

Edited by
Pedro A. Fuertes-Olivera
Henning Bergenholtz

e-Lexicography

The Internet, Digital Initiatives
and Lexicography

e-Lexicography

Also available from Continuum

Academic Writing, Diane Pecorari, Susan Hunston, Maggie Charles

Definition in Theory and Practice: Language, Lexicography and the Law, Roy Harris and
Christopher Hutton

e-Lexicography

The Internet, Digital Initiatives and Lexicography

Edited by
Pedro A. Fuertes-Olivera and
Henning Bergenholtz



Continuum International Publishing Group

The Tower Building
11 York Road
London SE1 7NX

80 Maiden Lane
Suite 704
New York NY 10038

www.continuumbooks.com

© Pedro A. Fuertes-Olivera, Henning Bergenholtz and contributors 2011

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without prior permission in writing from the publishers.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 978-1-4411-2806-5

Library of Congress Cataloguing-in-Publication Data

A catalog record for this book is available from the Library of Congress.

Typeset by Newgen Imaging Systems Pvt Ltd, Chennai, India
Printed and bound in Great Britain

Contents

<i>Preface</i>	vii
<i>Notes on Contributors</i>	ix
Introduction: The Construction of Internet Dictionaries	1
<i>Pedro A. Fuertes-Olivera and Henning Bergenholtz</i>	
1 Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries	17
<i>Rufus H. Gouws</i>	
2 Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries	30
<i>Henning Bergenholtz</i>	
3 Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction	54
<i>Sven Tarp</i>	
4 Filtering and Adapting Data and Information in an Online Environment in Response to User Needs	71
<i>Theo JD Bothma</i>	
5 A Multi-Layer Architecture for 'Pluri-Monofunctional' Dictionaries	103
<i>Dennis Spohr</i>	
6 Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools	121
<i>Patrick Leroyer</i>	
7 From Data to Dictionary	141
<i>Sandro Nielsen and Richard Almind</i>	
8 Internet Dictionaries for Communicative and Cognitive Functions: <i>El Diccionario Inglés-Español de Contabilidad</i>	168
<i>Pedro A. Fuertes-Olivera and Marta Niño-Amo</i>	
9 A Dictionary Is a Tool, a Good Dictionary Is a Monofunctional Tool	187
<i>Henning Bergenholtz and Inger Bergenholtz</i>	

10	The Technical Realization of Three Monofunctional Phrasal Verb Dictionaries <i>Birger Andersen and Richard Almind</i>	208
11	Online Dictionaries of English <i>Robert Lew</i>	230
12	e-Dictionaries in the Information Age: The Lexical Constellation Model (LCM) and the Definitional Construct <i>Aquilino Sánchez and Pascual Cantos</i>	251
13	Modelling Interactive Reading, Translating and Writing Assistants <i>Serge Verlinde</i>	275
14	Electronic Dictionaries as Tools: Toward an Assessment of Usability <i>Ulrich Heid</i>	287
15	Conclusions: Ten Key Issues in e-Lexicography for the Future <i>Eva Samaniego Fernández and Beatriz Pérez Cabello de Alba</i>	305
	<i>References</i>	312
	<i>Index</i>	332

Preface

On 17 June 2010, the University of Valladolid conferred the degree of Doctor of Lexicography, *Honoris Causa*, on Henning Bergenholtz. The governing bodies of the university made the decision on two grounds, both of which were stated in the *Laudatio* read at the graduation ceremony. First, Dr. Bergenholtz has been a professional lexicographer for more than 30 years. During these years, he and his colleagues from the Centre for Lexicography, Aarhus School of Business, University of Aarhus, have developed and set in motion a theory to approach dictionary-making and reviewing in a way that is very different from the linguistic colonialism espoused by linguists and the Wiegand model that is based on a contemplative approach to lexicography. The new model, which is called the *function theory of lexicography*, has allowed Dr. Bergenholtz and his colleagues to reshape the field of lexicography, changing the field completely. Secondly, Dr. Bergenholtz has collaborated and is collaborating with the University of Valladolid in a very productive and scholarly way. For example, he helped us to organize an international symposium on *e-lexicography*, which was held at the University of Valladolid (14–16 June 2010), and which can be considered a landmark in the field, as participants discussed innovative and workable proposals.

Some of the proposals presented have been included in this book, which has been prepared and edited according to accepted academic standards: all the contributions were subject to a peer-review process and discussed by well-known international scholars in the field.

The above claims allow me to thank the following:

- Dr. Bergenholtz for his help in preparing the academic programme.
- The participants in the symposium for bringing to the discussion new and provocative ideas. My special thanks go to the authors who have contributed in this book and to several colleagues who also participated and collaborated to create a friendly and relaxed atmosphere: María José Crespo; Klara Ceberio; Sahat Ugartetxea; Mercedes Jaime; Rocío Jiménez Briones; Ricardo Mairal; Ángel de los Ríos; Pablo Gordo; Jacek Lesinski; Francisco Ruiz de Mendoza; Sol Sta. María; Bernadette Borosi; and Ángeles Sastre.
- The reviewers of the different articles for their insights and thoughtful comments.

- The funding authorities, which provided us with the funds for carrying out the event:
 - Ministerio de Ciencia e Innovación (Grants. FFI2008–01703/FILO, and FFI2009–07109-E/FILO)
 - Junta de Castilla y León (Grant VA039A09, and BOCYL, de 12 de abril de 2010).
 - Universidad de Valladolid (Ayudas del Vicerrectorado de Investigación).

Pedro A. Fuertes-Olivera
Valladolid, 29 October 2010

Notes on Contributors

Richard Almind is a researcher at the Centre for Lexicography, Aarhus School of Business, University of Aarhus, Denmark. His main task is connected with designing lexicographical databases and preparing software for producing both printed and internet dictionaries. He has published extensively on dictionary layouts, lexicographical databases and several aspects connected with the distinction between information tools and information databases. He is currently working on the design of the database for the *Spanish Accounting Dictionary* and other internet dictionaries that are being planned and compiled at the Centre for Lexicography. He can be contacted at rab@asb.dk.

Birger Andersen is Associate Professor at the Centre for Lexicography, Aarhus School of Business, University of Aarhus, Denmark. His research interests cover areas such as grammatical information in dictionaries, particularly dictionary grammars, the functional rationale for various types of data in dictionaries, and specialized bilingual lexicography. His current research centers on devising and creating a database for the generation of three electronic English–Danish phrasal-verb dictionaries for reception, translation and production, respectively. He can be contacted at ba@asb.dk.

Henning Bergenholtz is Professor of Lexicography at the Aarhus School of Business, University of Aarhus, Denmark, where he is also Director of the Centre for Lexicography. Since 2005, he has been Extraordinary Professor at the University of Stellenbosh (South Africa). In 2010, the University of Valladolid conferred the degree of Doctor of Lexicography, *Honoris Causa*, upon him. Dr. Bergenholtz has published on grammar, lexicography and language policy, and has been the editor of more than 30 dictionaries (for example, the *Dictionary of Fixed Expressions*). He has published extensively within the field of lexicography (for example, the *Manual of Specialised Lexicography*), and more than 200 papers. He can be contacted at hb@asb.dk.

Inger Bergenholtz has been a piano teacher at Kolding Music School in Denmark since 1988. She graduated in musicology at the University of Copenhagen and afterwards took a degree as a pianist at the Hochschule für Musik in Berlin. After teaching the piano at music schools in Germany, she

went to Denmark in 1987 and worked as a piano teacher and as a chamber musician. She published a print dictionary of music, *Politikens Musikordbog*, in 1996 and she is the author of papers in lexicographical journals, for example, *LexicoNordica* and *Lexikos*. In 2006, she began editing an internet dictionary of music, first www.musikordbogen.dk, now *Musikordbogen* on www.ordbogen.com. She can be contacted at inger@bergenholtz.dk.

