

OORDRUK UIT / OFFPRINT FROM

LEXIKOS 16

(AFRILEX-REEKS / SERIES 16: 2006)

**BURO VAN DIE WAT
BUREAU OF THE WAT**

STELLENBOSCH

Rufus H. Gouws, Ulrich Heid, Wolfgang Schweickard and Herbert Ernst Wiegand (Editors). *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Special Focus on Computational Lexicography.* Berlin/New York: Mouton de Gruyter.*

The trilingual encyclopedia *Dictionaries. An International Encyclopedia of Lexicography* (edited by Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand and Ladislav Zgusta) was published between 1989–1991 in three volumes, consisting of a total of 335 articles divided into 38 chapters and printed on 3 355 pages including the index. The envisaged supplementary volume which will be published in English will consist of 22 chapters, containing 136 articles. This will not exceed 1 360 pages.

The supplementary volume pursues the objectives formulated in the preface of the first volume (HSK 5.1) which are complemented by others, following from the practice-internal and theoretical developments of the last 15 years. With the supplementary volume it is consequently endeavoured to achieve the following objectives:

- Expansion and extension of the presentation of lexicography — especially the printed and electronic linguistic lexicography — of all language circles of the world
- Continuation of the presentation of the status and function of lexicographic reference works up to the present within the cultural systems of societies, with special reference to the new digital forms of lexicography
- Continuation of the already presented history of lexicography by means of selected examples of larger cultures up to the present
- Expansion, deepening and modification of presented excerpts of a general theory of lexicography in its four sections, i.e. research into the use of dictionaries, research of dictionary criticism, and historical and systematic dictionary research, with the aim that metalexigraphy can document its academic status as scientific discipline with an independent formation of a theory, that has been established during the last decade
- Complementation of the description of all phases of the lexicographic process, especially by means of the presentation of the new methods based on electronic corpora, by means of the discussion of the new pos-

* The editors welcome suggestions for additions to these articles. Please send suggestions before 31 December 2006 to the following address: herbert.ernst.wiegand@gs.uni-heidelberg.de. The editors will give the necessary consideration to suggestions received.

sibilities of computer-assistance as well as the consideration of the linking possibilities of different types of lexicographic processes

- Presentation of lexicographic training and lexicographic institutions
- Presentation of new metalexigraphic methods, and, as focus of the volume: An extensive and detailed presentation of computational lexicography with regard to its historical development as well as all the important research fields, research perspectives, methods and tools of the last two decades.

The conception and grouping, with which the outlined objectives of the supplementary volume will be achieved, are briefly sketched in the following paragraphs. The Arabic numbers refer to the classification given at the end.

The supplementary volume has four chapters directed at new lexicographic concepts and the further development of new beginnings.

In Chapter I, **Dictionaries in Modern Society: Current Status and Perspectives**, the first article gives the present and future prospects, problems and development possibilities of printed dictionaries in their different typological forms. The collaboration in the lexicographic processes of printed dictionaries with corpus linguistic approaches is emphasized. In the second article a corresponding presentation is given of computational lexicography and its influence on linguistic lexicography and other fields of research.

In the last decade, textual structures of printed dictionaries have often been in the focus of lexicographic research. In Chapter II, **New Developments in Lexicographic Theory I: Textual Structures**, the eight articles (nr. 3-10) attempt to present this research as exhaustively as possible in the following manner: A review article briefly indicating all the types of textual structures is followed by a review article on the following selected types of structures: macrostructures, access structures, microstructures and mediostuctures as well as textual architectures. The next article treats addressing and addressing structures and in the tenth and final article of this chapter a presentation is given of the different types of dictionary articles.

In Chapter III, **New Developments in Lexicographic Theory II: Dictionary Types and Functions**, new dictionary types and new developments in selected dictionary types are presented in ten articles (nr. 11-20). Consideration is also given to the lexicography of language for general purposes and languages for special purposes. Dictionaries of association, new types of cognitive dictionaries, language contact dictionaries, bridge dictionaries and recent developments in English, French as well as German learner dictionaries are treated. There are also review articles of new linguistic dictionaries as well as all the existing special purpose dictionaries of lexicography and those in progress. Article 20 concludes the chapter with a treatment of dictionary functions.

