Annette Klosa-Kückelhaus and Carole Tiberius

3 The Lexicographic Process

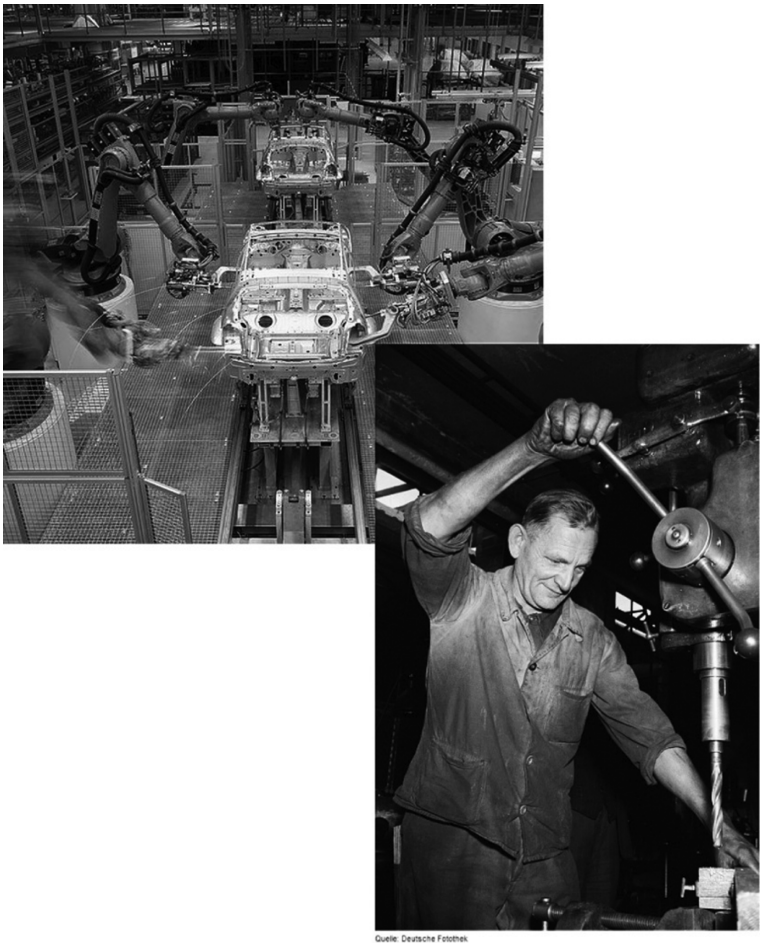


Fig. 3.1: Robots employed on an assembly line – machinist at work.

For many people nowadays, the assembly line symbolises a production process that runs according to a particular sequence of individual activities and was first introduced for the production of cars. Robots have now replaced people for frequently repeated steps but in

Annette Klosa-Kückelhaus, Leibniz-Institut für Deutsche Sprache, R5, 6–13, 68161 Mannheim, Germany, e-mail: klosa@ids-mannheim.de
Carole Tiberius, Instituut voor de Nederlandse Taal, Rapenburg 61, 2311 GJ Leiden, The Netherlands, e-mail: carole.tiberius@ivdnt.org

https://doi.org/10.1515/9783111233758-004

this type of manufacturing process, it is – still – a human being, here a machinist, who assembles and repairs the machines and whose specialist skills and expertise are indispensable. It is precisely within this tension – between automated work on the one hand and lexicographic activity on the other – that the lexicographic processes for born-digital dictionaries take place.

Until the mid-2010s, the lexicographic process was examined almost exclusively in relation to print dictionaries. But for Internet dictionaries, this process takes on an altogether different form: here, it is not a question of describing a linear series of individual production phases but rather individual tasks which are permanently intertwined and run in parallel with one another. A whole series of questions present themselves in this context, such as how subsections of the lexicon are chosen for editing, how new ways of gathering data from electronic text corpora influence the lexicographic process, what software can be used to support lexicographic processes, and what impact all of these changes have on users consulting dictionaries. There are also lexicographic portals in which different dictionaries are combined (→ Chapter 2.4) as well as centralised lexicographic databases from which single dictionaries can be generated, both of which have their own lexicographic processes.

3.1 Introduction

Since time immemorial, the form and content of dictionaries have been the focus of academic dictionary criticism and academic studies of dictionaries. Form and content are easily accessible from the outside because both the structure and substance are visible on paper in a printed dictionary or on the screen in the case of an electronic dictionary. However, the process by which a dictionary comes into being tends to take place much more in the background. By this process, we mean all of the tasks necessary to create a single dictionary or a (centralised) lexicographic database as well as all of the steps that need to be taken to compile a product (a dictionary or other language application) out of the data. These processes involve the participation of people with different skill sets and the use of different technical resources for different periods of time and in a particular order. Traditionally, dictionary editors scarcely provided any detailed information about these internal processes, something which applies equally to commercial and academic lexicography.

Nonetheless, these processes are worth investigating and describing for a variety of reasons:

– The form and content of a dictionary can be judged more appropriately against the background of knowledge about the circumstances in which it originated; this is important, for example, in the context of critical reviews of existing dictionaries. For instance, knowing that there were no financial resources in a dictionary project to be able to buy illustrations or produce their own can explain why re-

course had to be made to freely available images that may not have fulfilled their purpose to optimum effect. A more valid comparison can also be drawn between different dictionaries when information exists about their development timelines and lexicographic teams.

- When the individual processes and participants in the overall production process of a dictionary are known, meaningful suggestions can be developed on this basis for future improvements to the dictionary. For example, if possible ways of presenting data graphically are hardly exploited in an Internet dictionary, or not at all, this might be because the necessary technical know-how was missing from the project or the corresponding work stages were not planned into it. Accordingly, the dictionary editors can change their planning to include a possible visualisation of data if demand arises from users or dictionary researchers, for example.
- The impact of lexicographic processes on published dictionaries, (centralised) lexicographic databases, or lexicographic portals can make the planning of completely new lexicographic projects easier; new dictionaries, databases, or portals can be compiled more efficiently and consistently by learning from mistakes in the planning of existing projects. This may be the case, for example, when an existing dictionary team does not have any corpus-linguistic competence of their own at their disposal so that they are dependent on buying in this expertise from the outside in order to automatically identify corpus evidence. When similar evidence is planned for a new dictionary, this can be avoided if that project has its own corpus linguists in its team from the beginning.
- For users, a good understanding is necessary of specialised aspects of the lexicographic process for lexicographic portals and Internet dictionaries so that they can understand the new elements involved when using them. If they are interested, users should be able to find out about the lexicographic process so that they can assess, for example, to what extent a lexicographic resource is current and can be cited.
- Knowledge about the lexicographic process makes it possible to understand lexicographic resources as the product of many different cognitive, linguistic, computational-linguistic, corpus-linguistic, and editorial tasks and, therefore, as complex cultural goods. In this way, users can learn to assess the quality of edited Internet dictionaries and portals and support their continued existence by using them frequently.

Therefore, we are devoting a whole chapter of this book to the lexicographic processes behind Internet dictionaries, centralised lexicographic databases, and lexicographic portals. Some of what was previously known and well established concerning the lexicographic process in practical lexicography and dictionary research may no longer apply to dictionaries compiled for publication on the Internet. Thus, it is necessary “to unlearn a great deal of what we know” (Gouws 2011: 29; → Section 3.3). The initial foundation for our topic is provided by a brief summary of academic research into the pro-

duction process (→ Section 3.2) and a description of various lexicographic processes for Internet dictionaries of different types (→ Section 3.4). In addition, we consider the process of networks of Internet dictionaries (so-called dictionary portals, → Section 3.6) and centralised lexicographic databases (→ Section 3.7). Since electronic dictionaries cannot be produced without using computers, we show what software can be employed for this in → Section 3.5.

3.2 Research into the lexicographic process

Descriptions of the lexicographic process did not begin in essence until the 1980s, at which time they were restricted entirely to print dictionaries (e.g. Dubois 1990, Riedel/Wille 1979, and Schaeder 1987). To radically simplify and abbreviate matters, Landau (1984: 227) identified three phases that occur in every lexicographic process: "planning (30%), writing (50%), and producing (20%)". Here, the writing phase is said to last considerably longer for academic dictionaries in general than for commercial dictionaries.

Zgusta (1971: 223) defined the following stages in lexicographic work: "(1) the collection of material; (2) the selection of entries; (3) the construction of entries; and (4) the arrangement of the entries". According to Landau (1984), the following tasks are part of the core process, the writing phase, for print dictionaries: drafting a definition, editing it, preparing the text for typesetting (including typographical markups), or proofreading different stages of typesetting (proofs and wraps). As such, these activities include, on the one hand, core lexicographic tasks (explaining the meaning or meanings of a lemma) and, on the other, tasks that generally arise in the production of print media (proofreading). Generally, it is clear that the individual stages of work for print dictionaries are assumed to be linear, at least to some extent. However, it is less clear which individual tasks arise in the planning and production of a dictionary and who undertakes them.

It is to Wiegand that we owe the first complete description and theoretical conceptualisation of the lexicographic process. He defined the lexicographic process as follows (1998: 134):

> Ein abgeschlossener lexikographischer Prozeß [. . .] ist die Menge derjenigen prozeßzugehörigen Tätigkeiten, welche ausgeführt wurden, damit ein bestimmtes Wörterbuch entsteht. [A completed lexicographic process [. . .] is the set of process activities carried out to create a particular dictionary.]

Wiegand (1998: 135) divided the lexicographic process for producing print dictionaries into five phases: preparation, acquisition of material, editing of material, evaluation of material, and typesetting and print preparation. He distinguished generally between lexicographic processes without the use of computers and computer-aided lexicographic processes, i.e. those in which all phases of the process involve the use of computers (Wiegand 1998: 233; for more on the deployment of computers in the lexicographic process, cf. also Knowles 1990: 1648). However, the goal of both processes is

to compile a print dictionary. By contrast, the goal of a digital lexicographic process (involving the use of computers throughout the process) is to compile a dictionary that is not published in print but on an electronic data carrier (Wiegand 1998: 239).

