and applies it to the discussion of cases such as *My car drinks gasoline*, *a toy lion*, etc. His main point, perhaps, is that such meaning extensions are the 'norm' rather than the exception in language, and that they cannot be relegated to the realm of performance, a position that is very close to Patrick's theory of Norms and Exploitations.

James Pustejovsky's work on the Generative Lexicon has been generally ignored by lexicographers, unlike Fillmore's or even Mel'čuk's, but he has recently been introduced to the world of dictionary making through his collaboration with Patrick on the construction of the Pattern Dictionary of English Verbs (PDEV). In 'Mechanisms of sense extension in verbs' (twenty-two pages), he revisits the question of the determination of the number of meanings of polysemous words and how they can be identified. He takes the example of verbs of motion like *arrive* or *crawl* — the latter described by Fillmore and Atkins in terms of Frame Semantics — and shows that within the broad categories of conceptual metaphors described by cognitive linguists in general and Lakoff in particular, such as 'a conversation is like a journey', there are gradations, particularly in metaphoricity, according to the nature and number of meaning elements, or arguments of the 'original' sense that are preserved or modified. A lexicographer can conclude, as Patrick has repeatedly said, that the binary system of traditional dictionaries, same meaning *or* different meaning, is an inadequate representation of how language really works.

Igor Mel'čuk never worked with Patrick and differs from him in that he has always maintained that corpora cannot be expected to provide interesting data for the linguist. But he shares with him the conviction that the description of a lexical item implies the description of the other lexical items that are associated with it in discourse. His paper, 'The government pattern in the Explanatory Combinatorial Dictionary' (thirty pages), concentrates on the key lexicographical notion of 'government pattern', which can be used to distinguish the different uses of a form and to divide polysemous articles. He takes examples from English, Russian, French, Spanish and even Lezgian — how many languages does he master? Mel'čuk's text is of course highly Mel'čukian, formalized and requiring unfailing attention, but impressive in the logic of its demonstration.

David Wiggins met Patrick after Patrick wrote an article in the *Times Literary Supplement* that was a response to one of David's publications. Patrick's main point, apparently, was that language is a social 'thing'. In 'The paradox of analysis and the paradox of synonymy' (fourteen pages), Wiggins revisits the paradox of analysis: if the analysis means the same as the word being analysed it does not analyse anything. He shows that if one uses Frege's distinction between *Sinn* and *Bedeutung* the paradox can be shown not to be a paradox after all. For good measure he adds what he calls the paradox of synonymy: if two words have the same meaning, like *furze* and *gorse*, then saying *furze is gorse* is not more than saying *furze is furze* or *gorse is gorse*, that is saying nothing. Intriguing indeed, though one cannot help feeling that the two paradoxes cannot be allowed to stand in the way of the lexicographer writing definitions, in comprehension or by synonymy. The point might be that *is* in such sentences does not establish any identity of meaning but just says that the same concept can be designated by two words: *furze is gorse* says something like *The thing called furze can also be called gorse*, and the existence of two words for the same concept may be no more paradoxical than the fact that the French call *cheval* what the English call *horse*.

The second part opens with Kenneth Church's paper, 'More is more' (seven pages), in which he wonders, once again, whether a corpus must be balanced and whether size can assure balance, answering one of Adam Kilgarriff's (2007) recent publications on the subject. The answer is that accessibility must always be judged against size, and that all depends what one wants to do with the corpus. Kenneth Church evokes his work with Patrick, particularly for the preparation of the famous paper, 'Word association norms, mutual information, and lexicography' (1989). He notes that Patrick was interested in describing normal language rather than the beautiful language of literature: 'Many lexicographers study butterflies, but Patrick likes "central and typical" moths' (136). The title of Gregory Grefenstette's paper, 'Estimating the number of concepts' (fourteen pages), may surprise the reader. Grefenstette argues that estimating how many concepts are commonly used in a

language would be useful for the lexicographers of the future, and that it is possible to have a rough idea. The answer? About 200 million. A rough idea, admittedly, if one looks at the list of improvements to be considered in the mode of calculation, but an idea all the same. In 'Identifying adjectives that predict noun classes' (twelve pages), David Guthrie and Louise Guthrie investigate how far adjectives can be used to predict the semantic category of the nouns they qualify. The answer is that they 'do indeed provide a great deal of information about the semantic class of the nouns they modify' (166). Alexander Geyken's paper, 'Statistical variations of German support verb constructions in very large corpora' (seventeen pages), provides an answer to Kenneth Church's question: almost all the German verb constructions listed in the *Wörterbuch der deutschen Gegenwartssprache*, a large monolingual dictionary of German, are statistically salient in a 1-billion-token corpus but not in a 100-million-token one. The smaller corpus, however, has them all, though with frequencies that are too low to be significant. This reminds the author of Patrick's claim that a 100-million word corpus is enough for most usages except the very rare (181). 'A case study in word sketches — Czech verb *vidĕt* "see"' (twelve pages), by Karel Pala and Pavel Rychlý, lists a number of errors in what the Sketch Engine returns in their ongoing study of Czech. Most errors, they conclude, result from errors in POS-tagging, disambiguation of case homonymy, and incorrect lemmatisation (196).