Theo Bothma is Professor and head of the Department of Information Science at the University of Pretoria, Pretoria, South Africa, and chairperson of the School of Information Technology (2008–2012). His teaching and research focus on information organization and retrieval (including information literacy), web development and electronic publishing, as well as on curriculum development. He has published widely and presented numerous papers at local and international conferences. He was the editor of the *IFLA FAIFE World Report 2007* and *IFLA World Report 2010* (an interactive web-based publication). He is a member of the editorial boards of *Libri*, *SA Journal of Libraries and Information Science*, *International Journal for Information Ethics*, *Online Information Review*, *New Review of Information Networking* and *Education for Information*. He is member of LIASA, ACM, CSSA, as well as the IFLA Standing Committee for Knowledge Management and an expert resource person for IFLA FAIFE. He can be contacted at theo.bothma@up.ac.za.

Pascual Cantos is a Full Professor at the Department of English at the University of Murcia (Spain). He received his BA (1988) and PhD (1993) degrees in English Language and Literature from the University of Murcia (Spain), his MA in Computational Linguistics (1995) from Essex University (UK) and a Postgraduate Diploma in Multivariate Statistics (2000) from the UNED (Spain). His main research interests are in Corpus Linguistics, Quantitative Linguistics, Computational Lexicography and Computer Assisted Language Learning. He is a member of the LACELL (Lingüística Aplicada a la Computación, Enseñanza de Lenguas y Lexicografía) Research Group at the University of Murcia.

Pedro A. Fuertes-Olivera is Associate Professor at the University of Valladolid, where he was Vicerrector de Economía (2003–2006). He obtained his accreditation for Full Professor in 2011. He teaches specialized discourse, especially ESP. His research interest lies in specialized lexicography, translation, and language teaching. He has recently published several books and papers on lexicography, and is currently working in the planning and compilation of several dictionaries within the project known as the *Accounting Dictionaries*. He is a member of the editorial board of *Hermes*, *Journal of Language and Communication Studies*, *Revista Española de Lingüística Aplicada*, and *Odisea*. He can be contacted at pedro@tita.emp.uva.es.

Rufus H. Gouws is Professor in Afrikaans Linguistics and Chair of the Department of Afrikaans and Dutch at the University of Stellenbosch in South Africa, where he also coordinates the Postgraduate Programme for Lexicography. His research focuses primarily on theoretical lexicography and he complements his theoretical work with contributions in the lexicographic practice as editor and co-editor of various Afrikaans dictionaries. He is a member of the editorial board of the *Fachwörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* and of the journal *Lexicos* as well as co-editor of the journal *Lexicographica* and the book series *Lexicographica Series Maior*. He has published extensively in the field of lexicography and is a regular participant in international conferences and symposia. He is a former president of AFRILEX, the African Association for Lexicography. He can be contacted at rhg@sun.ac.za.

Ulrich Heid is a computational linguist and professor of language technology and computational linguistics at the University of Hildesheim, Germany. He did studies of romance linguistics and history and earned a PhD in computational linguistics at Stuttgart (1995); at the same place, he got his habilitation in 2001. Heid worked at University of Stuttgart since 1986, and at the University of Hildesheim since 2008; he has been involved in projects centred on dictionary structures (one in the Framework of the Stellenbosch Institute for Advanced Study), on corpus lexicography and on corpus linguistics. Since 2010, he has been involved in a basic research project on the treatment of ambiguities in corpus linguistic data extraction, a European project on terminology extraction and a national project on infrastructures for language resources.

Patrick Leroyer, PhD, is Associate Professor of French and Danish Specialized Translation and Business Communication at the Aarhus School of Business, University of Aarhus, Denmark, and is attached to the Centre for Lexicography. He is also Editor-in-Chief of *Hermes, Journal of Language and Communication Studies*. His current research is on theoretical lexicography in a broad, functional perspective, and includes the development of specific theories for data-access, -selection and -presentation in connection with communicative, cognitive, operative and interpretive functions in paper-based and e-lexicographic information tools. He has published within the fields of Specialized Translation and Business Communication, Terminography and Functional Lexicography. Patrick Leroyer can be contacted at pl@asb.dk

Robert Lew is Professor at the Department of Lexicology and Lexicography, Adam Mickiewicz University, Poznań, Poland. His current interests centre around dictionary use, and he is involved in a number of research projects, including topics such as access-facilitating devices, definition formats,

dictionaries for production, space in dictionaries and training in dictionary skills. He has worked as a practical lexicographer for various publishers, including Harper-Collins, Pearson-Longman and Cambridge University Press. He is Reviews Editor for the *International Journal of Lexicography* (Oxford University Press) and advisor to Macmillan Dictionaries. He can be contacted at rlew@amu.edu.pl.

Sandro Nielsen is affiliated with the Centre for Lexicography – Research into Needs-Adapted Information and Data Access, Aarhus School of Business, Aarhus University, Denmark, where he is Associate Professor. He has an MA in English (LSP for translators and interpreters) from 1987 and was awarded his PhD degree in specialized lexicography in 1992. He is the author and co-author of numerous publications on theoretical and practical lexicography, including *The Bilingual LSP Dictionary: Principles and Practice for Legal Language* (1994), a printed and an online bilingual law dictionary; three printed and five online accounting dictionaries; and is a major contributor to the *Manual of Specialised Lexicography*. His main research areas are principles for online LSP dictionaries, user guides in dictionaries, lexicographic information costs and academic dictionary reviewing. Teaching interests focus on lexicography and legal translation for translators and interpreters. He can be contacted at sn@asb.dk.

Marta Niño-Amo is Associate Professor at the University of Valladolid (Spain), where she teaching accounting. She has been working within the project of the *Accounting Dictionaries* and is currently the editor of *Memento Contable*, a respected Spanish publication featuring comments on accounting standards and legislation, with the aim of offering up-to-date information on accounting matters. She can be contacted at mna@tita.emp.uva.es.

Beatriz Pérez Cabello de Alba graduated in English studies from the University of Granada, Spain, and holds a PhD in Linguistics from the University of Córdoba. She has been a visiting scholar in the University of Amsterdam (1993) and the University of Verona (1994) and has lectured in Chulalongkorn University in Bangkok (1997–1998) and in the London School of Economics and Kingston University (1998–1999). In 2000 she worked for Bitext, one of the pioneering language solutions companies in Spain. In 2001, she joined the National University of Distance Learning in Spain, where she has been working since. She has collaborated in several research projects and carried out research on lexicology, lexicography and language technologies. She can be contacted at bperez-cabello@flog.uned.es.

Eva Samaniego Fernández is Associate Professor at UNED (Spanish Distance University), where she primarily teaches English for Specific Purposes. She is also a sworn translator and holds an MA in specialized translation. Her

research interests are translation, ESP, specialized lexicography and metaphor. She has published within the fields of metaphor translation, legal discourse and translation, translation-applied text analysis, translation and genre, etc. She also teaches English language courses for European Union judges and prosecutors within the framework of the Spanish Council of the Judiciary and Eurojust, and is an appointed linguist of the European Judicial Training Network within the project “Language Training on the Vocabulary of Judicial Cooperation in Criminal Matters”. She can be contacted at esamaniego@flog.uned.es.

Aquilino Sánchez studied at the University of Barcelona, Spain, and later at the University of Munich (*Germanistik*), Germany, and at Georgetown University, Washington, D.C. (*Applied Linguistics*). He began his university career at the University of Barcelona and three years later joined the Modern Languages Department at the Autonomous University of Barcelona, Bellaterra. He was also Director of the Official School of Languages in Barcelona for five years. In 1980, he obtained a Chair in English Philology at the University of Málaga and in 1981 he moved to the University of Murcia, where he has taught since. His research interests and teaching centre on Lexicology and Lexicography, Second Language Teaching Methodology and Corpus Linguistics. He has edited the *Gran Diccionario de uso del español actual* – 2001, SGEL – based on the corpus *Cumbre*, and also a bilingual *Spanish–English* dictionary (1993, SGEL). He has lectured in various universities in Spain and abroad and has published many books and articles on these subjects. He can be contacted at asanchez@um.es

Dennis Spohr studied computational linguistics at Universität Stuttgart and Dublin City University, with a focus on computational lexicography and semantic web technologies. In 2010, he completed his doctoral dissertation on multifunctional lexicon models under the supervision of Prof. Dr. Ulrich Heid at Universität Stuttgart. Shortly afterward, he joined the Semantic Computing group at the Center of Excellence Cognitive Interaction Technology (CITEC) in Bielefeld, where he is currently working on multilingual ontology localization and cross-language knowledge access and presentation using semantic web technologies. Prior to his appointment in Bielefeld, he had been involved in projects on lexical semantics and computational lexicography in Saarbrücken and Stuttgart.