Through critical engagement with Wiegand's proposed model, Müller-Spitzer (2003: 161) arrived at a further elaboration of different lexicographic processes based, first, on whether the dictionary is intended to serve human users or as a resource for language technology. A second distinction is made between a dictionary intended to be published purely on an electronic data medium and a body of data that is medium-neutral and from which both print and electronic dictionaries can be published. As such, all of the lexicographic processes for Internet dictionaries are either digital lexicographic processes or lexicographic processes that are conceived as medium-neutral.[1]

Reflection on involving the human user systematically in the lexicographic process (→ Chapter 8), and more specifically the digital lexicographic process, started at the beginning of the 21st century. Feedback from dictionary users during the development of a dictionary can bring about a clear improvement in quality (e.g. in relation to lemma coverage; cf. de Schryver/Prinsloo 2000a, 2000b). However, the involvement of users (→ Chapter 8) is also useful, for example, in the material acquisition phase if the dictionary is looking for first attestations or examples from sources which are difficult to find. Lexicographers then take on a stronger organisational role in the digital lexicographic process, especially in dictionaries that are compiled semi-collaboratively (Abel/Klosa 2014: 7). Careful planning and supervision of the process is essential, but it is also necessary to make this procedure transparent to dictionary users.

With more extensive dictionary teams, there is a change in the specific expertise needed, as indicated by Wiegand (1998). Thus, a series of other experts are involved in digital lexicographic processes in addition to lexicographers, for example, corpus linguists, computational linguists, text technologists, IT specialists, and designers (Klosa 2013: 504). Against this background, norms and encoding formats for digital lexicography have been established (cf. TEI Lex-0, OntoLex-Lemon, and the ISO LMF (Lexical Markup Framework) standard as well as ISO 1951:2007, and, most recently, DMLex).

3.3 The digital lexicographic process for Internet dictionaries and its particularities

If we use the media-specific characteristics of different types of dictionary to distinguish between them (→ Chapter 2), we can differentiate between the types of Internet dictionary listed in → Tab. 3.1. The original form of an Internet dictionary particularly (but

1 In what follows, we use the shorthand "digital lexicographic process".

not only) influences the dictionary's online structure, the density of hypertexts, the number of multimedia elements, and the means of access available. The criterion of completeness has a particular influence on the lexicographic process and it does so continuously until the dictionary is (possibly, but not necessarily) finished. Schröder (1997) thus introduced the distinction between "Abschlusswörterbuch" (completed dictionary) and "Ausbauwörterbuch" (dictionary under construction) while Lemberg (2001) referred to "statisch" (static) as opposed to "dynamisch" (dynamic) dictionaries. As far as dictionaries under construction are concerned (→ Section 3.3.2), we can further differentiate between those that initially appeared in print and were then digitised retrospectively before being (continuously) extended and revised (e.g. the OXFORD ENGLISH DICTIONARY [OED] or the DEUTSCHES RECHTSWÖRTERBUCH [DRWB]) and those that were planned directly for online publication and are continuously extended and revised (e.g. the ALGEMEEN NEDERLANDS WOORDENBOEK [ANW] or ELEXIKO, which, however, stopped as a project in 2017).

Tab. 3.1: Possible classification of Internet dictionaries (following Storrer/Freese 1996; Storrer 1998; Storrer 2001).

Characteristic	Type of Internet dictionary
Original form of publication	– first appeared as a print dictionary – first appeared as an electronic offline dictionary – appeared directly as an online dictionary
Completeness	– completed dictionary – dictionary under construction
Hypertexts	– dictionary with hypertexts – dictionary without hypertexts
User interaction	– dictionary with user interaction – dictionary without user interaction
Multimedia	– dictionary with text, illustrations, tables, diagrams – dictionary with text and audio data – dictionary with text, illustrations, and audio data – dictionary without multimedia
Access to the dictionary	– access by scrolling through the list of lemmas – access via a list of lemmas with specific characteristics in hypertext form – access via search options – combined forms of access

Finally, user interaction influences the lexicographic process insofar as dictionaries with user participation have to account for users' feedback in specific phases of their process (→ Section 3.3.3, → Chapter 8). Before examining the particularities of the lexicographic process for these different types of dictionaries, we can describe the digital

lexicographic process more generally in seven phases, each with many individual tasks (→ Tab. 3.2), using Wiegand (1998: 233ff.) as a starting point.

Tab. 3.2: Phases and tasks in the digital lexicographic process for Internet dictionaries.

Phase	Tasks (selected)
Preparatory phase	Dictionary outline, organisational plan (finances, workflow, timetable, staffing), pilot studies on lexicographic information types and the list of lemmas, plan for the dictionary (rough modelling of the data structure, editorial guidelines, model entry, planning the user interface and access structures, planning technical support, planning the versioning and archiving of dictionary data, planning user involvement and user studies)
Data acquisition phase	Acquiring primary lexicographic sources (corpus construction) and additional sources (e.g. reference dictionaries), acquiring further data (e.g. illustrations, videos, audio data)
Computerisation phase	Preparing corpus texts (tagging, lemmatisation), programming/acquiring a corpus research and analysis tool, programming/acquiring a dictionary writing system and implementing the data model, programming for the versioning and archiving of dictionary contents, programming for the user interface, acquiring and installing necessary hardware and any further software
Data processing phase	Compiling potential lemmas and frequency lists, defining frequency levels and classes, analysing co-occurrences, labelling image and audio data, integration of automatically identified data into dictionary entries (e.g. frequency, collocations)
Data evaluation phase	Linking dictionary entries and corpus, writing entries, inserting hyperlinks, integrating illustrations and other multimedia elements
Preparation phase for online release	Proofing content and form, testing the user interface (hyperlinks, multimedia elements, search options, etc.), writing user manuals for the dictionary, developing a “guided tour” of the dictionary
Maintenance and preservation phrase	Archiving different versions of the data and version control; maintenance of the online application

Different technical qualifications are needed to undertake the tasks listed. Ideally, everybody involved in the dictionary project participates in the preparatory phase, i.e. lexicographers and corpus linguists when planning the design of the corpus, or lexicographers and graphic designers or specialists in web design when conceiving the layout of the dictionary entries. Corpus linguists (acquiring corpus texts) and lexicographers (inspecting and gathering other sources) also work together in the data acquisition phase. The tasks involved in the computerisation phase lie primarily with corpus and computational linguists and, possibly, programmers, but most of these tasks cannot proceed without the agreement of the team of lexicographers.

The phase that takes up the most amount of time in the lexicographic process is the data evaluation phase in most dictionary projects. This is the lexicographers' main area of responsibility, but this only partially applies to Internet dictionaries if information compiled automatically from corpora, or with natural language processing software, or generated with the help of Large Language Models (→ Chapter 6) appears in the dictionary alongside lexicographic information prepared by editors. For example, it is the responsibility of corpus and computational linguists to ensure that examples are extracted automatically from the corpus and information about syllabification is automatically generated for the lemmas. Nonetheless, in many dictionary projects, information that has been compiled automatically is checked by the editors before it appears in the dictionary as part of what is called "post-editing lexicography" (cf. Jakubíček 2017). This checking and correction takes place in the phase when the dictionary is being prepared for online release, during which the lexicographers work together with the programmers to test the user interface. Once the dictionary (or dictionary portal) has been published, its website needs to be maintained and preserved as one of the last phases of the process (cf. Svensén 2009: 413; Tiberius/Krek 2014: 1).

Below we consider finished dictionaries that have been retrospectively digitised in → Section 3.3.1 and then two types of dictionaries under construction in → Section 3.3.2 before reflecting on the implications of dictionary users' feedback for the digital lexicographic process in → Section 3.3.3.

3.3.1 The digital lexicographic process in retro-digitised Internet dictionaries

Retro-digitised dictionaries are finished dictionaries originally published in print whose lexicographic process has already run its complete course. However, they are to be made newly available in an electronic medium (here, the Internet). As such, of the phases of the lexicographic process described above, only the following apply in the case of retrospective digitisation of a dictionary that is to be published online (for an example → Section 3.4.3):

- Preparatory phase: organisational plan (finances, workflow, timetable, staffing) and plan for the dictionary (rough modelling of the data structure, possibly in alignment with a centralised lexicographic database if needed; planning the user interface and access structures, planning technical support, and potentially planning for user involvement);
- Computerisation phase: acquiring and installing the necessary hardware and software, implementing the data model and the user interface;

– Preparation phase for online release: testing the user interface, developing a "guided tour" of the dictionary;
– Maintenance and preservation phase: → Tab. 3.2.

However, the phase with the greatest overall significance for the lexicographic process – the data evaluation phase including the core lexicographic, corpus linguistic, and computational linguistic tasks – is missing altogether, as is the data acquisition phase. As such, it is worth asking whether the process leading to the publication of a retrospectively digitised dictionary can be defined as a digital lexicographic process at all (→ Section 3.6). If a retro-digitised Internet dictionary is not simply a 1:1 replica of the underlying print dictionary but rather has an added lexicographic value online, we consider it legitimate to refer to at least a digital lexicographic sub-process. This added value exists, for example, if:
– the lexicographic information is presented in a different way (e.g., distributed on seperate windows instead of a simple reproduction of the print image),
– the dictionary can be searched in an innovative way,
– the mediostructure is extended systematically with hyperlinks.

Retro-digitisation projects primarily require the expertise of computational linguists and programmers as the tasks are mostly computational. However, lexicographic expertise is indispensable in a successful retro-digitisation dictionary project for planning the user interface and access structures as well as for modelling the data structure and extending the content.

3.3.2 The digital lexicographic process in dictionaries under construction

As mentioned above, when it comes to dictionaries under construction we must first of all distinguish between whether the dictionary is a completely new creation (for examples → Section 3.4.1) or whether a print dictionary is first being retro-digitised (in parts or as a whole), then published online, and continuously extended in that form. The latter raise the interesting case of a lexicographic process for a print dictionary, which may or may not be computer-aided, that has been completed and after which a digital lexicographic process is initiated for the Internet dictionary. In this scenario, certain phases in these processes may be omitted, others may replace them, and new tasks may be added:
– Preparatory phase: only questions specific to the medium of the Internet have to be clarified e.g. planning the retrospective digitisation, planning the further development of the dictionary in particular subsets of the corpus, potentially planning the development of the underlying data, and planning the user interface.