The next two articles are about Patrick's PDEV, that currently contains 678 'completed verb entries with patterns and the corresponding reference sample data' (202). In 'The lexical population of semantic types in Hanks's PDEV' (sixteen pages), Silvie Cinková, Martin Holub and Lenka Smejkalová investigate the semantic type labels given to noun collocates of verbs and try to determine how far they are consistent and reproducible. In 'From pattern dictionary to pattern-bank' (twenty-five pages), Elisabetta Jezek and Francesca Frontini wonder how the information contained in the PDEV can best be classified so as to be retrievable. They suggest the creation of a bank that would complement the dictionary.

Part III opens with Rosamund Moon's 'Words that spring to mind: idiom, allusion, and convention' (twenty-four pages). Moon, who worked with Patrick both at Collins for the preparation of the COBUILD and later at Oxford University Press, examines the occurrences of *spring to mind* in the 448-million-word Bank of English. Her observations go far beyond what any lexicographer can afford to consign to a dictionary, but they 'chime with points made by Patrick Hanks' (263), that words do not always mean what dictionaries say they mean, but that they are 'part of the way in which a language is being used'. Sue Atkins, in 'The DANTE database: Contribution to English lexical research, and in particular to complementing the FrameNet data' (thirty-one pages), explores the feasibility of a mapping between FrameNet and DANTE (for Database of Analysed Texts of English),

a lexical database compiled for a forthcoming English — Irish dictionary. DANTE draws on a corpus of 1.7 billion words explored by the Sketch Engine and has now about 42,000 lexical entries covering more than 92,000 lemmas (268). Unfortunately the plates illustrating the article are small and not sharp enough to be easily readable. Atkins notes that the two databases have much in common, particularly the fact that they list lexical units, LUs, single words or multiword lemmas in one of their senses. There are many difficulties — one being the fact that the two databases do not always distinguish the same LUs, as we can see for *observe* — but Atkins is confident that, with the help of 'geeks' (294), the two databases will one day communicate and enrich each other.

Adam Kilgarriff and Pavel Rychlý, in 'Semi-automatic dictionary drafting' (fourteen pages), explore a way in which the Sketch Engine could be used to identify the different meanings of polysemous words by grouping together their occurrences in the corpus according to the semantic categories to which the collocates (in the broad sense) belong. This, the authors say, is still highly experimental, but could be made to work. In 'Science without theory?' (ten pages), Paul Bogaards wonders whether a theory of lexicography exists and whether it is even possible. His answer is that there might be a number of theories that could apply to some aspects of dictionary-making, drawing on linguistics for the assessment of lexical data, on psychology and psycholinguistics for an evaluation of the needs and skills of the users, on information theory for the way the information is presented to the user in the dictionary, etc., but that a unified theory of lexicography is difficult to imagine. Can one say that lexicography is a science? Bogaards notes, interestingly, that the lack of a theory does not seem to have prevented lexicography from making progress over the centuries, and particularly since the late 20th century. 'The Polish COBUILD and its influence on Polish lexicography' (ten pages), by Mirosław Bańko, explains how the COBUILD model was used to compile a dictionary of Polish, the *Inny slownik języka polskiego*, published in 2000. This was not a commercial success, but it was then imitated by other dictionaries of Polish for native speakers, particularly in the use of full-sentence definitions — a fine example of the influence of one national tradition on another, and also of how innovations can turn out to be important in the long run, however disturbing for the dictionary users when first introduced. In 'ARGOT: The flesh made word' (sixteen pages), Jonathon Green discusses the French *argot*. First the name of a group of criminals, it was then used to refer to their language, that was collected in a number of records and eventually dictionaries, in a tradition that is different from the tradition of cant dictionaries in Britain. The last article is by Michael Rundell, 'Defining elegance' (twenty-seven pages), and it does just that: it tries to define the notion of elegance in writing in general and in writing dictionaries in particular. A vague notion, admittedly, but it seems to have something to do with simplicity,

brevity, and ingenuity. In lexicography it applies to the choice of examples, the determination and ordering of the senses of polysemous words, the writing of definitions, and the design of the dictionary as a whole. Elegance has always been important in lexicography, and will remain important in electronic dictionaries, where lexicographers will have to choose what is relevant and what is not. Rundell's paper is, of course, elegantly written, as usual. It ends with a plea for flexibility, for the creative part of the lexicographer's work.

The book as a whole adequately reflects Patrick Hanks's work and highlights what the community owes him, the full-sentence definitions of COBUILD, the new representation of polysemy in NODE, the idea that meanings are events rather than entities, that what counts is not the meaning but the meaning potential, the concept of Mutual Information as a way of exploring collocations, the idea that lexicographers must look for the more frequent patterns and the notion that differences in meanings are always associated, at one point or another, with differences in pattern. Patrick Hanks is, as Sue Atkins writes (267), 'someone whose contribution to the field of corpus-based lexical analysis is second to none', as this volume clearly demonstrates.

References

Church, K. W. and P. Hanks. 1989. "Word Association Norms, Mutual Information, and Lexicography." In *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, 26–29 June 1989*. 76–83. University of British Columbia, Vancouver.
Kilgarriff, A. 2007. "Googleology is Bad Science." *Computational Linguistics* 33(1), 147–151.

Reviewer's address

Henri Béjoint
CRTT (Centre de Recherche en Terminologie et Traduction)
86, rue Pasteur
Université Lyon 2
69007 Lyon
France

henri.bejoint@univ-lyon2.fr