

Van Sterkenburg, Piet. (ed.). 2003. *A Practical Guide to Lexicography, (Terminology and Lexicography Research and Practice, vol. 6)*. Amsterdam/Philadelphia: John Benjamins. ISBN 90-272-2330-0

Reviewed by Jean-Claude Boulanger

This impressive multi-author work is for the greater part devoted to general monolingual dictionaries. It consists of two parts. The first deals with forms, contents and uses of dictionaries; the second concentrates on linguistic corpora and dictionary compilation. Each part is divided into chapters: three for the first part and four for the second. The chapters themselves are subdivided into sections with separate contributions. There are, in total, twenty-nine articles, most of which use the English language for exemplification. What follows is a short description of the content of each article.

The book opens with a preface by Piet van Sterkenburg who reflects on the idea that an adequate description of the vocabulary requires links to databases and awareness of the most recent models of knowledge developed in semantics and pragmatics. In the editor's view it is impossible to describe the morphology and syntax of a language without these considerations. Such new orientations, therefore, would appear to demand a re-examination of lexicography, especially since linguisticians, who in the past had kept aloof from lexicographic questions, are now taking an interest in the field.

The first chapter deals with the foundations of lexicography and dictionaries. It consists of five sections. The first (Piet van Sterkenburg) is a historical survey. The author attempts a definition of the term *dictionary*, which cannot be undertaken without reference to a prototype, in this case the general monolingual alphabetical dictionary. He reviews some definitions, notably those of Henri Béjoint, Ladislav Zgusta and Bo Svensén all of which are concerned with the printed dictionary, thus not taking account of the electronic dictionary. Subsequently the author examines the requirements for a dictionary to be recognised as such. Three conditions are established: form, function and content. The criterion of form contrasts printed dictionaries with electronic dictionaries, including CDROMs; the function of a dictionary is to inform the user about the linguistic behaviour of the most representative words of the lexicon and of measuring their level of acceptability with respect to a standard; the criterion of content covers the different types of microstructural information: pronunciation, word class, sense, register, etc. On the basis of these conditions a definition emerges which makes it a reference work with the

objective of recording the words of a language with the purpose of supplying a user with the linguistic information necessary for producing and understanding his native language. The article concludes with a short history of dictionaries and its avatars, the glossary and the vocabulary.¹

The second article (František Čermák) deals with source materials for dictionaries. Where do these data come from? Normally, a language dictionary is built around primary sources (archives, corpora) and secondary sources (enquiries among users, encyclopedias, and, more recently, the web). Increasingly text corpora are used for elaborating primary sources. The author explains and specifies each type of source. The second part of the article looks at automated text corpora: the methodology of their setup, the distribution between written and oral source, text types (general or specialised language) and the proportions of each type. The structure of English and Czech corpora are used as examples.

The third article (Paul Bogaards) is on the various uses of dictionaries. Four aspects are examined: user surveys, metalexigraphy, models of dictionary usage and experimental research. The first topic deals with enquiries into the use of dictionaries. The chief reason for consulting a dictionary is the search for an unknown word one has heard or read in a book. The spelling is the first concern for written work. Metalexigraphic studies consist of the critical analysis of dictionaries. This is a job for specialists. There is a great deal of interest in learners' dictionaries, especially in English and German-speaking countries. Other areas of investigation are the nature and presentation of grammatical information, illustrations, examples, definitions, etc. The third topic synthesizes the process of dictionary consultation by means of a graphic model of use starting from a linguistic problem that requires a solution. The last paragraph presents an example of dictionary consultation on the basis of the described model.

The fourth article (Rufus Gouws) is occupied with types of dictionary articles, their structure and different types of lemmata. A dictionary contains three types of texts: the preliminary or front matter, the alphabetical dictionary itself and the back matter. The author limits his concern to the main part, i.e. the dictionary proper. A dictionary article has two fundamental structures; firstly, that of the lexical items (Josette Rey Debove's microstructural information) and, secondly, that of the diacritic metalanguage or structural indicators (special symbols, typographic characters, asterisks, brackets, etc.). Functional information is of two types: information about the form of words (spelling, pronunciation, word class...) and information about the meaning or meanings of words (definition, synonyms, antonyms...). Dictionary articles are of two types: articles with a main headword or lemma together with subheadings (the

principle of clustering) and articles with a simple structure (one word equals one article). In either case there may be a synoptic substructure which affects encyclopedic or other functional information. The author also discusses different types of lemmas or entries, i.e. word components, simple, compound and complex words. The article concludes with a glance at alphabetical ordering which is compared to a systematic ordering of entries.