Sven Tarp is Professor of Lexicography at the Aarhus School of Business, and affiliated with the Centre for Lexicography. Since 2008, he has been Extraordinary Professor at the University of Stellenbosch (South Africa). He is author and co-author of several publications on lexicography, including *Lexicography in the Borderland between Knowledge and Non-Knowledge: General*

Lexicographical Theory with Particular Focus on Learner's Lexicography. He can be contacted at st@asb.dk.

Serge Verlinde is Professor of French for Specific Purposes and Director of the *Leuven Language Institute* at the University of Leuven, Belgium. His main research interests are corpus linguistics, pedagogical lexicography and Computer Assisted Language Learning (CALL). He is co-author of the *Dictionnaire d'apprentissage du français des affaires*. Recently, he developed web applications for teaching and learning French vocabulary: the *Base lexicale du français* and *Alfalex*. He is currently working on online reading, translating and writing aids.

Introduction: The Construction of Internet Dictionaries

Pedro A. Fuertes-Olivera

Henning Bergenholtz

Many, perhaps a large majority of e-dictionaries, were not *e-dictionaries* in their own way, but mere printed dictionaries (*p-dictionaries*) made available on an electronic platform. Still, we can observe that only a few existing e-dictionaries really use the technical possibilities of the electronic medium in the conception and preparation of dictionaries, and in the access to and presentation of data in them. The practical explanation for this is simply that most lexicographers do not see the consequences of the difference between a lexicographical database and a dictionary and, therefore, continue the tradition of planning and compiling polyfunctional e-dictionaries, which are directly taken from or made similar to p-dictionaries.

The theoretical explanation for the above situation is that some lexicographers, for example H. E. Wiegand (1998), assume the necessity of new theories to construct e-dictionaries, since they consider that former theories of lexicography are only usable for p-dictionaries. Our view is different, not only because it will force us to accept a two-string theory of lexicography – one for p-lexicography, and another for e-lexicography – but also because we are convinced that what we need is the same theory(ies), although it is adapted to the different access to and data presentation possibilities of the two media. This is the main topic of this book.

All the articles in this book, therefore, explore the state of art in e-lexicography by studying how new lexicographical concepts, and their application in specific internet dictionaries, are expected to shape lexicographical innovations in the near future. Electronic dictionaries are typical products of what we call knowledge and information society, which demand a different approach to electronic lexicography from the one which started at the beginning of the electronic-dictionary age. For instance, only internet dictionaries, but not other types of electronic ones, assure quick and easy access to extra-lexicographical data. Consequently, old typologies of electronic dictionaries must be substituted by more informed ones, which can take into consideration the difference between an information database and an information tool, the

access process used and their integration within pedagogical environments; they should also explore the introduction of Boolean searches, and allow for maximizing and minimizing searches, to name just a few of the lexicographic characteristics that are being dealt with in this book, some of which are summarized in Gouws's chapter, and reproduced below:

- The use of data banks from which different types of dictionaries, and even different dictionaries of the same type, can be extracted.
- The mistake of including much more data than needed, a possibility that is of the utmost interest for e-lexicography, considering that e-lexicography is not hampered by space restrictions.
- The broadening of lexicographical theory to the development, planning, compilation and publication of other reference sources, which are also focused on the users of these sources, the data presented in them, the structures to accommodate the data and, ever so important, access to the data in order to achieve an optimal retrieval of information.
- A paradigm shift, which is also applicable to printed dictionaries, takes into consideration the fact that dictionary users are also internet users who are used to downloading and uploading all types of data. Uploading offers lexicographers the opportunity to enhance a spirit of lexicographic democratization.

These are central issues in the articles included in this manuscript, which is divided into two parts; one being more focused on general theoretical questions and their translation into specific dictionary projects, while the other being more concerned with presenting specific dictionary projects which have recently been out of the lexicographical drawing board with the aim of illustrating what a user-driven lexicography is about. This distinction is rather artificial and we, therefore, request the attention of our readers to view the book as a unified theoretical and practical attempt to discuss where we are and where we can be in the near future.

Part 1 Chapters 1–6 are specially concerned with the tenets of function theory, the theoretical construction initially developed by Bergenholtz and Tarp (2002, 2003, 2004, 2005a; see Tarp 2008a, and Tono 2010 for a review), which has also opened new ways in the construction of internet dictionaries, with the aim of facilitating the exact data users need in a quick and easy way. The main idea behind addresses the question as to how lexicographic data can be modelled in such a way that a lexicographic tool is capable of satisfying the different types of users in different types of situations. As several authors point out in this book, lexicographers cannot aim to satisfy the needs of each individual user. Therefore, 'what is needed is not only a lexicographical tool that is capable of dealing with *types* of users and situations, but one that provides the necessary mechanisms for *individualization* of dictionary content – in terms

of customizing the views that an individual user is given on the lexicographic data' (Spohr, this volume).

One of the solutions proposed in this book is to move beyond the term 'dictionary' and introduce the term 'information tool' as a kind of umbrella term with which researchers can design any tool, no matter what we call them, aiming to satisfy the needs users might have in the four use-situations described so far: communicative, cognitive, operative and interpretive. Furthermore, researchers within this theoretical framework are also claiming the necessity for a shift in paradigm, which will place lexicography in the realm of information science, and will focus on describing the defining elements of internet dictionaries in terms of their accessibility, their formal properties and their categorization as information tools with which the distinction between information tool and information database, which is central in many of the dictionary projects here described, is more easily understood.

In addition, each chapter in the book analyses more specific proposals and considerations in depth. They are framed within the general frameworks we have referred to in the previous paragraphs, and, hence, they add to the unit of purpose of this book. Below, we include summarize some of the main ideas of each chapter. By so doing, we are confident that readers of this book can really get a gist of the hotly debated main issues concerning e-lexicography in international lexicographical circles.

In Chapter 1, 'Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries', Gouws states that the future of e-lexicography 'should not be isolated from either the past or the present', and, therefore, states that e-lexicography should take what we have learnt from the past and move to the future guided by innovation and intelligent boldness. For example, the distinction between a contemplative and a transformative approach, which has been already applied when working with printed lexicography, is also useful for e-lexicography.

The above reflection translates into a number of issues that we have to learn and unlearn in connection with the dominant views in the three stages in which we can group the history of theoretical lexicography: an initial stage dominated by the language contents of the dictionary, and which has resulted in a kind of *linguistic colonialism* which is still much appreciated and defended in some areas; a second stage that is mostly concerned with dictionary structures, thus following Wiegand's main theses on the conceptualization of lexicography as an independent discipline and of the dictionary as a text in itself; and a third stage, which follows from Bergenholtz and Tarp's functional approach, which is centred on lexicographic workings and their interest in putting the dictionary user and the situation of use at the centre of the discussion.

Gouws claims that the advent of e-lexicography has not only made the fallacy of linguistic colonialism and the inappropriateness of Wiegand's stance evident, but has also emphasized a number of issues that we have to unlearn

with a view to understanding and explaining lexicography as an information science, and to presenting a general theory that is independent of the medium in which the dictionary is written.

