Annette Klosa-Kückelhaus

Introduction

Lexicography is a long-established academic and cultural practice going back many hundreds of years. However, the dramatic growth of the Internet since the late 1990s has led to fundamental changes in this practice. In the meantime, it has become possible to locate and browse through a wide range of lexicographic content relating to almost all of the major languages and many smaller languages and endangered languages in the world in a matter of minutes and free of charge. Just one generation ago, you would have had to undertake an elaborate library search and possibly order a book through an inter-library loan. Many historical dictionaries that were previously only accessible in specialist libraries have also been made freely available as part of comprehensive digitisation projects. This availability of more and more lexicographic content and in new formats is undoubtedly the principal change that users of reference works have seen.

However, behind the scenes of lexicographic practice and research, a great deal more has changed. These changes began as early as the mid-1990s, when the use of computers already began to radically modify the processes of lexicography (cf. Storrer 2001), with the publication of dictionaries in other media not far behind. But the dictionary landscape was altered much more decisively by the advent of free dictionaries on the Internet that were not produced by prestigious publishers. The range of freely accessible lexicographic content online may not have been able to entirely match academic and published dictionaries in terms of quality, but they still drew very high numbers of users and led to a collapse in the sales of publishers' print dictionaries. At the same time, many publishers found it difficult to identify a business model suitable for marketing digital lexicographic data on the Internet for money. Equally, academic dictionaries took a very long time to adjust to the altered media context. However, many of the lexicographers and researchers involved in these dictionaries were able to see the numerous opportunities offered by digital media and the Internet as a publication platform for lexicography (cf., e.g., de Schryver 2003). And yet, naturally, long-established practice does not change overnight, and even now much remains in flux, especially since Artificial Intelligence (AI) and Large Language Models took off in lexicography in the third decade of the 21st century.

This introduction is devoted to the opportunities and perspectives provided for lexicography by digital media and the Internet. Its aim is to communicate to students and academics at universities the central aspects of the research and practice of Internet lexicography. The emphasis lies less on unresolved research questions and specialist technical aspects of Internet lexicography and more on an easily accessible,

Annette Klosa-Kückelhaus, Leibniz-Institut für Deutsche Sprache, R5, 6–13, 68161 Mannheim, Germany, e-mail: klosa@ids-mannheim.de

https://doi.org/10.1515/9783111233758-001

introductory thematic overview of the individual areas of work, enriched through references to further, more in-depth reading. In the process, we have concentrated on key areas of Internet lexicography, with the goal of sharing fundamental concepts and methods in a way that is readily understandable, thereby embedding this important and innovative field of research and practice in university teaching and, above all, in the training of language teachers as well as future lexicographers.

More specifically, our compendium covers the following areas of Internet lexicography: first of all, Peter Meyer, Axel Herold, and Frank Wiegand provide an introduction to **The Technological Context for Internet Lexicography** (→ Chapter 1), explaining the most important technical requirements and processes that enable a dictionary to be provided and used online. Issues to do with logging, versioning, and persistence/identity are also discussed in this chapter. In this way, their contribution makes it easier to understand the technical questions addressed in other chapters (e.g. in relation to the processes involved in editing and publishing an Internet dictionary, different ways of accessing the lexicographic data, and possible approaches to researching dictionary use).

In the chapter on **A Typology of Internet Dictionaries and Dictionary Portals** (→ Chapter 2), Stefan Engelberg and Angelika Storrer develop the criteria for classifying online reference works that are applied in subsequent chapters of this volume. They discuss typological features of Internet dictionaries that are both specific to the medium and independent of it and also propose a typology of dictionary portals (which include several Internet dictionaries).

Chapters 3 to 8 provide insights into the development of an Internet dictionary: in **The Lexicographic Process** (→ Chapter 3), Annette Klosa-Kückelhaus and Carole Tiberius explain how the preparation and publication of an Internet dictionary (or dictionary portal or central lexicographic database) proceed. After introducing and providing an overview of research into the lexicographic process in general, they describe the particular details of the digital lexicographic process for Internet dictionaries, giving specific examples. In addition to discussing software that supports the lexicographic process, the question arises about the process that has to be described in order to develop lexicographic portals and central lexicographic databases.

Axel Herold, Peter Meyer, and Frank Wiegand then investigate the central question of modelling in the chapter on **Data Modelling** (→ Chapter 4), exploring a number of different possible options. They provide an introduction to data structures and formats of representation (e.g. XML documents), different data models (e.g. conceptual-semantic models), and attempts to standardise data modelling for Internet dictionaries (e.g. the Text Encoding Initiative, TEI).