- Data acquisition phase: this can be omitted if only the material from the print dictionary (not edited electronically) actually serves as the primary source for the further development of the dictionary.
- Computerisation phase: the development of the Internet dictionary cannot be realised without using computers in a consistent way, even if the print dictionary has already been compiled with the help of computers. The user interface also has to be implemented.
- Data processing phase: this phase is also omitted if no new sources are to be used for the dictionary.
- Data evaluation phase: the tasks already undertaken by lexicographers when evaluating material for the print dictionary are also necessary for the further development of the Internet dictionary. Further tasks might be added, e.g. editing and preparing hyperlinks.
- Preparation phase for online release: this phase replaces the typesetting and print preparation phase for a printed dictionary.
- Maintenance and preservation phase: this phase did not exist for the printed dictionary but is of major importance for the online dictionary under construction, which needs updates as well as version control.

For print dictionaries and finished (retro-digitised) electronic dictionaries (→ Section 3.3.1), the typesetting and print preparation phase, or the preparation phase for online release, only begins once the preceding phases in the process have been completed. By contrast, dictionaries under construction that are intended for Internet publication from the very start are published gradually, bit by bit. Furthermore, dictionaries under construction are not necessarily compiled in alphabetical order but may appear instead in modules since the alphabetical sorting and listing of lemmas has become redundant as an access structure (→ Chapter 4). The body of entries included in a module can be defined in different ways (e.g. on the basis of frequency or according to word class; → Section 3.3). Working on a module (or partial lexicon) chosen according to particular criteria is an advantage for the lexicographers because in this way the lemmas can often be edited more consistently (Storrer 2001: 61f.).

Within a dictionary under construction project, it is also possible to work on multiple modules in parallel. Hence, it might be the case that one module is still in the preparatory phase, another is under construction, and a third has already appeared online. Depending on the decisions taken about the publication cycle in the dictionary project, the different phases can overlap even more. For example, if a project decides to release each completed entry immediately instead of posting larger finished collections of entries online (e.g. every quarter), the following situation can arise: the data for some of the lemmas in a module are still being evaluated, other lemmas are nearly ready to be published online, and others have already been published. If a dictionary

combines the automatic compilation of lexicographic information with the manual editing of lemmas, the boundaries between the phases in the process will shift still further: a lemma may already have appeared online with its automatically compiled data while subsequent editorial treatment of the same lemma may only be entering the data evaluation phase. As such, the same lemma may find itself in two different phases of the lexicographic process at one and the same time.

These reflections make clear that dictionaries under construction are "offene Systeme" (open systems; Schröder 1997: 16). As such, the digital lexicographic process for these dictionaries is to be seen as more circular than linear. It is essential that this process is carefully planned and continuously checked; in addition, all of those involved must be in a position to undertake different tasks from different phases of the process in parallel. This is carried out more easily when a dictionary under construction is generated from an existing (centralised) lexicographic database. In such a scenario, data from retro-digitised resources, corpus data, or other linguistic information can be combined with newly compiled lexicographic information.

3.3.3 The digital lexicographic process and dictionary users

There is a range of options for involving users directly in the work on dictionaries (→ Chapter 8) and this can have an impact on the lexicographic process for Internet dictionaries:[2]

- Users are asked to report errors and/or to make suggestions for existing entries and new entries;
- Users can provide comments on a word entry that are answered by the editors (both the comments and answers are available online for other users);
- Users privately ask questions on the content of the dictionary that are answered by the editorial team;
- Users give each other advice (e.g. in forums) or are asked to assess the content provided by other contributors.

What happens after errors have been logged by users depends on the type of error reported. In → example (1) a user of the German online dictionary ELEXIKO informed the editors that there was an orthographic mistake in the morphological variant *Burschenschaftler* given for the lemma *Burschenschafter* 'member of a student fraternity'. The user also pointed out that not all entries for German words ending with *-schaftler* showed the correct syllabification ". . . schaft|ler" and asked whether this had been done on purpose:

2 For more on the digital lexicographic process in collaborative dictionaries such as WIKTIONARY, → Section 3.4.2.

(1) Sehr geehrte Damen und Herren,
beim Eintrag "Burschenschafter" hat sich bei der Worttrennung der morphologischen Variante ein (Tipp-) Fehler eingeschlichen: Bur|schen|schaf|lter

Es sollte wohl "Bur|schen|schaft|ler" heißen.

Einige Einträge auf schaftler weisen dabei keine Trennung von "schaft" und "ler" auf. Ist dies beabsichtigt? (Vgl. z.B. Kulturwissenschaftler)

The mistake in this example requires making a correction in a single entry and then systematically checking a whole series of words for corresponding corrections. In a case like this, a lexicographer has to put the relevant entry or the relevant groups of headwords back into the data evaluation phase, undertake the corrections, check them again in the online preparation phase, and finally re-release the relevant entry or entries online.

In → example (2) a user of ELEXIKO noticed that a search for *tausende* 'thousands' did not yield any result.

(2) hallo,
elexico findet nichts für die suche nach "tausende".
mfg

This error relates to the list of lemmas or the dictionary's search functionality rather than to individual lemmas or a group of lemmas. If the missing lemma really is absent, even though it should be on the list according to the original concept behind that list, the lexicographers will also return to the data preparation and evaluation phases here. For example, they will check in the corpus whether the missing word is present in the correct spelling in sufficient frequency. It is possible that all of the other stages of editing and revision will follow on here before the lemma can appear online. The process will be different if the word could not be found because of an error in the search functionality or the way the search functionality was realised. In this case, the relevant technical colleagues will have to return to the computerisation phase. If the programming of the search options needs to be corrected, this, of course, will need to be tested again before the improved version is released online.

In general, dealing both with errors reported by users and with their suggestions for improvements, additions, new lemmas, or the revision of selected lemmas and similar means that for dynamic Internet dictionaries, certain phases of the digital lexicographic process are undertaken again. Requests from dictionary users or discus-

sions in dictionary blogs may have a similar effect, namely when these serve as an impetus for the dictionary team to further develop the dictionary in a particular way or to undertake systematic corrections or additions.

Research into the use of the dictionary (→ Chapter 8) can have similar consequences. For example, unsuccessful searches in the log files may lead to the conclusion that potential headwords have been omitted and the dictionary project may decide for these words to be worked on retrospectively. In this case, the project has to go back to the phases of data processing and analysis as well as preparation for publication in the lexicographic process. If the results of a user survey indicate that the dictionary should be urgently updated with information about pronunciation in the form of audio data, but these were not originally envisaged and therefore have to be acquired only now, the project needs to go back to the data acquisition phase. Or the dictionary project has to return to the preparatory phase because an evaluation of the user interface in user studies requires a thorough revision of the interface.

To some extent this circularity is characteristic for the lexicographic process in Internet dictionaries under construction because of the involvement and consideration of dictionary users. However, it need not only be burdensome in terms of the dictionary team being forced to permanently correct and add to the dictionary; it can also be an opportunity because the dictionary can be developed and improved in terms of its user friendliness and overall usefulness. Furthermore, these steps can also make it possible for those consulting the dictionary to come to terms better with the changing state of the dictionary. However, the latter can only work if there is a careful versioning of the dictionary.

In the ideal case, all revisions, additions, or deletions in a dictionary under construction have to be marked in a transparent way for users (or at least for users in an academic context, where exact bibliographic details are needed for citations from the dictionary). However, as a bare minimum, the different stages of revision should be recorded (as in → Fig. 3.2 in the example from the OED) or the relevant most recent versions or date of revision (as in → Fig. 3.3 in the example from the Digitales Wörterbuch der deutschen Sprache [DWDS]).

It must be borne in mind that it is very costly in technical terms to keep older editions of the dictionary genuinely available. As such, the project team must weigh up whether the effort is worthwhile for information that may be accessed very rarely.

3.4 Examples of lexicographic processes for Internet dictionaries

In this section we consider and discuss the lexicographic practice and process for four Internet dictionaries: the Algemeen Nederlands Woordenboek (ANW), elexiko, and the German-language Wiktionary, which were planned for online publication only, and a retro-digitised dictionary, the Ordbog over det Danske Sprog (ODS).

Fig. 3.2: Record of the publication history of the entry *dictionary* in the OED.[3]

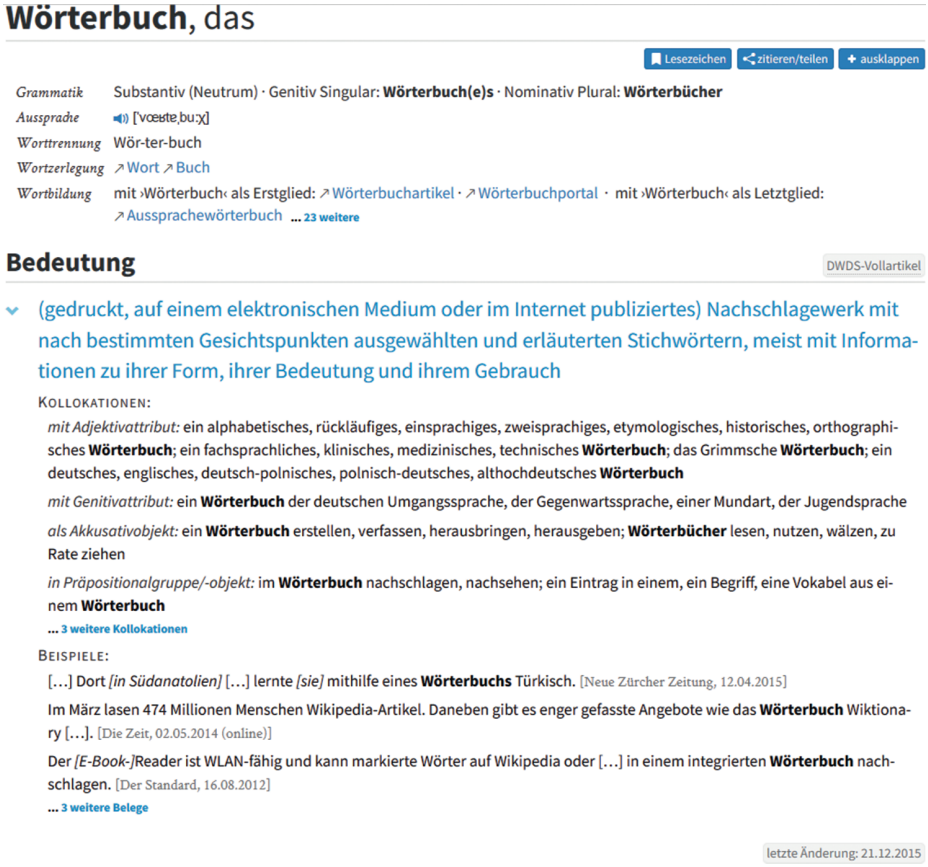
3.4.1 The lexicographic process for the ANW and ELEXIKO