Bergenholtz follows suit in Chapter 2, 'Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries', in which he claims that theoretical lexicography has not had a significant effect on practical lexicography so far, and adds that both theoretical and practical lexicography have had several important errors and misunderstandings, among which he highlights the linguistic colonialism of lexicography and the lack of real attempts to develop new presentation forms and access options. In particular, he criticizes the conclusions drawn from dictionary surveys, which are regarded as unscientific and, therefore, inadequate to advance in our understanding of theoretical lexicography, and mentions that log files show that all the discussion on access structure is of little importance, since each respondent in each individual case chooses a individualized search path. As we cannot describe an access structure for each individual, he proposes looking at the selected individual search path and see which dictionaries – with which macro and micro structures, different search-relevant graphic markers in printed dictionaries and research sequences in certain fields in the database of an electronic dictionary – are associated with a particular fast or particularly slow access.

Bergenholtz's proposal is investigated in two detailed case studies in which several questions are formulated in order to assess whether the users found help in an array of printed and internet dictionaries, and the search time spent consulting the dictionaries. With these two case studies, Bergenholtz exemplifies two important criteria when evaluating the use and quality of a dictionary, which are whether the user could find the item that contains the answer to the question that prompted the search, and how long the search took. For him, the best dictionary is probably the one rendering a usable result in a short period of time, for example the four dictionaries of fixed expressions that he describes, that are extracted from the same database, each with the aim of being helpful in searches adapted to the function demanded by the user and to the use situation in which the user performs his or her search. Among them he mentions several search options that mark a novelty in electronic lexicography, and signal the way ahead for internet dictionaries, which will be connected to the retrieval of the data the user needs in a specific use situation and no more. For example, users have the possibility of finding an expression with a meaning similar to the one just found, and have the option of finding an expression with a particular meaning in use situations in which they might not know the fixed expression or cannot exactly remember the idiom or saying it as it is actually known. Bergenholtz also discusses the default searches in the database and details the data presentations in agreement with the tenets of function theory.

Some of the theoretical issues raised by Bergenholtz are also present in Chapters 3, 4, 5 and 6, written by Tarp, Bothma, Spohr and Leroyer,

respectively. As also mentioned by Gouws, Tarp defends in 'Lexicographical and Other e-Tools for Consultation Purposes: Toward the Individualization of Needs Satisfaction' his belief that there is no need for a new theory at the highest level of abstraction. However, he adds that every discipline must cope with the epistemological process of acquiring new knowledge and, consequently, any general theory about any subject field must constantly be improved, and sometimes replaced. This is the current situation in lexicography where we are witnesses to a paradigm shift with which lexicographers are paying attention to elements that existed earlier but were 'hidden' or unnoticed, and of completely new elements which exist and are related to the new media and technologies.

The combination of both types of elements has given rise to the consideration of lexicography as a consultation discipline integrated into information science, which initially demands a true definition of two concepts: *e-lexicography* and *lexicographical e-tools*. Both are frequently used to refer to any reference work made available on an electronic platform. Tarp, however, believes that both concepts have to be understood in a narrower way and, hence, restricts his classification to what he calls Model T Fords, and Rolls Royces, which are lexicographical works that have been (or will be) constructed with the aim of offering dynamic articles with dynamic data that correspond to the specific types of information needs which specific types of users performing specific types of lexicographically relevant activities might have in any consultation situation. He adds that Model T Fords are already in use, for example the *Accounting Dictionaries* or the *Danish Music Dictionary*, whereas Rolls Royces are still in the drawing board – the main difference being the fact that the former allows access to the data selected in a prepared database with browsing on the internet – but do not allow a recreation and re-representation of the data made available in this way. In his contribution to this volume, Verlinde, however, contradicts Tarp's view by indicating that Rolls Royces might have already come into existence and signals his *Interactive Language Toolbox* – we will present it in the second part of this introduction – as an exemplar of a lexicographical Rolls Royce.

Finally, Tarp claims that a theory like the function theory cannot be built directly upon concrete and individual phenomena that might differ from each other in many aspects, but from an abstraction with which we can work by referring to types of users, types of user situations, types of user needs and types of data that might satisfy these needs. Within this general framework, however, each user, user situation, user need, data and consultation is an individual act and, therefore, the individualization of user-needs satisfaction is a question to be taken seriously, especially because the internet allows lexicographers to provide the necessary mechanisms for an *individualization* of dictionary content.

Tarp describes three methods to achieve individualization: the interactive method, which will allow users to be assisted in making a personal profile as

well as indicate the specific type of situation or activity where information needs and, even the specific type of need, occur; the active method, which will allow users to design their own 'master article' in terms of the types of data wanted and their arrangement on the screen; and the passive method, which consists of automatic tracking of the users' behaviour during a number of consultations, thus facilitating the creation of a profile of the type of data that the users generally look for in order to furnish the same type of data when the users once more consult the e-tool.

The individualization of need is having another important consequence: it is blurring the same concept of dictionary typology as in lexicographical e-tools conceived and, according to the principles of individualization, there are neither monofunctional nor multifunctional data access routes, but only individualized ones that translate in lexicographical e-tools viewed as one multifunctional dictionary with individualized search options within the framework of its defined functions. As a consequence, the best dictionary in terms of needs satisfaction is not necessarily a monofunctional dictionary, but any dictionary – whether monofunctional, pluri-functional or multifunctional – that allows either monofunctional access or individualized access in the framework of its specific and foreseen functions.

Tarp's vision needs to be implemented by means of an array of devices that are described in Chapter 4. Bothma's 'Filtering and Adapting Data and Information in an Online Environment in Response to User Needs' addresses two related and very connected issues, which are not only crucial for the future of e-lexicography, but also for advancing in the development of user-based theoretical and practical lexicography. He investigates to what extent modern information technology can facilitate the design and implementation of e-dictionaries and/or e-information tools for specific user groups and situations and can enable the user to 'create' his/her own e-dictionary and/or information tool(s).

Both questions are answered by reviewing a number of information technologies and techniques, namely, searching and navigating, user profiling, filtering, adaptive hypermedia, metadata markup, linked open knowledge, recommended systems and annotation systems – all of which can be used in e-dictionaries to customize, that is, personalize information access in response to user needs. Bothma's views on the above information technologies and techniques are, on the one hand, encouraging, considering that most of them are already in use in specific e-dictionaries, and on the other hand, rather disappointing, as they are not used to their full potential. Hence, he concludes his chapter by formulating a wish: 'If lexicographers were to embrace these technologies, it would be possible to provide customized information tools that can satisfy the user needs of all individual users. It would therefore be possible to create information tools that would address the information needs not only of the 'average' user, but those of a specific user in 'one out of a thousand

consultations', by providing 'dictionaries capable of meeting all the users' needs in specific types of situations'. In such a customizable e-information tool,

- the user will be able to:
 - set up a complex profile indicating his/her preferences;
 - change the profile based on specific information needs for any given situation; and
 - drill down to the required level of complexity and/or detail in any given situation;
- the system will:
 - further adapt the profile based on the user's information behaviour; and
 - present information to the user based on the characteristics of such a profile;
- the database will require that:
 - the data be marked up through a complex metadata schema
 - to enable matching the characteristics of the user's profile with the characteristics of the data;
 - there be links to external data sources (linking open knowledge)
 - either through direct linking by the lexicographer; or
 - by on-the-fly searching of such external data sources
 - to enable the user to get additional information on demand;
- The system will also be able to:
 - make recommendations to the user based on his/her profile and expressed information need; and
 - allow the user to make private, group or public annotations to the database to
 - enhance the user's future use of the data; and help the lexicographer to keep the database more current and up-to-date. (Bothma, this volume)

Spohr's Chapter 5, for example, 'A Multi-Layer Architecture for "Pluri-monofunctional" Dictionaries', is one of the first attempts to translate the aforementioned theoretical considerations into practice. He reports on recent efforts to develop a model for a lexical resource that enables the definitions of function-based views on lexicographic content (i.e. he offers a summary of the main ideas discussed in the previous chapters), and provides an infrastructure that can be extended so that it allows for an individualization of access and presentation of such content.

