Book Reviews

Lexicography and Natural Language Processing:
A Festschrift in Honour of B. T. S. Atkins

Marie-Hélène Corréard (editor)

Reviewed by
Woody Haynes and Martha Evens
Illinois Institute of Technology

This volume is a festschrift for Sue (B. T. S.) Atkins, who is perhaps best known for her work as general editor of the *Collins-Robert English/French Dictionaries* and a consultant-advisor for the *Oxford-Hachette English/French Dictionary*, which led to a revolution in the construction of bilingual dictionaries. She has also done a great deal to bridge the gap between professional lexicography, academic linguistics, and computational linguistics. Most recently she has been working with Charles Fillmore to build FrameNet and to adapt his ideas for use in a dictionary framework. The lead-off paper in this volume is Atkins's keynote address from the 1996 EURALEX meeting, the inspiration for this volume. The contributions of the other authors, all of them new papers, examine how lexicography is responding to Atkins's call for a "radical new type of dictionary" in her 1996 address. They consider the current state of dictionaries, discuss their strengths and weaknesses, and describe new computational tools that are facilitating exploration in new directions and providing new insights.

The increasing availability of diverse electronic text has had a profound effect on the process of creating dictionaries, both in the compilation process and in the depth and structure of the result. It is true that most of the papers in this volume talk about how to use natural language processing in lexicography rather than how to make use of various lexical resources in natural language processing. But we believe that the volume includes much useful material for all those whose work in computational linguistics makes them consumers of lexical resources, since it discusses many of the current ideas about how those resources should be constructed. Even those papers that focus on bilingual resources provide many interesting ideas about the practice of lexicography today. Most of the recommendations for bilingual dictionaries can equally address issues involved in the use of machine-readable dictionaries of all kinds and apply to deficiencies seen in all available lexicons.

Atkins's keynote address, entitled "Bilingual Dictionaries: Past, Present and Future," looks at the various types of information available and needed in various types of bilingual and monolingual dictionaries, categorizes it, and argues for a truly electronic dictionary that can adapt itself to the needs of the "multifarious users." Atkins identifies the strengths of current dictionaries in their wealth of information, scholarly work, and concern for the needs of the dictionary user. She sees weaknesses in the redundancy, coverage gaps, inflexible equivalence and collocational selection, distortion caused by disparate needs of source and target languages and by monolingual infor-

mation omitted from bilingual dictionaries, inextensibility of bilingual dictionaries to multilingual dictionaries, lack of integrated thesaural functions, and the user learning curve for dictionary metalanguage.

In "Use and Usability of Dictionaries: Common Sense and Context Sensibility?" Krista Varantola discusses the disparate needs of lay dictionary users and language professionals. She suggests adapting frame semantics, as proposed by Fillmore and Atkins (1998), to facilitate tailoring the electronic dictionary to give users what they need in terms they understand, relying less on context-free, impenetrable text definitions.

Alain Duval, in "La métalangue, un mal nécessaire du dictionnaire actif," the only paper in the volume not in English, addresses the problems of communicating with the user of a bilingual dictionary. The new bilingual dictionary tries to function as an "active dictionary" that supports the user who is trying to generate text in a second language, while still doing the job of the "passive dictionary" that helps the user who is merely trying to understand that language. Duval illustrates the differences between the old and new with examples from several older bilingual dictionaries and points out the advantages of the new approach for users as well as the demands on the user who must understand the expanded metalanguage.

In "Word Groups in Bilingual Dictionaries: OHFD and After," Richard Wakely and Henri Béjoint describe their approach to usage notes in the *Oxford-Hachette French Dictionary.* They discuss their method of identifying lexical sets exhibiting sufficient size, frequency, and behavior commonality. These sets could then be described once with the entries for each set member pointing to the page containing the usage note. They note the pluses and minuses of this approach in terms of practicality, convenience, and usability.

In "Examples and Collocations in the French 'Dictionnaire de langue,'" A. P. Cowie, current editor of the *International Journal of Lexicography,* looks at the treatment of examples in a number of French monolingual dictionaries, including *Dictionnaire du français contemporain, Le Petit Robert, Le Grand Robert,* and *Le Trésor.* He contrasts their methods of blending examples constructed by lexicographers with quotations, exact or adapted. He contends that "the richness, diversity and fitness for purpose of examples in *Le Grand Robert* and *Le Trésor,* especially, are among the finest achievements in modern lexicography."