Here, we combine our discussion of the lexicographic processes for the ANW and ELEXIKO since the two projects were relatively similar from the outset. Both are academic, corpus-based dictionaries that describe contemporary language use, the ANW for Dutch and ELEXIKO for German. The ANW is one of the projects run by the Instituut voor de Nederlandse Taal (INT) in Leiden, and ELEXIKO is a project realised at the Leibniz Institute for the German Language (IDS) in Mannheim. The ANW and ELEXIKO are Internet dictionaries that are (or were) open in the truest sense of the expression. From the very beginning, both were planned for online publication including continuous further development (in the case of ELEXIKO, until 2017, when the project ended).

→ Table 3.3 shows the different phases of the digital lexicographic processes involved in the ANW and ELEXIKO between 2001 and 2017. The table makes clear that the different phases no longer take place neatly one after the other, as they would for a print dictionary, but rather that they run in parallel and that it can be difficult to separate the individual phases from each another. In general, we notice many overlaps between the individual phases, except in the data preparation phase.

Determining the end of a phase is not always unequivocal. For instance, the phase of data processing may be reactivated whenever new technology becomes available, like for the automatic extraction of examples from a corpus or the com-

3 Webpage last accessed on 24 March 2024.



Wörterbuch, das

Lesezeichen | zitieren/teilen | ausklappen

Grammatik Substantiv (Neutrum) · Genitiv Singular: **Wörterbuch(e)s** · Nominativ Plural: **Wörterbücher**
Aussprache [ˈvœʁstə,bu:χ]
Worttrennung Wör-ter-buch
Wortzerlegung ↗Wort ↗Buch
Wortbildung mit ›Wörterbuch‹ als Erstglied: ↗Wörterbuchartikel · ↗Wörterbuchportal · mit ›Wörterbuch‹ als Letztglied: ↗Aussprachewörterbuch ... 23 weitere

Bedeutung DWDS-Vollartikel

(gedruckt, auf einem elektronischen Medium oder im Internet publiziertes) Nachschlagewerk mit nach bestimmten Gesichtspunkten ausgewählten und erläuterten Stichwörtern, meist mit Informationen zu ihrer Form, ihrer Bedeutung und ihrem Gebrauch

KOLLOKATIONEN:

mit Adjektivattribut: ein alphabetisches, rückläufiges, einsprachiges, zweisprachiges, etymologisches, historisches, orthographisches **Wörterbuch**; ein fachsprachliches, klinisches, medizinisches, technisches **Wörterbuch**; das Grimmsche **Wörterbuch**; ein deutsches, englisches, deutsch-polnisches, polnisch-deutsches, althochdeutsches **Wörterbuch**

mit Genitivattribut: ein **Wörterbuch** der deutschen Umgangssprache, der Gegenwartssprache, einer Mundart, der Jugendsprache

als Akkusativobjekt: ein **Wörterbuch** erstellen, verfassen, herausbringen, herausgeben; **Wörterbücher** lesen, nutzen, wälzen, zu Rate ziehen

in Präpositionalgruppe/-objekt: im **Wörterbuch** nachschlagen, nachsehen; ein Eintrag in einem, ein Begriff, eine Vokabel aus einem **Wörterbuch**

... 3 weitere Kollokationen

BEISPIELE:

[…] Dort *[in Südanatolien]* […] lernte *[sie]* mithilfe eines **Wörterbuchs** Türkisch. [Neue Zürcher Zeitung, 12.04.2015]

Im März lasen 474 Millionen Menschen Wikipedia-Artikel. Daneben gibt es enger gefasste Angebote wie das **Wörterbuch** Wiktionary […]. [Die Zeit, 02.05.2014 (online)]

Der *[E-Book-]*Reader ist WLAN-fähig und kann markierte Wörter auf Wikipedia oder […] in einem integrierten **Wörterbuch** nachschlagen. [Der Standard, 16.08.2012]

... 3 weitere Belege

letzte Änderung: 21.12.2015

Fig. 3.3: Information on the last update (bottom right) for the entry *Wörterbuch* 'dictionary' in the DWDS.[4]

puter-aided compilation of word families. The data in → Tab. 3.3 also indicate that the ANW was more dependent on computer support than ELEXIKO.

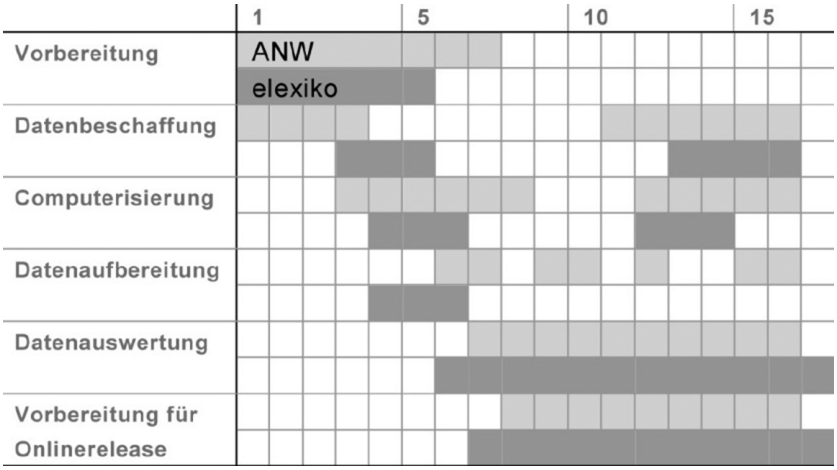
In what follows we provide a comparative analysis of the different phases in the lexicographic processes of the two projects.

Preparatory phase

In the field of academic lexicography, both the ANW and ELEXIKO broke new ground, and essentially the only useful experience that the projects could build on came from print lexicography. For that reason, some essential steps were missed out on or incorrectly im-

4 Webpage last accessed on 24 March 2024.

Tab. 3.3: Process phases (over the years) for the ANW and ELEXIKO (left column: preparatory phase – data acquisition phase – computerisation phase – data processing phase – data evaluation phase – preparation phase for online release; the last phase for maintenance and preservation is not shown but started as soon as the dictionaries were released online).



plemented in both projects. In the preparatory phase, both projects concentrated on developing the concept of the dictionary's content and on pilot studies, underestimating the importance of constructing a detailed organisational plan including information about finances, staffing, timetable, and workflow. For example, user studies were not initially envisaged or carried out, with log data being stored since the beginning of the online publication of the ANW in 2009 but not analysed systematically; only incidental detailed log analyses have been undertaken (cf. Tiberius/Niestadt 2015). Furthermore, as Abel/Klosa (2012) concluded, no proper market analysis of existing dictionary writing systems was carried out for ELEXIKO. As a result, both projects incurred unnecessary expenditure of time and money in order to improve or retrospectively undertake tasks that were initially poorly planned or not planned at all.

Data acquisition phase

Since the ANW and ELEXIKO are corpus-based dictionaries, the data acquisition phase was devoted above all to assembling the corpus. The ANW corpus was originally a closed corpus of 100 million words that was specially compiled for the project. This idea has since been reconsidered, and new material is being added to the corpus. The ELEXIKO corpus was a dynamic corpus constituted virtually from the German reference corpus (DEREKO) at the IDS Mannheim and subsequently updated continuously until the end of the project.

Both projects also invested much time in gathering image material (in particular, image material that can be used free of charge). The ANW uses online databases of public domain photographs, such as WIKIMEDIA, as well as other illustrations from the Internet (always with reference to the original source. The illustrations in ELEXIKO also came from online databases of public domain photographs such as WIKIMEDIA or PIXELIO and were only systematically added to the edited lemmas between 2012 and 2017 so that the project had to return to the data acquisition phase for illustrations after it had already been running for a long time.

The ELEXIKO project also planned, successfully, to illustrate the natural pronunciation of lemmas (in context) with the help of examples from sound recordings. Here, up to three examples per lemma had been selected for this purpose from the "Archiv für gesprochenes Deutsch" (AGD) held by the Leibniz Institute for the German Language (IDS), although the examples were only available to download and listen to from 2012 onwards.

These examples from the data acquisition phase demonstrate that the phases do not really run one after another but can repeatedly overlap. For most dictionaries under construction, the data acquisition phase runs until the end of the project, although the main period lies at the beginning of the project with the original conceptualisation and constitution of the corpus.

Computerisation phase

In the computerisation phase, both projects prepared and processed the corpus texts so that they could be used lexicographically and established an editorial system (also known as a dictionary writing system). The preparation and processing of the corpus involved, first, tokenising and enriching the corpus with lemmas and part-of-speech tags and, second, loading the corpus into a system in which it could be searched. For the ANW, an in-house corpus query system was used originally but this was replaced in 2007 by the commercial system SKETCH ENGINE (Kilgarriff et al. 2004) and now a combination of SKETCH ENGINE and an in-house system is used. ELEXIKO used corpus tools already available at the Leibniz Institute for the German Language (IDS).