The lexicographical data model he proposes argues for a hierarchical organization of the entities in a lexical database in order to express information on different levels of granularity, and highlights the benefits that the use

of semantic web standards like Resource Description Framework/ Resource Description Framework Schema (RDF/RDFS) and Ontology Web Language (OWL) has for the definition of the lexical resource model here described, which distinguishes between several types of lexical entities, ‘with *free* and *bound* units located at the highest level, and more specific subtypes below each of them’. Moreover, this structure is reinforced by the fact that the specifications of the user needs are not included directly in the database but on a separate layer, ‘in order to ensure modularity and thus extensibility’. For example, returning to the question of pluri-monofunctionality, the model deals with three specific questions related to the user’s needs: (i) in which language and vocabulary an indication should be presented; (ii) what should be presented and in which language(s); and (iii) which indications should be accessible for which users. Spohr answers these three questions by detailing the functioning of the model for the German user, and by commenting on a prototypical implementation of the proposed architecture, which includes a prototype of ‘a web-based electronic dictionary containing roughly 14,000 lexemes with 44,000 example sentences and almost 35,000 morpho-syntactic preferences’.

Finally, Leroyer’s Chapter 6, ‘Change of Paradigm: From Linguistics to Information Science and from Dictionaries to Lexicographic Information Tools’, is a kind of bridge between Chapters 1–6 and Chapters 7–14. On the one hand, he summarizes in a very direct and convincing way the flaws observed in linguistic approaches to lexicography – mostly those by Atkins and Rundell (2008), and Béjoint (2010) – which are still defending the only practical nature of lexicography and thus neglecting a theoretical basis to the practical activity of dictionary making. Hence, he formulates his claim for a shift of paradigm, which views lexicography as

an integrated part of the social and information science paradigm and refers to the interdisciplinary discipline concerned with the study, design and development of functional tools aimed solely at the gratification of human information needs and problems. The distinctive feature of lexicographic tools is the triangulation of three interrelated sets of social, logical and semi-otic parameters, corresponding respectively to the following dimensions of the tool: user, access and data.

Social parameters are in every single case determined by the systematic identification of the specific information problems, needs and profiles of the potential user of the information tool. The social parameters are decisive for both the functional genesis (communicative, cognitive, operative and interpretative functions) and the gratifying use of the information tool.

Semiotic parameters are in every single case determined by such data selection and presentation that ensures gratifying extraction of information in accordance with the specific problems, needs and profiles of the potential

user. Data are by nature semiotic and consist of verbal and non-verbal signs. The most frequently used symbols in data selection and presentation are words.

Logic parameters are in every single case determined by such structures, modes, indices, algorithms and computing technologies that ensure gratifying access to data in accordance with the specific problems, situations, needs and profiles of the intended user.

On the other hand, Leroyer presents four different models of lexicographically designed information tools that illustrate the change of paradigm and demonstrate how lexicography is currently moving towards the realm of information science:

- A patient e-dictionary, *Lexonco*, whose genuine purpose, which was to solve information problems for cancer patients and their families, was achieved by presenting a modular, functional configuration with three types of individualized access modes: the consultational access mode; the interactive and participative access mode; and the automated access mode.
- An e-dictionary of real property that is an e-lexicographic guide to French real estate aiming to solve the communicative and cognitive needs of Danes in two specific use situations: Danes without any command of French and Danes with some command of French.
- An e-lexicographic mobile tourist guide with data of two kinds: a conventional user-driven search and navigation mode; and a range of automated access modes.
- An e-lexicographic guide for scientific text production that aims to, first, build a complete and adaptive tool to provide assistance to scientific text production in English, and, secondly, to support the practical training of students of specialized translation.

In a word, Chapters 1–6 present some new arguments for considering lexicography an integral part of information science. Following suit, Chapters 7–10 reinforce this view by describing in detail how three internet dictionaries have been planned and compiled: the *Accounting Dictionaries* (Chapters 7 and 8), the *Danish Music dictionary* (Chapter 9) and a dictionary of *English Phrasal Verbs* (Chapter 10). They are a glaring example of the kind of e-tools that display ‘dynamic articles with dynamic data, which correspond to the specific types of information needs which specific types of users performing specific types of lexicographically relevant activities might have in any consultation situation’ (Tarp, this volume).

The *Accounting Dictionaries* actually consist of a set of two monolingual and three bilingual online dictionaries in the following languages: Danish, English

and Spanish. The theoretical basis underlying the project gives priority to lexicographic functions, that is, the help these dictionaries can give to users in specific types of situation where users require knowledge to resolve issues relating to accounting. In Chapter 7, 'From Data to Dictionary', Nielsen and Almind first present and discuss the practical, technical basis of the project, which is made up of two distinct components: database and dictionary; thus, they analyse the theoretical framework of the project (already summarized in Chapters 1–6 of this book), and finally show that the theoretical and practical bases are integral features of the project, and consequently, neither can stand alone.

Regarding the database of the *Accounting Dictionaries*, Nielsen and Almind trace the chartered and difficult history of the conversion of the database, from its origin as a two-tier system for Danish, into its current hub-and-wheel structure for Danish, English and Spanish. They also add that the *Accounting Dictionaries* are best described as a triadic construction in that the structure consists of three main components: a database; (a) dictionary website(s); a search engine. Describing the dictionaries in terms of the above triadic structure has a number of practical and theoretical implications. First, the database is the source of several dictionaries. Secondly, the dictionaries contain, or might contain, several independent components. Finally, the *Accounting Dictionaries* do not have macrostructures in the text–linguistic sense, but a 'data presentation structure' that 'is supported technically by an output device which arranges the data retrieved from the database according to type, and presents these data in a predetermined order depending on user needs as identified by the type of help sought'.

Regarding functions, intended users groups, and data selection, both Nielsen and Almind and Fuertes-Olivera and Niño-Amo, in Chapter 8, 'Internet Dictionaries for Communicative and Cognitive Functions: *El Diccionario Inglés-Español de Contabilidad*', explain that these dictionaries have two main types of function, for example, communicative and cognitive functions, which aim to satisfy the needs of three main user types: (i) translators and language staff; (ii) accounting experts and semi-experts; and (iii) students and laypersons interested in Danish, English and Spanish accounting matters. They add that the selection of data starts with making an external and internal subject classification. The external subject classification puts into focus accounting texts dealing with the rules for accounting as defined by the International Financial Reporting Standards (IFRSs), with which all companies listed on stock exchanges in the European Union must comply. The internal subject classification was originally limited to financial and management accounting. The preparation of the external and internal subject classification resulted in three language-specific electronic text corpora containing authentic texts such as International Financial Reporting Standards, national financial reporting and bookkeeping standards and statutes, and financial statements published by national and international companies.

In addition, Fuertes-Olivera and Niño-Amo add a new plane to their description of one of the *Accounting Dictionaries*, the *Diccionario Inglés-Español de Contabilidad*. They envisage the way ahead by exploring two possible developments that could contribute to the elaboration of more focused internet dictionaries at no extra cost: (i) to increase reliability by widening the use of hyperlinks in order to minimize or even eliminate the stress users have when working with texts that are in the forefront of new knowledge, a typical situation when translating, reading and producing specialized texts; and (ii) to write a *systematic introduction*.

Bergenholtz and Bergenholtz, and Andersen and Almind employ the same theoretical framework in Chapters 9 and 10, 'A Dictionary Is a Tool, a Good Dictionary Is a Monofunctional Tool', and 'The Technical Realization of Three Monofunctional Phrasal Verb Dictionaries', respectively.