In a fifth article, Piet Swanepoel presents a typology of dictionaries from a pragmatic point of view. First he explains the theoretical foundation of his typology. He distinguishes between monolingual and bilingual dictionaries. He comments on existing typologies and explains their usefulness. He describes each type of dictionary and places it in a hierarchical order. In some detail the author comments on his typology which is based on the one established by Ladislav Zgusta in 1971. Each type or subtype is exemplified by one or several titles of dictionaries.

The theme of the second chapter of the book is descriptive lexicography, i.e. the content of dictionaries. The first contribution (Johan de Caluwe and Ariane van Santen) deals with the phonological, morphological and syntactic data in monolingual dictionaries. The authors give a traditional presentation of the content and the role of each type. They mention the pronunciation of words outside a particular context with regional variations which occasionally occur in English. Morphological information is of two types: grammatical data proper (gender, number, conjugation...) and morpholexical data (derivation, word composition...). The authors also discuss the inflection of nouns, adjectives and verbs. The different modes of words are analyzed from the point of view of their syntactic information: for example, countable and non-countable nouns, determiners in conjunction with an adjective, verbal categories (transitive, intransitive, etc.).

The second article of this chapter (Dirk Geeraerts) explores the domain of semantics: definitions and meanings. The author identifies five types of semantic information in an article and proposes to examine them by formulating five questions:

1. Do I focus on the senses of the individual words or consider that words do not occur in isolation?
2. Which readings of a word do I consider relevant? In other words, how does one make the choice of meaning of polysemic words, including the LSP meaning?
3. Which type of meaning do I have to define and how do I formulate the definition: with respect to the conceptual reference (the denotation) or with

respect to the more metalinguistic criteria, which may correspond to the morphosyntactic definition?

4. Which linguistic perspective do I take, the logical definition by intension or extension?
5. Which definitional format do I use, the definition by synonymic phrase or the definition by complete sentences which includes the definiendum, as, e.g., in school dictionaries: “A X is a Y which...”?

The third article (Stanisław Prędotą) studies the dictionaries of proverbs. The author gives a brief history of anthologies of proverbs and proposes a typology of these works for English. He also examines the macrostructure of these dictionaries from the point of view of their alphabetical or thematic ordering. This is followed by an analysis of the microstructure of English proverb dictionaries.

In the fourth article Igor Burkhanov studies pragmatic data such as usage and register labels and examples. He also analyzes dictionaries of style and collocations. The first of these are similar to writing manuals and deal with rules of composition. We may ask whether they can really be called dictionaries? In his typology of usage labels, he explains that the categories and the number of indicators may vary with the range of the vocabulary in the dictionary. Regarding exemplification, currently discussion centres around the created example versus the citation. For Burkhanov examples serve mainly to complement the definition by incorporating semantic features like hyponyms or by contributing referential information of an encyclopedic nature, for example, data which go beyond purely functional information.

The fifth article (Johan de Caluwe and Johan Taldeman) is focused on morphological data. It deals with the treatment of derived/compound words as opposed to simple words. This question is studied from the point of view of reception — decoding of words — and that of the production of dictionaries — encoding of words. A further distinction is also made between the printed and the electronic dictionary. Among the topics concerning production we find a discussion of the concentration of entries under headwords against single entries, the relation between the derivative and its base word which can be established via etymology or word formation, and the decision whether to include affixes as headwords. From the point of view of reception, two questions arise. The first concerns derivatives and compounds whose morpho-semantic structure seems irregular or special, i.e. opaque and unpredictable. Since the elements constituting such words do not necessarily have compositional meaning, such opaque words must be given as headwords so that users can decode

them. The second question concerns the regular and predictable nature of form and meaning; that is, whether these words should be included in the dictionary. The author adduces two reasons for retaining these words: their polysemy and their conditions of usage. From the point of view of production or word formation, the dictionary must provide lists of derivatives in order to illustrate the word-forming vitality of the root words. The same applies to compounds: they should be listed in order to serve as models for future lexical creations.

In the sixth article, Piet van Sterkenburg deals with onomasiological data and traces a short history of the onomasiological dictionary. In a first part, he analyses different dictionaries of this category: thesaurus, synonym dictionaries, reverse dictionaries and picture dictionaries. These dictionaries cannot escape alphabetical ordering since they need indexes. Then, the author lists the onomasiological data contained in semasiological dictionaries. He goes into some detail on the question of definitions. The history of these dictionaries is told in the second part of the article. Starting with the thematic glossaries of the 13th-14th centuries, he moves on to Gabriel Girard (1677–1748) whose work was very influential all over Europe. He does not forget Mark Roget's thesaurus, first published in 1852, which became the model of this type of book.