A lexicographic editing system was developed in-house in 2007/2008 for the ANW and the system has been used since 2008 (Niestadt 2009; Tiberius et al. 2014). The INT's editing system is regularly improved on the basis of new insights provided not only by lexicographers but also by computational linguists and programmers. It is also used for other projects at the institute as well as for a dictionary project at the Frisian Academy. The ELEXIKO project did not use a dictionary writing system in the strict sense, but instead an existing XML editor (OXYGEN[5]) was adapted to the needs of the project. The edi-

5 OXYGEN is a commercial XML editor: http://www.oxygenxml.com/.

tor was supplemented with additional software, namely an in-house linking tool called "Vernetziko" (cf. Meyer 2011) and an administrative tool.

Backing up data and versioning are important in the computerisation phase. For example, the Norwegian dictionary project NORSK ORDBOK established that if the daily backup of data did not work, this basically meant losing data that would take six person-weeks to compile (Grønvik/Smith Ore 2013: 255). In the ANW, Git[6] software is used for versioning the source code of the editing system and the web application. The dictionary entries are stored in a MySQL database. Backups are made every day that are stored for three months. Each quarter a backup is made that is stored for a year, and once a year a backup is made that is stored for ten years. This all takes place in order to avoid delays in the project.

Data preparation phase

As a rule, the first task that has to be completed in the data preparation phase is compiling a list of potential lemmas. This is a semi-automatic task as often lexicographers have to manually check the candidates. Although the possibilities for automatically extracting lexical data from corpora have improved significantly in the last decades through improvements in language technology (→ Chapter 6), the ANW still is and ELEXIKO was always relatively conservative when it came to including automatically extracted data in the dictionary, and their corpus data still are/were predominantly analysed manually (although, of course, the analysis is computer aided). However, both projects contain automatically extracted data. The ANW contains data that is dynamically derived from other lexical resources, such as information on orthographic form from the official guidelines on Dutch spelling. In the ELEXIKO project, for example, three examples identified in the ELEXIKO corpus are displayed for all the lemmas that were not edited before the project came to an end.

Data evaluation phase

Originally, the ANW had a larger lexicographic team at its disposal than the ELEXIKO project. The ANW had five full-time lexicographers until mid-2015 in addition to three full-time lexicographic assistants, an editor-in-chief, and a project manager. This has now been reduced to one full-time lexicographer assisted by two lexicographers who dedicate a small amount of their time to the ANW. Four full-time lexicographers (at one time, five) were employed on the ELEXIKO project (including the editor-in-chief). Sometimes, students work(ed) as interns with the ANW or ELEXIKO.

6 Git is a free, open source tool for version control: http://git-scm.com.

The ANW had a more complex workflow in the data evaluation phase than ELEXIKO since the lexicographers and lexicographic assistants worked closely together as can be seen in → Fig. 3.4. On the ELEXIKO project there was no support from lexicographic assistants, and a lexicographer edited the complete entry for any word. The situation is now the same for the ANW.

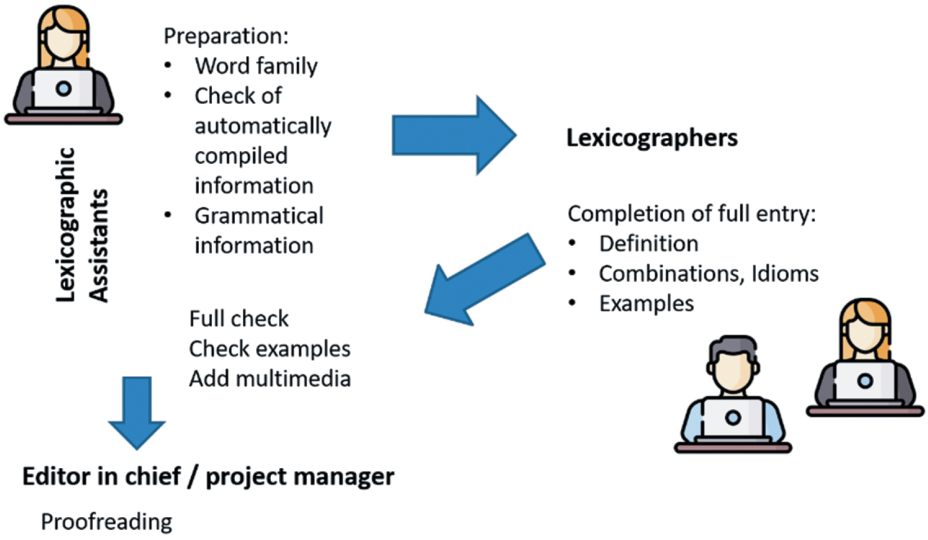


Fig. 3.4: The original ANW workflow.[7]

In this phase, the work of the lexicographers is not fundamentally different from the data evaluation phase in a print dictionary. However, in all cases the analysis is computer aided; that is, functions such as concordancing, filtering, and sorting make it possible for lexicographers to analyse large amounts of data efficiently.

Preparation phase for online release

In this pre-final phase the dictionary entries are proofread. In the early years of the ANW, a new online version of the dictionary was compiled every three months. The editor-in-chief and project director would do the proofreading after which the articles would go back to the lexicographers for final revisions. As soon as that was done, the editor-in-chief and project director would check the entries again and then change the status of the dictionary entry to "going online". Spelling and hyperlinks were

7 Icons designed by Freepik: https://www.freepik.com/.

checked automatically and then corrected manually. After that, a new version of the dictionary was uploaded into a test environment for one week. Errors and inconsistencies, etc. could still be corrected. If, after a week, the version in the test environment was approved, an updated version of the dictionary was published on the public website. Nowadays, the ANW is much more dynamic, and updates are done overnight so that changes to the data are visible online the very next day.

In ELEXIKO, all entries first underwent a double reading for content (by the editor-in-chief and a different lexicographer than the one who compiled the entry, as seen from the different handwriting in → Fig. 3.5) followed by formal proofreading (e.g. for orthographic errors). All the revisions were checked again, and generally all the hyperlinks in the entry were checked, illustrations were opened to test them, and audio data played, etc. before the finished article was published online overnight.

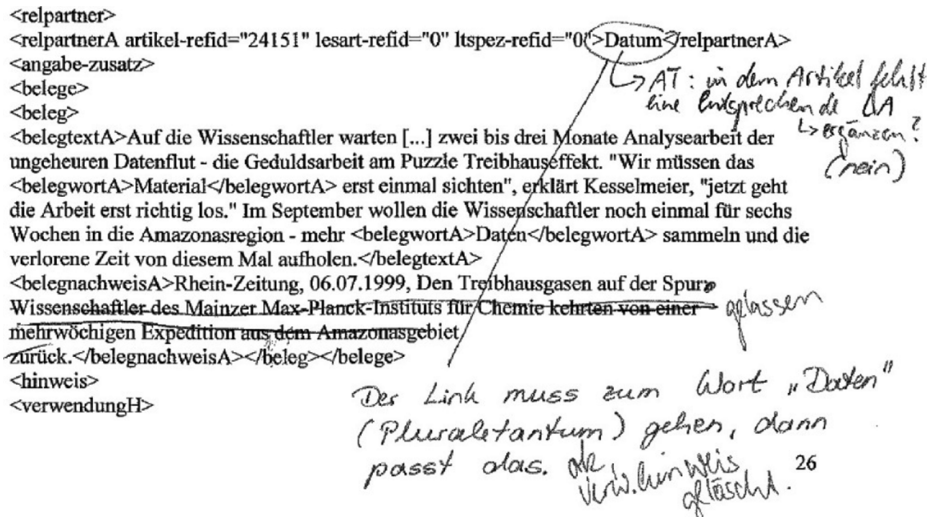


Fig. 3.5: Extract from the double content-editing of a manuscript version (with XML markup) of the entry for *Material* 'material' in ELEXIKO (excerpt from paradigmatic relationships with corrections hinting at wrong linking).

Before a dictionary is first published on the web, the outer texts for the dictionary also need to be written like user instructions, information about the content, and information about the corpus, etc. For the ANW, these dictionary outer texts have remained largely unchanged since first published in 2009, and are only changed when necessary to reflect the current status of the project. For ELEXIKO, these texts had been online since 2003 in their initial, very brief form before being completely revised and extended in 2007.

Maintenance and preservation phase

Because ELEXIKO is part of the dictionary portal OWID at the Leibniz Institute for the German Language (IDS), all technical tasks of maintenance and preservation are carried out in this context. However, at different times, editors had to update entries, for example in 2011, when a new version of the official German orthographic rules was released. These comprised new rules for the syllabification of loanwords so that this information in many entries in ELEXIKO had to be changed accordingly.

Maintenance of the ANW is an integral part of the workflow of the project to ensure that the online application is continuously up and running. To this end software updates are carried out regularly and the data is updated on a daily basis.

Summary

As we noted at the outset, work on the ANW and the ELEXIKO project began without having a complete overview of the digital lexicographic process necessary to realise the projects successfully. As a result, important planning stages were overlooked. Because both dictionaries were compiled completely from scratch and were conceived exclusively for the Internet medium, the necessary experience could only be acquired in practice in order to ultimately set out the lexicographic process in full. In doing so, it was possible to translate existing experience from print lexicography.

3.4.2 The lexicographic process for the German WIKTIONARY

In this section we consider the digital lexicographic process for another type of dictionary under construction, namely for WIKTIONARY, the collaboratively compiled online dictionary. WIKTIONARY is a freely available, multilingual dictionary for the vocabulary of different languages, the content of which is compiled collaboratively. Meyer/Gurevych (2016) describe the lexicographic process for the German WIKTIONARY and compare it to the lexicographic process for dictionaries compiled by an editorial team. In their study, they concluded that the phases recognisable from print dictionaries are only replicated to a small extent in WIKTIONARY. The data acquisition and data preparation phases merge strongly with the data evaluation phase while the data evaluation phase is almost impossible to distinguish from the preparation for online publication phase since the markup language is automatically translated into fully formatted dictionary entries.