Chapter IV, **New Developments in Lexicographic Theory III: Selected Dictionary Subjects**, consists of eight articles (nr. 21-28). From the numerous themes, directed especially at aspects of dictionary subject matter, treated in recent metalexigraphic research, relevant and innovative aspects have been selected in such a way that consideration is given to a relative wide spectrum of problems. Culture-bound lexical items in monolingual and bilingual lexicography, language policy and lexicography, sensitive items, and meaning in prototype theory and in frame theory are treated in separate articles. Furthermore the concept of semiotaxis is presented as well as new developments in the lexicographic treatment of collocations and lexicographic examples.

The last chapter dealing with new developments in the theory of lexicography is Chapter V, **New Developments in Lexicographic Theory IV: Research in Dictionary Production and Use**, consisting of eight articles (nr. 29-36). The chapter starts with a presentation of the structure of lexicographic processes and their phases, contains a presentation of dictionary management, an article on the theory of dictionary use and review articles on empirical usage research and on methods developed and employed in recent research on dictionary use. Furthermore there are articles on the concept of simultaneous feedback, research on usage in the domain of electronic dictionaries and ways and means of establishing a dictionary culture.

The sixth chapter deals with organizational questions and is entitled: **New Developments in Lexicographic and Metalexigraphic Organization**. The five articles (nr. 37-41) present review articles on the training of lexicographers and training possibilities, the establishment of lexicographic units, the spectrum of existing lexicographic associations, principles of the evaluation of dictionaries and dictionary projects and aids playing a role in metalexigraphic research.

Chapter VI is followed by six chapters dealing with recent developments in the lexicography of single languages and the lexicography of single languages that could not have been dealt with in HSK 5.1-5.3.

Chapters VII to XII are dedicated to the lexicography of individual languages. The mutual aim of the descriptions are to document the development of lexicographic research since the start of the nineties of the 20th century. The point of departure of these supplementary articles is the treatment of the lexicography of individual languages presented in HSK 5.1-5.3. With this background the latest developments with regard to methodology and form, from both a synchronic and a diachronic perspective, are discussed and commented on. The following focal areas have been identified for the supplementary volume 5.4:

Chapter VII, **New Developments in the Lexicography of Individual Languages since 1990 I: The Ancient Languages of the Near East and the Classical Languages**, treats, by means of examples, the lexicography of the ancient Eastern languages as well as Greek and Latin. Chapter VIII, **New Developments in the Lexicography of Individual Languages since 1990 II: The Ro-**

mance Languages, presents the lexicography of the Romance languages (with the focus on the Iberoromance languages, French, Italian and Romanian). Chapter IX, *New Developments in the Lexicography of Individual Languages since 1990 III: The Germanic Languages*, is dedicated to the complete field of Germanic languages (Scandinavian languages, English and American English, German, Dutch and Afrikaans). Chapter X, *New Developments in the Lexicography of Individual Languages since 1990 IV: The Slavic Languages*, treats the lexicography of selected Slavic languages, divided into the domains of Eastern, Southern and Western Slavic languages.

Chapter XI, *Lexicography of Selected Asian Languages*, illustrates by means of the lexicography of Chinese, Japanese and Korean a selection of the total domain of the Asian languages. Chapter XII, *Lexicography of Selected African Languages*, contains eight articles (nr. 64-71) focusing on the lexicography of certain language families (the Nguni and Sotho languages), individual languages (Shona, Chiluba, Fang and Swahili) and the languages of certain geographical regions (Central Africa and Western Africa). The articles in this chapter complement the limited presentation of the lexicography of the African languages presented in HSK 5.3 and recognize the rapid development of lexicography on the African continent.

After the treatment of the lexicography of selected single languages follow ten chapters on computational lexicography.

The section on computational lexicography starts with Chapter XIII, *The History of Computational Lexicography*, giving an overview of the computational lexicographic work of the 1950s and 1960s and, in a second article, an overview of the time before 1970-1980.

The subsequent two chapters deal with typologies of electronic dictionaries, one for interactive dictionaries and one for dictionaries for natural language processing (NLP).

Chapter XIV, *Typology of Electronic Dictionaries I: Electronic Dictionaries for Human Use*, consists of nine articles (nr. 74-82). The first three cover design criteria for interactive electronic dictionaries, the 'added value' that the user may get from the computational medium, as well as principles of user interface design for electronic dictionaries for human use. The subsequent articles are devoted to electronic versions of printed dictionaries, Internet dictionaries, electronic learner's dictionaries and context-sensitive dictionaries, i.e. those that guide the reader of an online text to the relevant (e.g. idiomatic) reading of a word. Finally, the overview is completed by articles on large-scale documentary electronic dictionaries, electronic encyclopedias and products that combine dictionary look-up with corpus search.