The third chapter is concerned with the different special types of dictionaries. The first article (Mike Hannay) deals with bilingual dictionaries and describes their microstructural organisation in terms of user needs. The author accounts for the direction of dictionary consultation, i.e. from the user's language to the foreign language (from the known to the unknown) as well as from the perspective of understanding of the foreign language word (from the unknown to the known). Another section deals with the unidirectional or bidirectional nature of bilingual dictionaries. Finally, the author examines the status of the user and investigates the environment of usage of learners' dictionaries of a second language.

In her article, Lynne Bowker studies the relationship between lexicography and specialised dictionaries. She singles out the peculiarities of this type of research in contrast to general lexicography; then she examines some characteristics of LSP dictionaries before discussing the different stages of compilation of this type of dictionary. The article finishes with a quick look at other types of specialised dictionaries such as dictionaries of regional variants. LSP lexicography deals with terms related to a field of knowledge. The author relates this type of research to terminology, even admitting that it is very difficult to distinguish the one from the other. She reminds us also that a LSP dictionary describes groups of terms linked to a terminology, that such a dictionary can be monolingual, bilingual or multilingual and that its users range from experts to lay

persons. Finally she specifies some elements of macro and microstructure. This type of reference work appears in the form of books, CDROMs, and on the web, especially for term banks.

This article concludes the first part of the book. The second part is more specifically devoted to linguistic corpora and the compilation of dictionaries. In the fourth chapter, John Sinclair presents two articles in which he is interested in the creation of corpora and their treatment for lexicographic purposes. Firstly, the author presents the conditions for assembling a text corpus: content, variety of texts, preserving the integrity of the input, sophisticated typology of texts, etc. In the second article, dealing with computational aspects, he presents such problems as the recognition of words as well as the subtle distinctions between lower case and upper case letters. He also discusses two features of text, namely linearity and legibility, automatic tagging of texts by means of systems such as SGML and its subset XML. Besides, conservation requires archives in order to protect the data and ensure their permanence. The topic of annotation is dealt with from two viewpoints: document type description (DTD) and annotation of words by means of labels in order to record all manner of decisions with reference to the analyses. The author discusses the advantages and disadvantages of types of annotation.

The third article in this chapter (Truus Kruyt) explores multifunctional linguistic databases by which the author means lexicographic data, such as text corpora, dictionaries and thesauri. By multifunctional he means the re-use of linguistic material for other purposes than those for which it was originally assembled. The term also means that the linguistic data have been conceived for multiple uses. One part of the article traces the history of automated lexicographic data since 1960. Another section illustrates the multifunctionality of data on hand of four examples of re-use; the last section is meant as an outlook, especially by examining the new standards like the Text Encoding Initiative (TEI) and EAGLES (Expert Advisory Group on Language Engineering Standards), by discussing the preferred theoretical approach from the linguistic point of view and by assessing evaluation and legal implications.

The fourth article (Daniel Ridings) is occupied with the history of a project for developing an Afrikaans dictionary. It gives a detailed analysis of the methodology and software specification necessary for the preparation of this task which started in 1926 in the traditional manner. The software package 'Onoma' which is used for the production of dictionary articles is described in great detail.

The fifth chapter deals with dictionary design, i.e. the planning of electronic dictionaries. The first article in this chapter (Lineke Oppentocht and Rik

Schutz) scrutinizes the effect of technological developments on the design of electronic dictionaries. The authors show the advantages of this type of tool for the user; electronic dictionaries can be used for translation, Internet searches, automatic summaries or abstracts, etc. In order to support their claims the authors describe existing achievements, current projects and ideas in the stage of planning. Currently existing works are quasi-copies of printed dictionaries whereas the new ones will be quite independent from the older models and have their own format. So, since space is no longer a constraint, abbreviations will no longer be justified and symbols can be made more explicit, variants and feminine forms of nouns and adjectives can be placed in their alphabetical sequence, crossreferences will become hyperlinks, repetitive phraseology will be controlled, the recognition of complex units will become easier, etc. In addition the electronic dictionary permits the presentation of articles under different aspects: semantic hierarchies only, citations and examples only, definitions only, synonyms in different sequences, etc. It also permits the construction of onomasiological fields and the diversification of specific searches (words with identical final endings, a suffix, for example, or words whose pronunciation of the final syllable is the same). Among new developments we note the facility for frequent updates, linking of dictionaries, incorporation of other reference works into the main dictionary, etc.