Another important difference is that the lexicographic process for WIKTIONARY is shaped strongly by revisions and by discussion. The writing process for entries in the dictionary is based on multiple revisions by different authors. The revisions include additions, more specific details, re-formulations, corrections of errors, the integration of examples and sources, and deletions of what some editors believe is irrelevant ma-

terial. There are special discussion pages relating to the overall concept of the dictionary as well as to individual entries that can be used for preparing, assessing, and implementing changes and that represent an important tool for tracking changes. Meyer/Gurevych (2016) have proposed a new model to better describe the lexicographic process for WIKTIONARY (→ Fig. 3.6): in this model, different collaborating users take already part in the preparatory phase with the conceptualisation of the dictionary. In the compilation phase, articles are written, and the lexicographic instructions are continuously updated. Entries and instructions alike are discussed by users, which may lead to revisions of entries. Revised entries can be the topic of user discussions as well. Thus, collaborating users are part of every step in WIKTIONARY's lexicographic process.

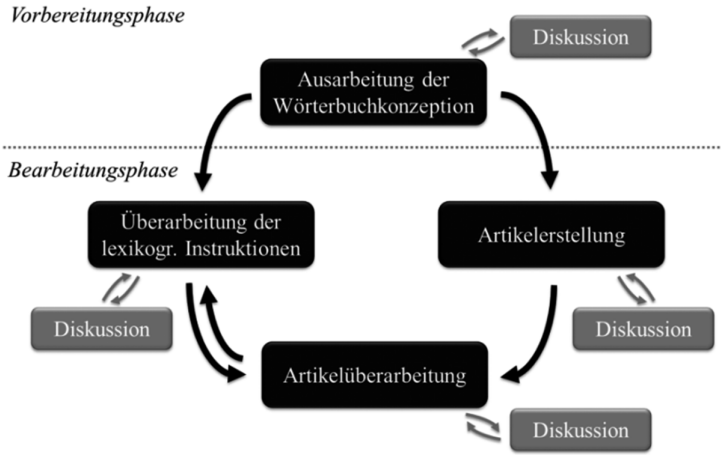


Fig. 3.6: Process model for WIKTIONARY (following Meyer/Gurevych 2016: 67); top: preparatory phase: development of the dictionary concept with discussion; bottom: editing phase: revision of lexicographic instructions (left) with discussion, compiling entries (right) with discussion, revision of entries (bottom) with discussion.

3.4.3 The lexicographic process for the ORDBOG OVER DET DANSKE SPROG

Finally, we consider the lexicographic process for the ORDBOG OVER DET DANSKE SPROG (ODS), the retrospectively digitised Danish language dictionary. The ODS is a historical dictionary comparable to the great national dictionaries like Jacob and Wilhelm Grimm's DEUTSCHES WÖRTERBUCH (DWB), the WOORDENBOEK DER NEDERLANDSCHE TAAL (WNT), the OXFORD ENGLISH DICTIONARY (OED), and the SVENSKA AKADEMIENS ORDBOK (SAOB). It was originally published between 1918 and 1956 in 28 volumes and has since

been extended with five additional volumes (1992 to 2005). There has been a digital version since November 2005 available at ORDNET.DK, a portal of dictionaries for the Danish language.

The digitisation of the ODS began relatively late compared with other projects. The SAOB began being digitised as early as 1983, the OED and WNT in 1984, and the DWB between 1998 and 2003/4. As a result, the ODS was able to benefit from the digitisation experiences of other projects. The process began with a preparatory phase in 2004 and was supposed to last for five years. In broad terms, two main tasks could be distinguished in the subsequent data preparation phase: raw digitisation and structural markup.

The first step in the digitisation process was carried out in collaboration with the Center for Digital Humanities at the University of Trier, which was also responsible for digitising the DWB. Dual text capture without proofreading was the method used for the raw digitisation, that is, the print version of the dictionary was typed up by two people (in Asia). In order to achieve good results, a handbook was drawn up to ensure consistent coding for special characters and symbols. After the text capture, the two versions were compared automatically in Trier, and a list of differences was generated, checked, and corrected. This process of text capture took nine months.

In the second stage, the digitised version was marked up structurally using scripts. First, a rough markup was created in which only lemmas, homograph numbers (if present), and word classes were explicitly indicated as such. This version took around two years to complete and appeared online in November 2005. The markup was subsequently refined in order to identify other units of information in the microstructure of the dictionary text that could be derived from typographical features (examples, for instance, are always coded in italics). In addition, the supplementary volumes were to be integrated into the digitised dictionary. For this, the project returned to the data preparation phase following the online release of a first stage of data processing.

In contrast to the original plan to complete the retro-digitisation within five and a half years with the online release of the ODS, it has to be recognised that, although an end point for the necessary processes is theoretically possible, it has not been achieved in practice. This is undoubtedly the result of the Internet as a medium and its general capacity to enable quick corrections and additions to websites. We can also see from this example that further optimisation of a dictionary like the ODS – and also, of course, other types of Internet dictionary – can, in principle be undertaken, ad infinitum so that the digital lexicographic process in this kind of scenario is never complete.

3.5 Software to support lexicographic processes

Nowadays, lexicographic work is characterised by increasing automated support (cf., among others, Abel 2022, Abel/Klosa 2012, Rundell 2023, Rundell/Kilgarriff 2011). Modern lexicographic work is inconceivable without two tools, namely a dictionary writing system and a corpus query system.

3.5.1 Dictionary writing systems

A dictionary writing system is a software application that facilitates the lexicographic process and, preferably, optimises and streamlines it. It allows lexicographers to write a dictionary entry (data evaluation phase) and it makes both project management and publication easier (preparation phase for online release). A dictionary writing system usually has three components:

- a text-editing interface that allows the lexicographers to edit dictionary entries – this can be either a WYSISWYG ('what-you-see-is-what-you-get') view or simply an XML view (→ Fig. 3.7);
- a database to keep the data secure; relational databases like ORACLE, MySQL, and PostgreSQL as well as native XML databases are often used in a lexicographic context;
- a series of administration tools for project management and publication.

A dictionary writing system can be developed in house during the computerisation phase of a particular project (as is often the case for academic dictionaries such as the ANW), or a commercial system can be acquired. There is also the possibility to work with open-source systems, e.g. LEXONYMY (cf. Měchura 2017; Váradi et al. 2022).

The advantage of an in-house system is that it is easier to adapt it to the lexicographic process of a particular project and improvements can be made in-house, if necessary. However, this assumes that the project has the necessary human and financial resources at its disposal. The advantage of a commercial system is that many users will have contributed collectively to improving the system, which allows for the rapid development of new functions (such as a component added in 2023 to the system TLex that facilitates collaboration with ChatGPT, cf. de Schryver 2023: 2). Open-source systems also profit from the feedback that its many users give to the developers. Nonetheless, in commercial and open-source systems, the project is tied to the available data model (cf. Tiberius/Niestadt/Schoonheim 2014 and Abel 2022 for further discussion of the advantages and disadvantages of all options).

The more complex the administration and publication tools are that are offered by the system, the greater the control that the editors have over the lexicographic process. In this respect, the dictionary writing system developed in house for the *Norsk Ordbok* (NOB) offers an interesting feature. It makes it possible to monitor how much

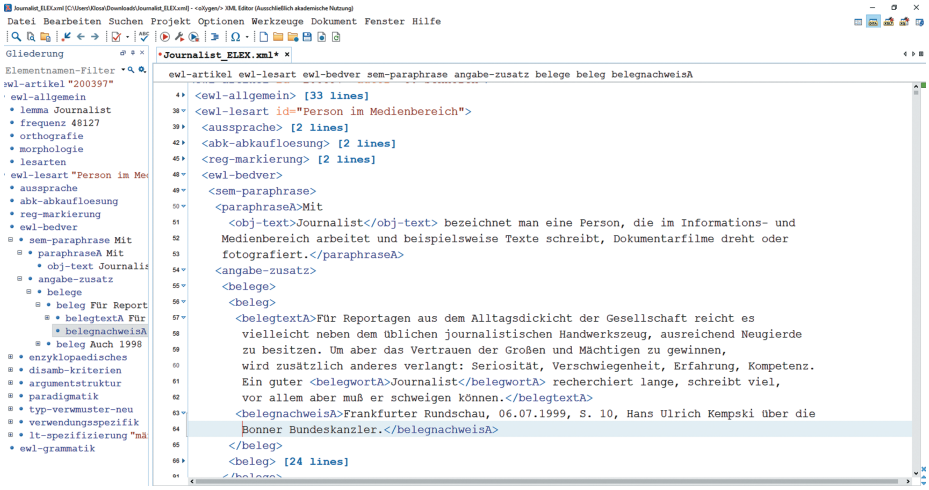


Fig. 3.7: The XML editor OXYGEN, which was used in the ELEXIKO project to prepare dictionary entries (here an extract from the lemma *Journalist* 'journalist').

text should be written for an article in order to maintain an appropriate distribution of article lengths in a particular alphabetical category. When a new word is selected for inclusion, a maximum length for the entry is suggested, based on the amount of data that is available at that point in time for compiling the whole dictionary. While the entry is being written and edited, the actual length of the entry is constantly compared with the maximum length that has been calculated so that the lexicographers can see precisely whether they have remained within the proposed length of the entry (Grønvik/Smith Ore 2013: 254).

3.5.2 Corpus query systems

Corpus query systems are used in the data preparation and analysis phase of the digital lexicographic process. Corpus query systems are tools with which corpus texts can be queried in ways that are linguistically relevant. Lexicographers are probably the most demanding corpus users – they need and regularly use the highest number of features in corpus tools. In fact, several features of corpus tools were originally designed especially for lexicographic purposes, only to be found useful by linguists, teachers, and others as well (Tiberius et al. 2022). SKETCH ENGINE is an example of a corpus query system which is often used in lexicographic projects (Kilgarriff et al. 2004), but there are many other corpus tools available (a comprehensive up-to-date list of these tools can be found at https://corpus-analysis.com/tag/concordancer).

The basic functionality that a corpus query system provides to support lexicography is (KWIC) concordancing, which displays all the occurrences of a keyword found

in the corpus with around 20 words of surrounding context. KWIC stands for "Keyword in Context" and refers to the concordance being displayed with the keyword in the middle of the screen. Most corpus query systems make it possible to sort and filter concordance lines and also, when necessary, to display more context.