Chapter XV, *Typology of Electronic Dictionaries II: Electronic Dictionaries for Machine Use*, is still typological in nature, but focuses on NLP applications. Individual articles describe lexical needs, representation techniques, descriptive approaches and major realizations in each of the following subdomains of NLP: speech recognition and speech synthesis, text processing and

spelling or grammar checkers, information retrieval, information extraction and data mining, question answering, dialogue systems, natural language understanding systems, text generation systems, machine translation and computer-assisted language learning. The last article discusses generic large-scale electronic dictionaries intended for multiple applications.

Chapter XVI focuses on lexical presentation formalisms: *Models for the Representation of Dictionaries: The Form Aspect*. In seven articles (nr. 93-99), it deals with the major representation formalisms for electronic dictionary resources: XML (and SGML), databases, feature structure-based formalisms, formalisms of knowledge representation as used in artificial intelligence, and hypertext. One article discusses dedicated formalisms for lexical representation and one describes objectives and topics of international standardization in the field of representation formalisms for electronic dictionaries.

Chapter XVII, *Models for the Representation of Linguistic Data in Electronic Dictionaries: The Content Aspect*, is subdivided into twelve articles (nr. 100-111). The first seven articles are each devoted to one 'level of linguistic description', its needs, data types, coding and representation schemes. These articles cover pronunciation, inflectional and word formation morphology, syntax, semantics and pragmatics, as well as dictionaries which explicitly attempt describing and representing the interaction between syntax and semantics. The next article treats idiomaticity in electronic dictionaries, followed by one on multimodal dictionaries. Two articles are devoted to international standardization work for the linguistic description of lexical material in monolingual as well as in bilingual and multilingual electronic resources. The last article of this chapter describes typologies of metadata for electronic dictionaries.

Chapter XVIII, *Models for Non-alphabetical Computational Dictionaries*, is entirely devoted to WordNet and its developments and applications, to resources with world knowledge (such as CYC) and to both linguistic ontologies and ontologies constructed in research towards the Semantic Web. Besides reporting facts about the current state-of-the-art, these four articles (nr. 112-115) are also intended to allow for a comparison between the different approaches.

So far, all chapters of this part of the handbook concern the electronic dictionary as a product: its types, its contents, and its form. Chapters XIX, XX and XXI are devoted, however, to dictionary making and to all its computational aspects. As most dictionary making is nowadays supported to some degree by computational tools, these chapters obviously concern both electronic and paper dictionaries, for men and machines. One chapter is devoted to corpus design for lexicography, one to tools for data acquisition and one to support for dictionary writing.

Chapter XIX, *Computer-based Dictionary Making I: Acquisition of Lexical Data from Corpora — Corpus Design*, encompasses four methodological articles, two on corpus design criteria for lexicography (monolingual and bilingual), one on existing large corpora for lexicography (as e.g. many of the na-

tional corpus projects), and a last one on possibilities and limitations of the use of the World Wide Web for lexicography. Chapter XX, **Computer-based Dictionary Making II: Acquisition of Lexical Data from Corpora and Machine-readable Dictionaries — Tools and Procedures**, deals with tools for lexicographers. Six of its nine articles (nr. 120-125) cover the major aspects of corpus-based data provision for dictionary making: corpus pre-processing, the corpus-based design of a macrostructure, concordancing, the extraction from corpora of data on syntactic and collocational properties of words, and tools for the use of parallel and comparable corpora. The three other articles (nr. 126-128) concern the use of electronic versions of dictionaries as data sources: retrodigitization, the analysis of electronic versions of print dictionaries and the reversal of bilingual dictionaries A → B to make them usable as a starting point for a dictionary of the language pair B → A. Finally, Chapter XXI, **Computer-based Dictionary Making III: Computational Support for the Writing of Dictionary Articles**, gives a state-of-the-art overview of tools for dictionary writing and for consistency control, of workbenches for corpus-based lexicography and of tools supporting lexicographers in dictionary updating by means of a comparison between data from an existing dictionary and the results of corpus analysis.

Computational work in terminology is the topic of Chapter XXII: **Computational Terminography**. As terminology and terminography are extensively dealt with in the HSK handbook on specialized language, only truly computational aspects of terminography are dealt with here: the representation of terminological data in terminological databases and in formats for the exchange of terminological data, and approaches to and challenges in the extraction of terminological data from texts. Finally, the chapter includes an article on existing large-scale termbanks.