In the second article, Krista Varantola establishes connections between linguistic corpora and the compilation of dictionaries. In the first section she approaches her topic from the point of view of the professional user and his needs. The solutions she proposes have been applied in existing dictionaries or they could be implemented in future ones. Her main criterion is the 'usability' of the dictionary. She also explains how the electronic dictionary introduces major changes in dictionaries and in the techniques of their compilation. Since these dictionaries are based of text corpora, these text should also be available to the dictionary user. The author also lists sources of frustration for the user, for example when he searches for information which is not lexicographic in nature, the neglect of front matter with the pretext that there is no time for this, the excess of information when the user is looking for a very precise answer. Text corpora will also permit refining the semantic zone of words by a neater specification of the collocation. The author also deals with the active use of the dictionary — writing — and the passive use — reading. The second part of this article describes the importance of the front matter of dictionaries for the user. Some further points are covered, such as abbreviations and symbols, the actual appearance of the book, the layout and typography, illustrations.

The third article (Sean Michael Burke) deals with the nature of on-line dictionaries. The author shows first that on-line lexicography uses the principles and methods of traditional lexicography while perfecting and adapting them to the new medium. Secondly, he wants to demonstrate the advantages of on-line dictionaries over the printed book. Then he examines the macrostructure by which he understands the mode of access to the entries and thus to the various details of functional information about the word. He explains how a query can achieve a satisfactory answer even though the question contained errors. An algorithm takes care of establishing the proper connections. It can also reconstruct the lemmatized form on the basis of conjugated or declined forms. Then he examines the content of the microstructure. He emphasizes, in particular, the new data which the on-line mode permits to access. The question of available space is discussed. From the necessarily limited space of the dictionary one moves to the almost infinite space of the on-line dictionary. Among the effects of the new medium we note the abandonment of abbreviations and the more frequent starts of new lines which makes it easier to itemize information. Among new types of information we note complete paradigms, especially those of declensions in Latin dictionaries, the increase in the number of examples and encyclopedic information, the addition of illustrations, the “sonorisation” of pronunciation, the sound picture of certain noises and animal voices, etc.

The sixth chapter addresses the production of dictionaries. In the first article Geert Booij is interested in the codification of phonological, morphological and syntactic information. The approach is that of an electronic dictionary which will lead to a printed dictionary. The electronic version is easier to consult than the printed version since it is no longer necessary to rely on the alphabetic sequence. Furthermore, the electronic dictionary permits access to corpora of spoken language, whereas the printed dictionary is essentially or almost entirely based on the written language. The author then discusses the selection of phonological data, i.e. the pronunciation of words outside a particular context. The electronic dictionary permits sound recording of pronunciation and listening to it rather than having to be content with an IPA transcription. The new technology can also assist in providing information on accents and syllable division, phenomena which are relevant for writing. Morphological information is important for both the lexicon — word formation — and for grammar — conjugation and inflections, for example. The author favours a description of affixes and affixoids. Syntactic information has a bearing on restrictions on constructions, contextual constraints as well as collocations.

In the second article of this chapter, John Simpson focuses on the constructions of examples and the choice of citations. He recalls the tradition of handwritten index cards and compares it with the modern methods based on automated corpora. He also shows how examples and citations complement the definitions.

The third article (Fons Moerdijk) provides an opportunity to examine the codification of semantic data, especially one or several senses or one or several denotations of a word. Definitions are supplemented by networks of semantic relations (synonyms and antonyms), domain indicators and, in the case of bilingual dictionaries, equivalents. The author deals with three aspects: the identification of the sense or senses, their ordering and the definitional formulae. The process is based on a text corpus. The contexts of a word are regrouped and analysed in order to extract the possible meanings, always avoiding the selection of one meaning for each occurrence where the context differs only little from similar ones. Equally, it is necessary to identify the border which separates polysemy from homonymy. The ordering of senses can be carried out in one of three ways: historical, frequency or logico-semantic ordering. Finally the conclusion reminds us that definitions take the form of analysis, description of the concept or simple synonymy.