There are also various strategies for linking lexicographic data and providing access to those linked data. These are presented by Stefan Engelberg, Carolin Müller-Spitzer, and Thomas Schmidt in the chapter on **Linking and Access Structures** (→ Chapter 5). They show how lexicographic information in Internet dictionaries can be interconnected and describe onomasiological and semasiological structures for accessing data alongside

other methods (e.g. grapheme-based searches). In the process, the differences between Internet dictionaries and print dictionaries become particularly clear.

In → Chapter 6 on **The Design of Internet Dictionaries**, Annette Klosa-Kückelhaus and Frank Michaelis present some general thoughts on the design of dictionaries and discuss differences between print and online publications. They also explain design dependencies (e.g. on data modelling, on the user) and elaborate on specific aspects of Internet dictionary design such as content-centric presentation vs. user-/human-centric design. Ideas on the design of search functions and the design process as a whole are discussed as well.

Alexander Geyken and Lothar Lemnitzer provide an introduction to one particular aspect of compiling lexicographic content in their chapter on **The Automatic Extraction of Lexicographic Data** (→ Chapter 7), where they explore the different possibilities of extracting word-based information from electronic corpora. Corpora are central in the typology of possible data sources, and the chapter shows in detail what information can be extracted from them to generate particular lexicographic data. The limits of automatic processes are also discussed, in addition to desirable future developments, such as access for users to the primary sources themselves.

In the chapter on **User Participation** (→ Chapter 8), Andrea Abel and Christian M. Meyer report on how users can be involved in the lexicographic process. They distinguish between direct user participation (e.g. forms for entering new word entries), indirect user participation (e.g. feedback forms), and complementary participation (e.g. dictionary blogs), using a range of specific examples to discuss their specific advantages and disadvantages as well as their effects on the lexicographic process involved in creating dictionaries.

A published Internet dictionary can be the subject of **Research into Dictionary Use**, a topic which is introduced by Carolin Müller-Spitzer and Sascha Wolfer in → Chapter 9. Empirical research into dictionary use concerns itself with actual instances of use or, more generally, with observations and experience of dictionary use. As such, it must draw on methods of empirical research in the social sciences, the basic elements of which are elaborated in the chapter. The main part is dedicated to user research in relation to Internet dictionaries, which are the focal point of this introduction.

This book (including the extensive → Index) gives insights into the state of research and its development since the Internet's first phase of popularisation and up to 2024. We have sought to position these developments in the wider tradition of lexicography as a cultural practice and also to illuminate its connections to dictionary research in the typographical age. However, the focus lies on innovations that are connected to digital media and the Internet. Today, lexicography is once again standing "at a turning point in its history" (Granger 2012: 10). We can certainly assume that human beings will always have linguistic questions and needs in the distant future and that some form of tool will be required to deal with them. It is less clear, however, whether dictionaries as we know them today will continue to exist or whether they will be increasingly integrated into the context of smart reading and writing tools and other

digital resources (cf. Lew 2015: 7) and disappear as such. The role that digital dictionaries can and will play for Large Language Models and the other way round is now (in 2024) in the process of being researched and defined as well. This introductory volume should provide the foundations to be able to trace future developments in practice and research.

Our experience of the last three decades of digital lexicography has demonstrated that a cultural practice like lexicography only changes slowly at its core and mostly only as a result of external pressure. As such, we have strong grounds to assume that the present volume will provide a good overview of the field, at least for the coming years. And yet, at some point, this volume, too, will represent but a historical snapshot of Internet lexicography in the mid-2020s.

Bibliography

de Schryver, Gilles-Maurice (2003): Lexicographers' Dreams in the Electronic Dictionary Age. In: *International Journal of Lexicography* 16, 143–199.

Granger, Sylviane (2012): Introduction: Electronic lexicography – from challenge to opportunity. In: Granger, Sylviane/Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 1–11.

Lew, Robert (2015): Dictionaries and Their Users. In: Hanks, Patrick/de Schryver, Gilles-Maurice (eds.): *International Handbook of Modern Lexis and Lexicography*. Berlin/Heidelberg: Springer, 1–9 (manuscript version).

Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, Ingrid/Schröder, Bernhard/Storrer, Angelika (eds.): *Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher*. Tübingen: Niemeyer, 53–69.