I. Dictionaries in Modern Society: Current Status and Perspectives

1. Printed dictionaries in the electronic era
2. The impact of computational lexicography

II. New Developments in Lexicographic Theory I: Textual Structures

3. Textual structures in printed dictionaries: An overview
4. Macrostructures in printed dictionaries: An overview
5. Access structures in printed dictionaries: An overview
6. Microstructures in printed dictionaries: An overview
7. Mediostructural components in printed dictionaries: An overview
8. Textual architectures in printed dictionaries: An overview
9. Addressing and addressing structures in printed dictionaries: An overview
10. Types of dictionary articles in printed dictionaries: An overview

III. New Developments in Lexicographic Theory II: Dictionary Types and Functions

11. Dictionary of associations
12. New types of cognitive dictionaries
13. Dictionary of language contact

14. Bridge dictionaries
 15. New developments in learners' dictionaries I: English and French
 16. New developments in learners' dictionaries II: German
 17. New developments in lexicography for special purposes I: An overview of linguistic dictionaries
 18. New developments in lexicography for special purposes II: An overview of dictionaries of lexicography
 19. New types of historical dictionaries
 20. Lexicographic functions
- #### IV. New Developments in Lexicographic Theory III: Selected Dictionary Subjects
21. The treatment of culture-bound items in dictionaries
 22. Language policy and lexicography
 23. The treatment of sensitive items in dictionaries
 24. The treatment of meaning in dictionaries and prototype theory
 25. The treatment of meaning in dictionaries and frame theory
 26. The concept of semiotaxis
 27. New tendencies in the treatment of collocations
 28. New developments in the selection of examples
- #### V. New Developments in Lexicographic Theory IV: Research in Dictionary Production and Use
29. The lexicographic process
 30. Theory of dictionary management
 31. Theory of procedures of dictionary use
 32. Empirical research in dictionary use since 1990
 33. Methods in the research of dictionary use
 34. The concept of simultaneous feedback
 35. User research in the field of electronic dictionaries: Methods, first results
 36. Ways and means to establish a dictionary culture
- #### VI. New Developments in Lexicographic and Metalexigraphic Organization
37. Lexicographic training: An overview
 38. Establishing lexicography units
 39. Lexicographic associations
 40. Evaluation of dictionaries
 41. Aids in metalexigraphic research
- #### VII. New Developments in the Lexicography of Individual Languages since 1990 I: The Ancient Languages of the Near East and the Classical Languages
42. The ancient languages of the Near East
 43. The ancient languages of Greek and Latin
- #### VIII. New Developments in the Lexicography of Individual Languages since 1990 II: The Romance Languages
44. Iberoromance I: Historical and etymological lexicography
 45. Iberoromance II: Synchronic lexicography
 46. French I: Historical and etymological lexicography
 47. French II: Synchronic lexicography

48. Italian I: Historical and etymological lexicography
 49. Italian II: Synchronic lexicography
 50. Romanian
- IX. New Developments in the Lexicography of Individual Languages since 1990**
III: The Germanic Languages
 51. Scandinavian languages
 52. English and American I: Historical and etymological lexicography
 53. English and American II: Synchronic lexicography
 54. German I: Historical and etymological lexicography
 55. German II: Synchronic lexicography
 56. Dutch
 57. Afrikaans
- X. New Developments in the Lexicography of Individual Languages since 1990**
IV: The Slavic Languages
 58. Eastern Slavic languages
 59. Southern Slavic languages
 60. Western Slavic languages
- XI. Lexicography of Selected Asian Languages**
 61. Chinese
 62. Japanese
 63. Korean
- XII. Lexicography of Selected African Languages**
 64. Lexicography of the Nguni languages
 65. Lexicography of the Sotho languages
 66. Lexicography of Shona
 67. Lexicography of the languages of Central Africa
 68. Lexicography of Chiluba
 69. Lexicography of Fang
 70. Lexicography of Swahili
 71. Lexicography of the languages of Western Africa
- XIII. The History of Computational Lexicography**
 72. The early history of computational lexicography: The 1950s and 1960s
 73. Computational lexicography and corpus linguistics until ca 1970–1980
- XIV. Typology of Electronic Dictionaries I: Electronic Dictionaries for Human Use**
 74. Design criteria and 'added value' of electronic dictionaries for human users
 75. Principles of user interfaces for electronic dictionaries for human users
 76. Electronic versions of printed dictionaries
 77. Electronic pedagogical dictionaries
 78. Context-sensitive electronic dictionaries
 79. Monolingual and bilingual electronic dictionaries on the Internet
 80. Large-scale documentary dictionaries on the Internet
 81. Electronic encyclopedias
 82. Combined products: Dictionary and corpus