Most systems also support a variety of powerful search queries for the lemma itself, a particular word form, or a phrase (in combination with a particular word class), right down to searching for all of the occurrences of a word in a specific lexical context. For example, it is possible to search for all occurrences of the verb *to speak* that can be found in a range of five words before or after the keyword *language* (as in *Which languages do you speak?*). A further feature that is particularly useful for lexicography (and that is available in advanced corpus tools) is the so-called lexical profile. A lexical profile is a statistical overview of the most important facts about a word and its customary combinations with other words (Atkins/Rundell 2008: 109). The word sketches in SKETCH ENGINE are a type of lexical profile, providing collocate lists per grammatical relation. In the German DWDS, it is not only lexicographers who use the "Wortprofil" ('word profile', its name for lexical profiles), but also users who are able to display the word profile of a word as part of the dictionary entry (→ Fig. 3.8).

Fig. 3.8: Word profile for the word *Buch* 'book' in the DWDS with collocates such as *lesen* 'read', *schreiben* 'write', *veröffentlichen* 'publish' or *dick* 'fat', *neu* 'new', etc.

Most functions in a corpus query system only work if the corpus data have been correctly processed in the computerisation phase. Processing corpus texts involves two steps: preparing the metadata and preparing the texts. The metadata contains information about the source such as the author, date, genre, and domain, which enables lexicographers to assign labels like "primarily spoken language" to particular lemmas with

greater confidence. Preparing the texts means adding linguistic annotations to the raw corpus texts, i.e. lemmatisation, tagging (annotation according to word class), or parsing (annotation according to syntactic structures) (→ Chapter 5).

Corpus software is improved and updated regularly on the basis of new insights and requirements that come from users. *Good Dictionary Examples* (GDEX), *Tickbox Lexicography* (TBL), *one-click copying*, *trend* analysis, and *word sense induction* in the word sketch are examples of this kind of functionality in SKETCH ENGINE which were added over time. GDEX is a function that seeks to automatically sort sentences in a concordance according to their usability as an example sentence in the dictionary based on a set of quantifiable heuristics, such as sentence length, frequency of words, and lists of vulgar words (Kilgarriff et al. 2008). In this way, the best example sentences will appear at the top of the concordance list, and these are the ones that lexicographers see first. TBL is a function through which lexicographers are able to select examples from a list of (good) candidates and export them directly into the editing system (→ Fig. 3.9).

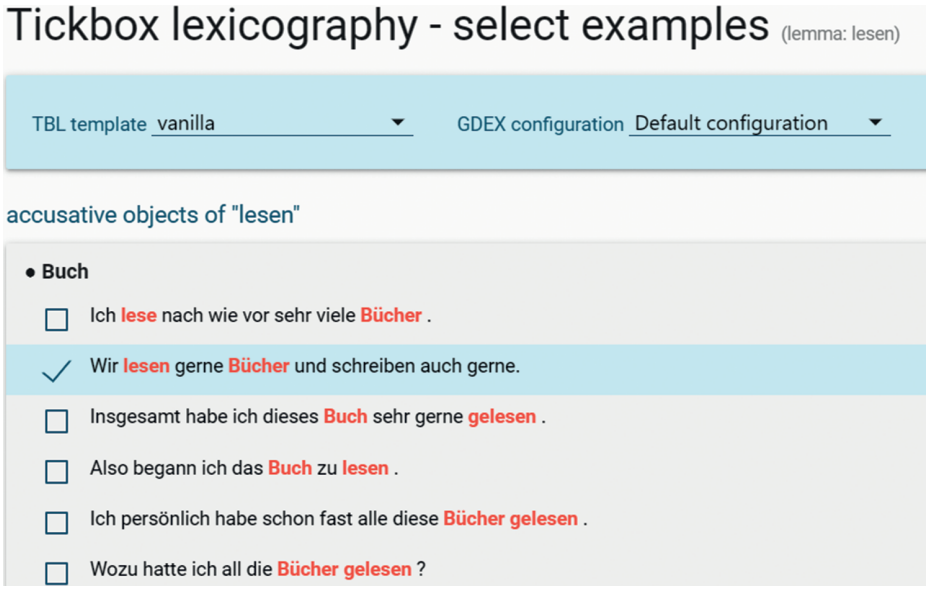


Fig. 3.9: Example of TickBox lexicography (for the verb *lesen* 'to read') in SKETCH ENGINE.

One-click copying is another useful feature for lexicographers. It supports easy transfer of data from the corpus query system into the editing system. Not only is the concordance line copied, but also the whole sentence (and potentially its metadata) is transferred onto the clipboard to be pasted into the editing system. *Trends* is a feature for detecting words that undergo changes in the frequency of use in time. It can be used to identify words whose use increases or decreases over time. *Word sense induction* is a functionality that has recently been added to the word sketch tool and that identifies

word senses automatically. This function categorises the collocations identified by a word sketch into groups corresponding to the different senses of a word.

Although it would be beneficial to have the dictionary writing system and the corpus query system integrated in one tool, they are often separate tools (although they are able to communicate with one another, as SKETCH ENGINE and the editing system used in the ANW project). There are only a few examples of systems in which the editing system and corpus query system are integrated in a single tool e.g. the TLEX system (Joffe/de Schryver 2004).

Choosing the most appropriate software tool to use for a new dictionary project may not be straightforward as there are many different dictionary writing systems and even more corpus query systems. See Kallas et al. (2019) for a list of systems that were mentioned in the context of the surveys that were carried out within the ELEXIS project to obtain an overview of existing lexicographic practices across Europe. These surveys also showed that in-house solutions are still very common for dictionary writing systems whereas for corpus query systems, commercial systems tend to be used most. The fundamental consideration in the choice of software is whether or not it meets all of the necessary requirements of the project. Basic aspects like price and the availability of academic licences can also play a role in choosing a system. Particularly in the planning phase, a thorough analysis of the available software is important in close collaboration between the lexicographic team and the IT experts in order to evaluate which work tasks can be carried out with existing software and which technical developments can be implemented in house under certain circumstances. Unfortunately, this kind of analysis is often absent, even now, in the planning stages of many dictionary projects.

3.6 The lexicographic process for dictionary portals

In this section, we consider the lexicographic process for dictionary portals (→ Chapter 2.4). Engelberg/Müller-Spitzer (2013) define a dictionary portal as a data structure that is represented as a website or a series of interlinked pages, that provides access to a series of electronic dictionaries, and where these dictionaries can be consulted as independent products. They distinguish between three main types that they refer to as dictionary collections, dictionary search engines, and dictionary networks.

Dictionary collections only give external access to the dictionaries in the portal. This means that they consist of links to the home pages of the dictionaries in the portal, and as a rule these dictionaries are not owned by the institution that runs the portal. The ERLANGER LISTE is one such dictionary collection.

Dictionary search engines are a little more sophisticated. They provide the option of searching all of the lemmas in the integrated dictionaries. However, the data in the various dictionaries are not interconnected with links. No dictionary content is presented within the portal, and the owners of the portal and the dictionaries are not

usually the same. The website OneLook and the European Dictionary Portal are examples of this kind of dictionary search engine.

Dictionary networks go one step further and make it possible for users to search for particular information in the entries in the dictionaries that are contained in the portal. Examples of a portal like this are OWID and the Trierer Wörterbuchnetz.

It is clear that the lexicographic processes for these three types of dictionary portal differ from one another. The lexicographic process for dictionary collections is the simplest. There is a preparatory phase, a data acquisition phase, possibly a short computerisation phase to set up the website, and a preparation phase for online release that involves writing the outer texts (what the portal is and what it is not). For a dictionary collection, the phases follow one another in a more or less linear way. First, it must be decided which dictionaries are to be included, then the website must be planned, and finally the portal has to be made available and maintained online. Once the dictionary collection is online, there are very few changes to make.

The lexicographic process for dictionary search engines is similar to the process for dictionary collections. However, the computerisation phase is more elaborate since a combined list of lemmas has to be compiled for all of the dictionaries in the portal, a search function implemented, and the data indexed. By clicking on a link, users move into an individual dictionary and leave the portal.

Dictionary networks have the most complex lexicographic process among dictionary portals. During the preparatory phase, decisions have to be taken about which dictionaries should be included in the portal and about the network connections between the dictionaries in the network. The dictionaries that are integrated in the network may be existing published dictionaries or dictionaries under construction. This influences the tasks that have to be completed in the data preparation and data evaluation phases. Here the computerisation phase involves interlinking the search engines and website. The preparation phase for online release involves writing the user guide.

Generally, consideration must be given as to whether we can refer to a digital lexicographic process at all in relation to dictionary portals and networks. Wiegand (1998: 134) defines such a process as being carried out so that one particular dictionary is created. However, compiling dictionary portals and networks does not involve a dictionary coming into existence but rather a website through which a variety of dictionaries can be accessed. Central tasks in the digital lexicographic process, e.g. corpus construction (data acquisition phase), processing corpus texts (computerisation phase), compiling a list of lemmas (data processing phase), and in particular writing dictionary entries (data evaluation phase) are absent here. Nevertheless, creating dictionary search engines and networks does involve tasks that require lexicographic competences, e.g. combining different lists of lemmas in an overarching access structure or conceiving and realising the linking between the individual contents of a dictionary network.

In this sense we tend towards extending Wiegand's (1998) definition: according to our understanding, a digital lexicographic process is the collection of activities belonging to the process that have to be undertaken so that a particular dictionary or a group

of dictionaries comes into being on or in an electronic carrier medium. This process can be divided into different phases in which computers are used consistently.

3.7 The lexicographic process of centralised lexicographic databases

The production of dictionaries is a meticulous and detail-oriented task that requires a great deal of time, effort, and expertise. In the light of shorter project running times as well as the changing role of lexicographic institutions and publishing houses (cf. Tiberius et al. 2023), which are becoming more of a data provider and less of a dictionary publisher, dictionary projects and publishers have started to move away from the production of stand-alone Internet dictionaries towards centralised lexicographic knowledge bases from which different end products (e.g. Internet dictionaries, dictionary apps, data for NLP tools) can be derived (for example, DUDEN dictionaries, cf. Alexa et al. 2002, or lexicographic institutions such as the Estonian Language Institute, cf. Tavast et al. 2018).