The next article is signed by three authors: Henk Verkuyl, Maarten Janssen, and Frank Jansen. Their contribution focuses on the codification of usage labels. They start by defining 'usage label' which they limit to the indicators which accompany a semantic element: register, figurative use, LSP domain indications, etc. The two sections of the article deal with types of labels and their functions respectively. The typological classification is conventional whereas the discussion of indications of political correctness breaks new ground. The authors also comment on the extent to which the functions of these indicators are often misunderstood or poorly managed.

The fifth article (Nicoline van der Sijs) deals with the codification of etymological information. On the basis of fourteen arbitrarily selected dictionaries in several languages (English, German, Dutch, Swedish, French, Italian and Spanish) the author attempts to answer four questions:

1. Are all, or only some of the entries given an etymology?
2. What choices have been made in the treatment of native words and loan-words?
3. Has attention been paid to both form and meaning changes?
4. Have dates of first occurrence been provided?

The author examines the etymological strategies of each dictionary. Then she analyses the changes which electronic dictionaries make to the representation of etymological information, especially because of the additional space and the establishment of hyperlinks with historical documents.

The seventh chapter contains four contributions under the heading of examples of design and production criteria for major dictionaries. The first article (Wim Honselaar) is aimed at bilingual dictionaries. The author describes all the stages of a dictionary project, from the idea of its formal architecture, to the target usergroup, the contents, the financing of the project, the selection of the team of collaborators, etc. The organisation of the work is dependent upon the editor's manual, the relationship of the projected dictionary with existing ones, the up-dating of the dictionary, the procedures for checking the content as well as with the preparation and checking of the final product.

The second article (Willy Martin and Hennie van der Vliet) focuses on terminological dictionaries. The authors start by comparing LSP dictionaries with general language ones. Then, the link is established between specialised dictionaries and databases which nowadays has become inescapable. The rest of the article concentrates on this aspect of terminological dictionaries with emphasis on the stages of developing a conceptual model, and especially on the place, the role and the impact of such a model on data bases. Each stage is carefully presented in great detail.

The following text (Ferenc Kiefer and Piet van Sterkenburg) deals with the design and production of monolingual dictionaries. The hypothesis is based on the idea that there is as yet no agreement among lexicographers about the ideal macro and microstructure. The authors think that the centuries of experience gathered in the production of dictionaries are a guarantee that there is nevertheless a basic framework and a minimal structure which permit the allocation and description of words in the dictionaries. They seek answers to the following questions: which factors have to be considered for the planning of a dictionary and which are the variables which may influence the development of the project with respect to the original conception? For their study they put aside text corpora and electronic processing, adopting a traditional approach. Then they present the different stages of a dictionary project: the place of the dictionary on the book market, the cost of the enterprise, the target users, the choice of words, their definitions and their onomasiological relationships, the macro and microstructure. Each section of a dictionary article is precisely described. Aspects more concerned with policy are also listed: the editor's guide, the layout, the publicity supporting the publication, and general planning (team, timetable, cost).

The fourth article in this chapter (Stefania Nuccorini) explores the notion of an ideal collocation dictionary for English. The author mentions terminological problems, for example the difference between the terms ‘collocation’, ‘phraseology’, and ‘idiomatic expression’. Her contribution consists of a description and exemplification of seven English phraseological dictionaries. She critically analyses the content of these works and the place of collocations in them. She also considers the introductory texts to these dictionaries which are intended to guide the user’s understanding of how these dictionaries function.

This set of articles is followed by an important glossary of the terminology of lexicography; a substantial general bibliography listing all the references given in the different articles and a general index round up this work.

While innovative in certain of its aspects, especially by favouring an approach linking text corpora to lexicographic work itself, this book leaves one with the general impression of ‘déjà vu’, most of the topics of this state-of-the-art survey, distributed over twenty-nine articles, already being abundantly documented in the relevant literature. Equally, the content is not without its repetitions, as these summaries have shown. But, in the end, the team assembled here nevertheless offers a solid introduction to current lexicography. The book is user-friendly and intended to be a safe and efficient guide for those who want to know how dictionaries are made. At the level of principles and methods, the team clearly demonstrates that in lexicography there are “universals” of dictionary production, at least with respect to western languages.

Note

1. On this topic see Boulanger, J.-C. 2003. *Les inventeurs de dictionnaires. De l’eduba des scribes mésopotamiens au scriptorium des moines médiévaux*. Ottawa: Les Presses de l’Université d’Ottawa.

Reviewer’s address

Jean-Claude Boulanger
Département de langues, linguistique et traduction
Université Laval
Québec (Québec) G1K 7P4
Canada

Jean-Claude.Boulanger@lli.ulaval.ca