- XV. Typology of Electronic Dictionaries II: Electronic Dictionaries for Machine Use**
 83. Electronic dictionaries for speech recognition, for its applications and for speech synthesis
 84. Electronic dictionaries for text processing and correction systems
 85. Electronic dictionaries for information retrieval, information extraction and text mining
 86. Electronic dictionaries for question answering and natural language database access
 87. Electronic dictionaries for computer-assisted language learning
 88. Electronic dictionaries for natural language understanding
 89. Electronic dictionaries for spoken and multimodal dialogue systems
 90. Electronic dictionaries for applications of natural language generation
 91. Electronic dictionaries for machine translation and machine-aided translation
 92. Generic electronic dictionaries for multiple uses in NLP
- XVI. Models for the Representation of Dictionaries: The Form Aspect**
 93. Representing human and machine dictionaries in markup languages (SGML, XML)
 94. Representing computational dictionaries in relational databases
 95. Representing computational dictionaries in feature structure-based representation formalisms and typed feature logic
 96. Representing computational dictionaries in AI-oriented knowledge representation formalisms: Dictionaries within expert systems
 97. Representing computational dictionaries in hypertextual form
 98. Representing computational dictionaries in dedicated formalisms: DATR, Word-Manager
 99. Standardization of the formal representation of lexical information for NLP
- XVII. Models for the Representation of Linguistic Data in Electronic Dictionaries: The Content Aspect**
 100. Coding pronunciation in computational dictionaries
 101. Coding inflectional morphology in computational dictionaries
 102. Coding word formation morphology in computational dictionaries
 103. Coding syntactic properties of words in computational dictionaries
 104. Coding semantic properties of words in computational dictionaries
 105. Computational dictionaries relating semantic and syntax: the FrameNet approach
 106. Coding pragmatic properties of words in computational dictionaries
 107. Coding idiomatic expressions in computational dictionaries
 108. Coding of multimodal dictionaries
 109. Standardization of the coding of monolingual linguistic information in computational dictionaries
 110. Standardization of the coding of bilingual and multilingual linguistic information in computational dictionaries
 111. Typologization of metadata of electronic dictionaries

XVIII. Models for Non-alphabetical Computational Dictionaries

- 112. Representing world knowledge: CYC
- 113. WordNet: Principles, developments and applications
- 114. Ontologies for natural language processing
- 115. Ontologies for the Semantic Web: Principles, formal considerations, resource building, tools and applications

XIX. Computer-based Dictionary Making I: Acquisition of Lexical Data from Corpora — Corpus Design

- 116. Criteria for the design of corpora for lexicography I: Monolingual dictionaries
- 117. Criteria for the design of corpora for lexicography II: Bilingual dictionaries
- 118. Large corpora designed for lexicographic work
- 119. The World Wide Web as a resource for lexicography

XX. Computer-based Dictionary Making II: Acquisition of Lexical Data from Corpora and Machine-readable Dictionaries — Tools and Procedures

- 120. Corpus pre-processing for lexicography
- 121. Tools to support the design of a macrostructure (nomenclature)
- 122. Concordance tools for lexicography
- 123. Tools and procedures for the acquisition of morphological and syntactic information from corpora
- 124. Tools for the acquisition of significant lexical combinatorics
- 125. Tools for the lexicographic use of parallel and comparable corpora
- 126. Analyzing machine-readable dictionaries — Reformatting and reinterpretation
- 127. Reversal of bilingual dictionaries
- 128. Retrodigitization: From paper dictionaries to electronic dictionaries

XXI. Computer-based Dictionary Making III: Computational Support for the Writing of Dictionary Articles

- 129. Tools for dictionary writing and consistency control
- 130. Workbenches for corpus-based lexicography
- 131. Tools for dictionary updating: Comparing dictionary and corpus data

XXII. Computational Terminography

- 132. Models for the representation of terminological data on the computer I: Terminological databases
- 133. Models for the representation of terminological data on the computer II: Terminological exchange formats
- 134. Large terminological databases
- 135. Acquisition of terminological data from text I: Approaches
- 136. Acquisition of terminological data from text II: Term variation, concept relations, etc.