Bergenholtz and Bergenholtz also discuss the characteristics of a good dictionary. Their answer, which is supported within the tenets of the function theory, is that dictionaries are tools and, hence, they 'must be designed for a specific and limited purpose'; that is, they must be monofunctional dictionaries, and be the result of lexicographical-based questions such as the following: access to the data in dictionaries; access time; dictionary structures; dictionary functions; dictionary users; use situations and so on. Furthermore, they illustrate the workings of some of the above issues relating to the *Danish Music Dictionary*, a set of three dictionaries, one for reception, one for knowledge and one polyfunctional, which are the result(s) users can extract from a single database whose technical possibilities are similar to the ones already described in Chapters 2 and 7.

Andersen and Almind, first, detail the data input of *a work in progress*: the construction of a lexicographical database with which users can access several dictionaries of English phrasal verbs that target Danish students and professional translators, to whom the expected dictionaries will offer explicit grammatical data and several linguistic labels, for example, diaphatic, diachronous, diatopical, diatextual labels and so on, as well as English collocations and Danish translations.

Secondly, they indicate that the database will be capable of 'generating three different dictionaries with three different functions': one for checking the meaning of English phrasal verbs, one for translating English phrasal verbs into Danish, and one for assisting users in producing grammatically and stylistically English phrasal verbs. As well as having the option of accessing the types of data in the individual articles in the database that are relevant for the satisfaction of the given need, users will also have the option of getting access to all types of data in the article in the database, 'since it may be consulted by users with other needs than the three communicative needs specified here'.

In addition, there are four more chapters, which are devoted to reviewing internet dictionaries for English and Spanish and to presenting the *Base Lexicale du Français*, and to offering a case study on evaluating the *usability* of

internet dictionaries. Although English and Spanish are two world languages, the level of development of e-lexicographical tools in both languages is so different that we are unsure about why such a situation really exists.

Lew's 'Online Dictionaries of English' (Chapter 11) presents an overview of the spectrum of available online English language dictionaries, and offers some general comments on a few selected key issues. In his overview, Lew indicates that English lexicography is influencing the workings of lexicographers in every corner of the world, and reports on prominent and representative exemplars of specific types of online dictionaries, which he classifies by making use of well-known criteria, for example, general vs. special purpose dictionary, and newer criteria related to the characteristics of the internet, for example, the distinction between institutional and collective, free vs. paid, and number of dictionaries retrieved when searching. The combination of both criteria results in a proposal of *ad hoc* categories, which contemplate the existence of General English Dictionaries, Learner's Dictionaries, User-involved Dictionaries, Diachronic (historical) Dictionaries, Subject-field Dictionaries, Dictionaries with Restricted Macrostructures, Dictionaries with Restricted Microstructures and Onomasiological Dictionaries.

Lew also calls the attention to a new generation of internet dictionaries that are especially designed to satisfy specific user's needs in specific use situations. For example, the main novelty of the *Louvain EAP Dictionary*, which is being developed as a dictionary for non-native writers, 'is that it is customizable in terms of field domain (business, medicine) and mother tongue (French, Dutch). As a consequence, usage notes and equivalents match the L1 of the user, and some of the examples are domain-specific'.

The second part of his chapter is devoted to analysing some issues with online dictionaries, which he considers to be relevant. For example, dictionary aggregators usually retrieve long and very similar entries, which results in 'highly unhelpful, many-times redundant, tortuous assemblages of disconnected lexicographic data'. Similarly, he comments on the 'step-wise approach to outer access', and claims that the option taken by *myCOBUILD.com*, namely, the partial entries which are listed, seems more adequate than other options. He also alerts against the view espoused by some researchers who confound dictionaries and databases, and exemplifies uses of corpus interfaces and wrappers that are similar to dictionaries. Although the latter are so sophisticated, Lew claims that 'there is not much hope that their popularity will extend much beyond a relatively small group of power users; the others will just increasingly Google for any answers, irrespective of the nature of the problem and I fear that this tendency presents a real threat to more specialized reference tools, including dictionaries'.

Sánchez and Cantos' 'e-Dictionaries in the Information Age: The Lexical Constellation Model (LCM) and the Definitional Construct' (Chapter 12) is

amazing in as much as it is a vivid example of a change in focus due to external circumstances. They should have presented a critical review of internet dictionaries of Spanish, but found that these are almost non-existent, and that the few they managed to find were not proper internet dictionaries but paper dictionaries that have been uploaded to the internet. Fortunately, Sánchez and Cantos were able to reorganize their contribution and develop an interesting proposal, one centred on the possibility of using the Lexical Constellation Model (they presented this model initially in Cantos & Sánchez, 2001) for producing better lexicographic definitions.

Sánchez and Cantos's main dissatisfaction with how many dictionaries deal with definitions is their conviction that words 'are typically defined as if isolated and somehow monolithic units, with only occasional reference to contextual elements (especially collocates) and with scarce information on the intricate web of semantic relationships that shape the units of meaning we call words, or the complex semantic relationships among words associated to specific lexical fields'. This view, which can be disputed by making reference to internet dictionaries in their own right, for example the *Accounting Dictionaries*, translates into constructing more precise definitions by signalling semantic attraction among words, especially in cases in which such an attraction was apparently unexpected (Cantos & Sánchez, 2001). It results from the clustering of specific semantic features that are perceived as units by the speakers, although this perception can vary when attending to personal and/or contextual factors. Hence, they propose definitions that are tailored to the user's needs, make use of every possible resource (words, pictures, photographs, videos, etc.), and result 'from a modular, and hierarchical approach increasing in structure, data and complexity, which goes hand in hand with the user's demand, user's access to data and user's access to (non)lexical tools'. Their proposal is illustrated with the Spanish word *mano* [Eng: hand], which is defined differently when attending to the condition of the user as a pupil, semi-expert or expert, and a list of resources – most of which are already in operation in internet dictionaries such as the *Danish Music Dictionary* and the *Accounting Dictionaries* – that constitute a complete novelty in Spanish lexicographical circles, which are still dominated by linguistic colonialism and the consideration that lexicography and lexicology are related disciplines.

Verlinde's 'Modelling Interactive Reading, Translation and Writing Assistants' (Chapter 13) presents the *Base Lexicale du Français*, which is hailed as a web-resource that has a task-oriented approach offering 'an alternative to the individualization of lexicographical e-tools by recreating and re-representing data on user profiles to optimize electronic dictionaries' (Tarp, this volume).

Verlinde initially expresses his doubts on the usability of user-involved, bottom-up, collaborative lexicography that lacks critical thinking, leads to confusion and makes certain the *more is less* paradox, with which scholars, for

example Schwartz (2004), refer to the traditional view espoused by linguistic-oriented lexicographers, who defend the aggregation of as much data as possible in a dictionary article, no matter its usefulness for the potential user. Thus, he formulates the question he addresses in his chapter, and which is concerned with the kind of user interface that will give as many users as possible access to the information that is relevant in a given context in a user-friendly way. His answer, as in many other recent works by Verlinde and colleagues (e.g. Verlinde, Leroyer & Binon 2010), is the *Base Lexicale du Français (BLF)*, which originated as a database for French only, but which ‘is being expanded to become a multilingual application which will be renamed *Interactive Language Toolbox (ILT)*’.

The interface of the *BLF* is similar to the one described for the *Danish Music Dictionary* and the *Accounting Dictionaries*, and hence considers the possible needs of a user in order to create various small, monofunctional dictionaries. He adds that on the homepage the user does not find a text box for entering his or her search string, but several possibilities that are forcing him or her to identify his or her consultation situation and needs (Tarp, 2008a). Furthermore, the current version of the *BLF* homepage not only offers access based on the user’s needs, but also on specific tasks performed by the user (typically, reading, translating or writing).

Verlinde also shows some advantages of the *BFL* in a teaching and learning environment. Regarding the *BLF* reading assistant, Verlinde compares the *BLF* with the *Alexandria* tool and advances the possibility of incorporating syntactic analyses of the sentences included in the assistant as well as a contextual translation for multi-word expressions. The *BLF* translation assistant has the ability to adapt to the topic of a text, and provides the most common word combinations for many academic words. The *BLF* writing assistant does not correct the submitted text, as spelling and grammar checkers do, but identifies ‘syntactic and lexical patterns which may contain errors’ Moreover, it also contains a tool, already available for Academic Dutch, which enables users to expand their vocabulary (words and word combinations), suggests (near) synonyms and hyperonyms as academic alternatives to general language, and displays specific word combinations for certain words in the text.