Juri Apresjan has been a leading figure in lexicography in the Soviet Union and Russia for over 30 years, since he worked with Igor Mel'čuk on the development of the *Explanatory-Combinatory Dictionary.* More recently he has been head of the major Russian machine translation project. In his paper, "Principles of Systematic Lexicography," he argues for the importance of building a systematic lexicon that can interact effectively with a system of grammar rules in the ECD tradition, and he sketches a linguistic basis for this effort.

Charles Fillmore, the creator of frame semantics and the father of the FrameNet lexical resource (Fillmore and Atkins 1998), discusses the problem of "Lexical Isolates," lexical items that "appear to be of unique semantic or syntactic type." He illustrates some of these behaviors with those problem children *let alone, mention, else,* and *ilk.*

In "Sketching Words," Adam Kilgarriff and David Tugwell describe their method of identifying English word sketches from a corpus with part-of-speech tags and a shallow parse, producing an automatic summary of a word's behavior that can assist lexicographers in describing that behavior and can help NLP systems subsequently to perform word sense disambiguation reliably. Each word sketch consists of one of twenty-six word relations, with one, two, or three operands. The salience of a word sketch is defined as a function of mutual information and log frequencies.

318

In "Good Old-Fashioned Lexicography: Human Judgment and the Limits of Automation," Michael Rundell, editor-in-chief of the *Macmillan English Dictionary*, considers whether the advances in automation of dictionary development will lead to the demise of the lexicographer. He argues against this view with several compelling examples, suggesting that each advance identifies new layers of complexity that depend on the lexicographer for analysis.

Patrick Hanks, lead editor on a number of Collins and Oxford dictionaries, suggests, in "Mapping Meaning onto Use," that frame semantics provides a "richer schema for representing meaning than is used in any current dictionary." He advocates using a syntagmatic organizing principle in adjective and verb dictionary entries "rather than (or rather, in tandem with) perceived meaning."

Gregory Grefenstette is probably best known for his work on cross-language information retrieval and its application to Internet text. Here he presents "The WWW as a Resource for Lexicography" in a wide range of languages and the tools needed to extract lexical information effectively. He argues that it is feasible to port a number of tools such as shallow parsers to other languages, especially those using some variant of the Roman alphabet, and that it is time to get to work on this project, as significant amounts of text begin to appear in a number of previously unrepresented languages.

Most of the papers in the volume view natural language processing as a tool for building lexicons. In "Lexical Knowledge and Natural Language Processing," Thierry Fontenelle talks about what is needed in a lexical database to support natural language processing and discusses where that knowledge can be found in existing lexical resources, especially collocational dictionaries, thesauri, and semantic networks. This leads naturally to the problems of representing knowledge about verb alternations and other collocations, using the lexical functions of the *Explanatory-Combinatory Dictionary* (Apresjan, Mel'čuk, and Zholkovsky 1970) and Fillmore's frame semantics.

Annie Zaenen, principal scientist and area manager for Multilingual Theory and Technology at the Xerox Research Centre in Grenoble and co-author of several books about lexical-functional grammar and natural language understanding, makes a convincing case for a depressing conclusion in her "Musings about the Impossible Electronic Dictionary." She looks at the complexity and pressures stifling progress in the creation of multifunctional lexicons and concludes that current trends will continue to produce disparate resources for disparate consumption rather than a unified lexical database.

This book is a EURALEX production in every way, and it is certainly a success. Anyone interested in lexicography should read this volume. It might have been even better, however, if the editors had given some of Sue Atkins's many admirers on other continents a chance to join in. The occasional typographical error should certainly be overlooked in view of the bargain price, which should allow many readers to buy copies of their own.

References

Apresjan, Juri D., Igor Mel'čuk, and Alexander Zholkovsky. 1970. Semantics and lexicography: Towards a new type of unilingual dictionary. In Ferenc Kiefer, editor, *Studies in Syntax and Semantics*, Reidel, Dordrecht, pages 1–33.

Fillmore, Charles J. and B. T. S. Atkins. 1998. FrameNet and lexicographic relevance. In *Proceedings of the First International Conference on Language Resources and Evaluation*, Granada, pages 417–423.

Woody Haynes is working on problems of word-sense discrimination, which led to his participation in the latest SENSEVAL. *Martha Evens,* his former thesis advisor and a former president of the Association for Computational Linguistics, is the author of a book on lexicography and the editor of the Cambridge University Press book *Relational Models of the Lexicon.* Their address is Department of Computer Science, Illinois Institute of Technology, 10 West 31st Street, Chicago, IL 60616; e-mail: skhii@mindspring.com, evens@iit.edu.