Verlinde finishes his contribution by commenting on two future developments: (i) to have access to rich databases; and (ii) to carry out precise analyses of submitted texts.

In ‘Electronic Dictionaries as Tools: Toward an Assessment of Usability’ (Chapter 14), Heid claims that if electronic dictionaries are to be understood as (software) tools, they should also be designed according to the principles applicable to software tools. One such principle is *usability*, a concept developed within information science with the aim of assessing the effectiveness and efficiency of the tool when used in a particular situation and for a particular task.

Heid reports on a set of usability tests applied to three electronic dictionaries: the ELDIT, *Elektronisches Lernerwörterbuch Deutsch–Italienisch* (Abel & Weber 2000; Abel & Campogianni 2005); the BLF, *Base Lexicale du Français* (Verlinde, this volume; Verlinde et al., 2010, etc.); and the OWID, *Online-Wortinformationssystem* (Müller-Spitzer, 2010). Heid's case study is precisely described, and hence it can serve as inspiration for future studies on, say, the quality of internet dictionaries. In sum, Heid not only illustrates the close connection between lexicography and information science, an idea already discussed in several chapters of this book, but also offers an interesting and different methodology than the one used by, say, Bergenholtz and colleagues, who have also reported on several methods for assessing empirically whether a user gets what he or she needs and how long this process takes. In particular, Heid's proposal is really provocative as it tries to relate the basic objectives of usability testing with the concerns of lexicographic theory, and to propose steps towards usability engineering of electronic dictionaries.

Finally, the book includes a summary (Chapter 15) – written by Samaniego Fernández and Pérez Cabello de Alba – of the key issues in e-lexicography that were hotly debated by the participants in the international symposium on e-lexicography. They follow Andersen and Nielsen (2009) and, consequently, raise a list of issues, some of which were initially discussed at an International 2008 Symposium hosted by the Centre of Lexicography at the University of Aarhus, and which are expected to generate controversy in the near future.

To sum up. The contributions in this book have the following demands in common: A good information tool should be easy to use, easy to learn to use and be able to provide a result in a short span of time.

Most of the contributions, except the one by Theo Bothma (Chapter 4), have focused on e-dictionaries, and have emphasized that the dictionary concepts described and analysed are broader than those typically commented on in relation to printed dictionaries. We can summarize them by indicating that a dictionary might have more than dictionary articles (for example, the *Musikordbogen* or *Danish Music Dictionary*), as dictionaries should be planned to satisfy the user's needs in use situations. Hence, a question arises: are we still calling dictionaries the e-tools of the future, especially when some of these are also including systematic introductions, grammar books and other components that are very far from traditional dictionary articles? Or if the question is asked in connection with Grefenstette's (1998) 'Will There Be Lexicographers in the Year 3000?'. Probably not. But our reasons for this answer are different from the ones typically espoused by Grefenstette and most British and American lexicographers, as we no longer see the future of lexicography connected more with linguistics than with many other disciplines. At the end of the day, the future of lexicography, and this is what this book is about, starts by planning, compiling and studying information tools, regardless of what we call them. We can call them dictionaries. And this is the name of a famous

and well-known information tool, an information tool that has the following specifications:

- it is small and handy; it is always available (outside, inside, on holiday, etc.); it provides answers to all questions; and it provides answers in a way that the user understands them easily; it gives the correct answer, but not a more detailed answer than the one that is necessary in order to solve the problem, for example, when its users, namely Huey, Dewey and Louise, do not know the directions, they must look for help to find the right way; when they do not know whether a particular dish is poisonous or not, they look for help; and when they need help in order to talk to a gorilla, or to understand an inscription carved in an ancient language, their ‘dictionary’ provides them with assistance.

This ‘dictionary’ is a very special one; it is a special information tool that is called a personal dictionary, that is, a tool that is only used by Huey, Dewey and Louise. Although we could call it an e-dictionary, it is in fact an intelligent computer, which is able to intuit what Huey, Dewey and Louise want to know and give them the right answers in a way that they can understand. It is a tool that can read their thoughts, and respond to them in a hand-held computer, which has a book jacket as protection (usually red, sometimes yellow). The answer comes up quickly; it is always true and contains all the relevant data – and nothing more.

Of course, we do not have such an e-tool and we will not have one in the future. But what we will see in the near future is that the often-quoted division between dictionaries, lexicons, encyclopaedias, handbooks, manuals and so on, will be replaced by different monofunctional tools for special user types with special needs. We think that many of these tools can be used without having to pay them, whereas some others will be more specialized and will be aiming at solving the specific needs of specialists who will pay for them provided that they are updated immediately. This book deals with both types.

Chapter 1

Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries

Rufus H. Gouws

1.1. Introduction

When talking about the future of e-lexicography, that future should not be isolated from either the past or the present. In this paper, I show how future lexicography, with regard to both electronic and printed dictionaries as well as dictionary research, can benefit from the past by being cognizant of positive and negative aspects of both types of dictionaries. Therefore, I will be taking a present- and past-based look at the future. Although many remarks in this paper, and many comments regarding the development of lexicographic theory, are not explicitly directed at electronic dictionaries, it is implied that electronic dictionaries fall within the scope of these remarks and within the scope of a general theory of lexicography.

We often make a well-motivated distinction between a *contemplative* and a *transformative* approach in lexicography. While the former focuses on an investigation of the prevailing situation, the latter opts for a development from a current situation to something new. It is a valid distinction indicating two approaches based on diverse final goals. Although when employing a contemplative approach one could gain from the experiences of existing dictionaries, such an approach too often lacks new ideas. On the other hand, a transformative approach works with innovative ideas but sometimes fails to fully recognize the value of already accomplished success. When planning electronic dictionaries, and when I am using the term ‘electronic dictionary’ I am referring to online/internet dictionaries and not, unless otherwise stated, to CD-ROM dictionaries, the obvious approach could be to argue that we are dealing with something totally new and, therefore, we need to follow an exclusively transformative approach, without any attention to what has already been achieved in the area of printed dictionaries. With regard to CD-ROM dictionaries, one has to accept the fact that many of them are only paper dictionaries

on electronic platforms and that no new theory is needed for them. Therefore, they will not be the target of this discussion. With reference to the occurrence of real electronic dictionaries, it could also be argued that we merely have a change in medium and that a dictionary remains a dictionary. Paul Kruger, a nineteenth-century South African statesman, once remarked that one should take from the past what is good and use it to build the future. This belief also applies to lexicography. Although the electronic medium has its own demands, possibilities and challenges, it does not imply that lexicographers compiling an electronic dictionary should categorically eschew the work done in the production of printed dictionaries. One can benefit from learning some things from printed dictionaries, and the lexicographic wheel should not be re-invented when embarking on the planning and compilation of electronic dictionaries. What should be learned from the past, and this applies to both printed and electronic dictionaries, is to conscientiously avoid similar traps and mistakes, especially in cases where what are now seen as mistakes were then regarded as the proper way of doing things. This is where the transformative approach with its acknowledgement of the past, but also with its future visions, plays an important role. In these new endeavours, we as lexicographers are still bound to make mistakes in the future, but we have to restrict ourselves to making only new mistakes. Therefore, the planning of new dictionaries should not be dominated by tradition, but *innovation* and *intelligent boldness* should constitute the guiding principles.

A good transformative approach should allow a reflexive component that partially overlaps with a contemplative approach. In this regard, it is important to be cognizant of an early remark by Zgusta (1971: 18) that lexicography 'is an activity in which tradition plays a great role', but he also states that 'things may change slightly if more interest is given to lexicographic theory and if new work procedures or ways of presentation are developed, tried out, and used' (19). He further continues 'The lexicographer's work is always creative, in a greater or a lesser degree, because he must always try to find new solutions to problems as yet unsolved' (20). This is the challenge faced by lexicographers working within both the printed and the electronic medium.

Many modern-day lexicographers are blessed or cursed with an overactive methodological memory that continuously reminds them of the ways in which they and their colleagues performed their earlier lexicographic endeavours. When employing a transformative approach in the planning of a new dictionary, such a methodological memory could be detrimental, and lexicographers might need to be compelled to adopt an unlearning mode (delete, delete, delete) to rid themselves of some of the traditions enforced by the methodological memory. Where lexicographers embarking on the planning of new, especially electronic, dictionaries are familiar with the traditions and practice of printed dictionaries, their new assignment might also demand the

unlearning of certain established habits which have no place in the electronic medium.

With its practical and theoretical components, lexicography can be regarded as a two-legged animal. The lexicographic practice resulting in printed dictionaries, irrespective of being produced on clay tablets, that is, the real hard-cover dictionaries, papyrus leaves or paper, developed in what can cautiously be called a pretheoretical era. However, this does not imply that there was no theory supporting some of these dictionaries. The history of the lexicographic practice clearly shows that some dictionaries had not been compiled in a haphazard way but according to a well-devised plan. In some cases, this plan had even been published, cf. Samuel Johnson's famous *The Plan of a Dictionary* (1747) that preceded the work on his *A Dictionary of the English Language* (Johnson 1755). Such a plan or model can be regarded as the theoretical framework for such a dictionary. Unfortunately, too many dictionaries did not bear evidence of such a plan or a clear theoretical underpinning. Metalexicography or dictionary research, representing the formal theoretical component of lexicography, is a relative latecomer to the playing field. Consequently, in contrast to many of their predecessors, modern-day dictionaries have the advantage of a sound theoretical basis and lexicographers have no extenuating circumstances for not relying on the available theoretical framework when compiling their dictionaries. This is of special significance to electronic dictionaries. Today's dictionaries must be better than their predecessors.

A little while ago, a university colleague who knows I am involved in lexicography put a valid question to me. He wanted to know whether all the research in theoretical lexicography has led to an improvement in the quality of dictionaries. I immediately responded in a positive way, perhaps too positively. But his question made me think quite a bit. We pride ourselves that lexicography is an independent discipline with dictionaries as its subject matter. If modern-day dictionaries, including electronic dictionaries, are not really regarded by their intended and loyal users as being better than their older counterparts, some serious questions must be raised regarding the relevance and future of our discipline.

In contrast to the early work on printed dictionaries, the practice of electronic dictionaries developed in an era where well-established theoretical frameworks are available. If electronic dictionaries do not utilize this advantage, then something is rotten in the state of Dictionopolis. However, it is important that the development of electronic dictionaries should not be isolated from that of printed dictionaries. New ideas formulated for new electronic dictionary projects will often also have applicability in printed dictionaries. Lexicographers involved in the planning of these electronic products should refrain from jealously guarding their new ideas in order to restrict their application only to this medium. These ideas should be made available to

the printed medium, as well, so that lexicography in general can benefit from these new developments.

1.2. Looking at the History of Theoretical Lexicography

The distinction between a contemplative and a transformative approach should not be regarded as only applicable to the planning of dictionaries. It can also be interpreted in a wider sense, where it applies to how one approaches the development of lexicographic theory. When revising, changing, adapting and improving lexicographic theory, our main focus should be on a transformative approach in order to reach new goals. For the future of electronic dictionaries, it is important that new theoretical horizons be identified and investigated. To do this, one should not totally eschew a contemplative approach, because a number of lessons are to be learned from the history of lexicographic theory.

A historical view of the development of theoretical lexicography, a contemplative approach, gives evidence of some distinct phases in this ongoing progress, but also of some dominant influences (see Gouws, 2004). This forms an important background to a transformative approach that looks at innovative ideas for further developments, and contains a reflexive component to make provision for the inclusion of positive and valid aspects from the current and even older theory to support new ideas, for example, many of the suggestions introduced in the theoretical approach of Ščerba (1940) that are still valid today.

Although there are a few exceptions, the introduction of mainstream lexicography has been as a subsection of linguistics. Some of the early publications in this regard were Chapman (1948), Doroszewski (1954), Garvin (1955), Országh (1962) and the significant Householder & Saporta (1967), a book in which the majority of contributions focused on problems of a linguistic nature. This approach was maintained in Ladislav Zgusta's famous and ground-breaking book, *Manual of Lexicography* (1971). In the foreword of this book, Zgusta states, 'we also hope that a coherent statement and discussion of lexicographic problems will help to clarify them, and to demonstrate the importance of their being conceived in the framework of the linguistic theory more effectively'. He starts the introduction to his book as follows, 'There can be no doubt that lexicography is a very difficult sphere of linguistic activity'. The *Manual of Lexicography* gave a brilliant introduction to various aspects of lexicography, but the discussion was presented within a strong linguistic framework, the then-prevailing way of looking at lexicography.

Zgusta's book had a big influence on the development of lexicographic theory in the 1970s. As a result, the main focus in earlier lexicographic research had been on the contents, especially the linguistic contents, of dictionaries, with emphasis on the different types of linguistic data presented in dictionary

articles. The linguistic bias is especially clear in the research regarding the treatment of meaning in dictionaries. Following the clear distinction between semantic and encyclopaedic data, as typically prevailing in the linguistics of the era of structuralism, linguistically biased theoretical lexicographers analyzed dictionaries, for example, to find the occurrence of encyclopaedic data in lexicographic definitions. They did not question the need for this type of data or the extent that might be permissible but they merely rejected it as non-linguistic entries. I am not pleading for encyclopaedic data in dictionaries – in many cases the linguists were quite right in condemning the specific entries, but they condemned them on the wrong grounds. They are inappropriate not because they are of an encyclopaedic nature but because they go beyond the needs of the typical user of a given dictionary. As Bergenholtz & Gouws (2007a) have indicated, the distinction linguistic and encyclopaedic might be of interest to linguists but from a lexicographic perspective it has little relevance. Also in electronic dictionaries, where space restrictions do not play such an important role, the decision regarding the inclusion of data must be based on the needs of the users and not on the possibilities of the medium.

The strong linguistic focus in at least some dictionaries can also be explained from another perspective – and this is where lexicographers of electronic dictionaries must play a guiding role. The early theoretical discussions of dictionaries mainly had linguists as participants, seeing that theoretical lexicography developed within a linguistic fold. The lexicographic practice is much older than theoretical lexicography, with ‘theoretical lexicography’ referring here to the scientific discipline. It must be accepted that theoretical lexicography had its foundations in the lexicographic practice, and not vice versa. The development went from the practice to the theory. Lexicographic theory developed as a response to the lexicographic practice and the original interest was primarily in the linguistic contents of dictionaries. In this regard, Afrikaans offers an interesting illustration. The development of meta-lexicographical discussions with regard to Afrikaans dictionaries runs parallel to the work on the comprehensive multivolume monolingual Afrikaans dictionary, the *Woordeboek van die Afrikaanse Taal* (henceforth abbreviated as the WAT). According to Gouws (1997: 19) the WAT played a central role in the development of the Afrikaans meta-lexicographical discussion; see also Botha (2003). This is due to the fact that early meta-lexicographical work in Afrikaans was directed at this dictionary, for example, a suggested model for this dictionary (Boshoff, 1926) and, especially later on, various critical reviews of the early volumes of the WAT, for example, Combrink (1962; 1979), Grobler (1978), Odendal (1979), Gouws (1985). Linguists wrote these discussions and reviews, and their comments focused specifically on the linguistic contents of the WAT and the success or failure of this dictionary as a source of linguistic data. In a similar way, linguists prompted the meta-lexicographic discussion regarding the dictionaries in other languages and the initial attempts to formulate a coherent lexicographic