Having a single pool of data has several practical advantages. It supports the reusability of the lexicographic data for different end products, both dictionaries and language technology applications. It avoids having multiple dictionaries with duplicated and/or conflicting information and helps to minimise redundancies. Furthermore, it ensures consistency, which again leads to more efficient maintenance of the lexicographic data in a homogenous manner. On the other side of the coin, having a single pool of data also presents new challenges for lexicography. Integrating all lexicographic data into one centralised database naturally results in a more complex data model (cf. e.g. Tavast et al. 2018, Depuydt et al. 2019, and Gantar 2020 for the challenges faced at different institutions when creating a centralised lexicographic database).

In this new constellation, there seems to be not just one lexicographic process but at least two: one for the centralised lexicographic database and one for each end product that is compiled out of the data. The compilation of a central database is by default media-neutral, and requires a media-neutral lexicographic process in which all of the phases outlined in → Tab. 3.2 are still relevant, except for the phase of preparation for online release as publication is expected to occur in the process for individual end products. For the creation of the individual end products, a subset of the phases is needed. In particular, for each individual end product there will be a preparatory phase (to develop a concept for a specific end product), a preparation phase for online release, and a maintenance and preservation phase.

Furthermore, it should be noted that compiling data for a centralised lexicographic database changes the traditional organisation of the work: instead of editing per headword, generic content creation requires more task-based editing (e.g. editing morphological information for a set of headwords in the database). This requires a turnaround

in the way of working for lexicographers. New editing tools are also needed to support this task-based editing, allowing different views on the data and supporting a more relational encoding of lexicographic data (cf. Měchura et al. 2023, Meyer/Eppinger 2018). Finally, it remains to be seen how much work is needed to customise the content from such centralised lexicographic knowledge bases for specific end products.

In the context of recent developments, we tend towards extending Wiegand's (1998) definition of the lexicographic process to cover lexicographic databases (instead of dictionaries alone) to better reflect the digital aspects: a (completed) lexicographic process is the set of process activities carried out to create a particular lexicographic database. Lexicographic databases can be used in language technology, for creating dictionaries for human users, or can be integrated in a dictionary portal.

Bibliography

Further reading

Abel, Andrea (2022): Dictionary writing systems. In: Hanks, Patrick/de Schryver, Gilles-Maurice (eds.): *International Handbook of Modern Lexis and Lexicography*. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-45369-4_111-1. *An overview of lexicographic editing systems.*

Kilgarriff, Adam/Kosem, Iztok (2012): Corpus Tools for lexicographers. In: Granger, Sylviane/Paquot, Magali (eds.): *Electronic Lexicography*. Oxford: Oxford University Press, 31–55. *An overview of corpus tools for lexicographers.*

Müller-Spitzer, Carolin (2003): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: Lexicographica 19, 140–168. *A continuation and development of H.E. Wiegand's thinking for electronic dictionaries*.

Tiberius, Carole et al. (2023): *A Lexicographic Practice Map of Europe.* In: International Journal of Lexicography. https://doi.org/10.1093/ijl/ecad023. *The results of various surveys that give an insight into lexicographic processes*.

Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Berlin/New York: De Gruyter. *Chapter "1.5. Computer, wissenschaftliche Lexikographie und Wörterbuchforschung" (pp. 133–246) provides the first analysis and description of digital lexicographic processes*.

Literature

Academic literature

Abel, Andrea (2022): Dictionary writing systems. In: Hanks, Patrick/de Schryver, Gilles-Maurice (eds.): *International Handbook of Modern Lexis and Lexicography*. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-45369-4_111-1 [last access: April 26, 2024]

Abel, Andrea/Klosa, Annette (2012): Der lexikographische Arbeitsplatz – Theorie und Praxis. In: Fjeld, Ruth Vatvedt/Torjusen, Julie Matilde (eds.): *Proceedings of the 15th EURALEX International Congress in Oslo 2012*. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 413–421.

Abel, Andrea/Klosa, Annette (2014): *Einleitung: "Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess"*. Mannheim: Institut für Deutsche Sprache, 3–8.

Alexa, Melina, et al. (2002): The Duden Ontology: An Integrated Representation of Lexical and Ontological Information. In: *LREC Workshop on WordNet Structures and Standardisation, and How These Affect Wordnet Applications and Evaluation 2002*. Las Palmas, Gran Canaria. https://publica.fraunhofer.de/entities/publication/efe658dc-eda0-4915-ab40-1aeafb79c2c6/details [last access: April 26, 2024].

Atkins, B. T. Sue/Rundell, Michael (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Culy, Christopher/Lyding, Verena (2010): Double Tree: An Advanced KWIC Visualization for Expert Users. In: *Information Visualization, Proceedings of IV 2010, 2010 14th International Conference Information Visualization*, 98–103.

Depuydt, Katrien/Schoonheim, Tanneke/de Does, Jesse: *Towards a More Efficient Workflow for the Lexical Description of the Dutch Language*. http://videolectures.net/elexisconference2019_depuydt_dutch_language/ [last access: April 26, 2024].

de Schryver, Gilles-Maurice (2023): Generative AI and Lexicography: The Current State of the Art Using ChatGPT. In: *International Journal of Lexicography* 36(4), 355–387. https://doi.org/10.1093/ijl/ecad021 [last access: April 26, 2024].

de Schryver, Gilles-Maurice/Prinsloo, Daniel Jacobus (2000a): The Concept of "Simultaneous Feedback": Towards a New Methodology for Compiling Dictionaries. In: *Lexikos* 10, 1–31.

de Schryver, Gilles-Maurice/Prinsloo, Daniel Jacobus (2000b): Dictionary Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. In: Heid, Ulrich, et al. (eds.): *Proceedings of the Ninth EURALEX International Congress*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, 197–209.

Dubois, Claude (1990): Considérations générales sur l'organisation du travail lexicographique. In: Hausmann, Franz-Josef, et al. (eds.): *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin/New York: De Gruyter, 1574–1588.

Engelberg, Stefan/Müller-Spitzer, Carolin (2013): Dictionary Portals. In: Gouws, Rufus H., et al. (eds.): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Computational Lexicography*. Berlin/Boston: De Gruyter, 1023–1035.

Gantar, Polona (2020): *Dictionary of Modern Slovene: From Slovene Lexical Database to Digital Dictionary Database*. Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje, 46 (2), 589–602. https://doi.org/10.31724/rihjj.46.2.7 [last access: April 26, 2024].

Gouws, Rufus H. (2011): Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. In: Fuertes-Olivera, Pedro Antonio/Bergenholtz, Henning (eds.): *e-Lexicography. The Internet, Digital Initiatives and Lexicography*. London/New York: Continuum, 17–29.

Grønvik, Oddrun/Smith Ore, Christian-Emil (2013): What should the electronic dictionary do for you – and how? In: Kosem, Iztok, et al. (eds.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17–19 October 2013, Tallinn, Estonia*. Ljubljana/Tallinn: Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, 243–260.

Jakubíček, Miloš (2017): The advent of post-editing lexicography. In: *Kernerman Dictionary News*, 14–15.

Joffe, David/de Schryver, Gilles-Maurice (2004): TshwaneLex: A State-of-the-Art Dictionary Compilation Program. In: Williams, Geoffrey/Vessier, Sandra (eds.): *Proceedings of Eleventh EURALEX International Conference*. Le Paquebot: Lorient, 99–104.

Kallas, Jelena et al. (2019): *ELEXIS deliverable 1.1 Lexicographic Practices in Europe: A Survey of User Needs*. https://elex.is/wp-content/uploads/2020/06/Revised-ELEXIS_D1.1_Lexicographic_Practices_in_Europe_A_Survey_of_User_Needs.pdf [last access: April 26, 2024]

Kilgarriff, Adam (2008): GDEX: Automatically finding good dictionary examples in a corpus. In: Bernal, Elisande/De Cesaris, Janet (eds.): *Proceedings of the Thirteenth EURALEX International Congress*. Barcelona: Universitat Pompeu Fabra, 425–432.

Kilgarriff, Adam, et al. (2004): The Sketch Engine. In: Williams, Geoffrey/Vessier, Sandra (eds.): *Proceedings of Eleventh EURALEX International Conference*. Le Paquebot: Lorient, 105–116.

Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus H., et al. (eds.): *Dictionaries. An International Encyclopedia of Lexicography. Supplementary Volume: Recent Deve lopments with Focus on Computational Lexicography*. Berlin/Boston: De Gruyter, 501–508.

Knowles, Francis E. (1990): The Computer in Lexicography. In: Hausmann, Franz Josef, et al. (eds.): *Wörterbücher, Dictionaries, Dictionnaires. Ein internationales Handbuch zur Lexikographie*. Berlin/ New York: De Gruyter, 1645–1672.

Landau, Sydney (1984): *Dictionaries. The Art and Craft of Lexicography*. New York: Scribner's.

Lemberg, Ingrid (2001): Aspekte der Online-Lexikographie für wissenschaftliche Wörterbücher. In: Lemberg, Ingrid/Schröder, Bernd/Storrer, Angelika (eds.): *Chancen und Perspektiven computergestützter Lexikographie*. Tübingen: Niemeyer, 71–91.

Měchura, Michal (2017): Introducing Lexonomy: an open-source dictionary writing and publishing system. In: Kozem, Iztok, et al. (eds.): *Electronic lexicography in the 21st century: Proceedings of the eLex 2017 conference*. Brno: Lexical Computing, 662–679.

Měchura, Michal, et al. (2023): *Relations, relations everywhere: an introduction to the DMLex data model. Video presentation at eLex conference 2023*, *Brno, Czech Republic*, *27–29 June 2023*. https://www.youtube.com/ watch?v=b6Lkv-3D5C0 [last access: April 26, 2024].

Meyer, Peter (2011): 'vernetziko': A Cross-Reference Management Tool for the Lexicographer's Workbench. In: Kosem, Iztok/Kosem, Karmen (eds.): *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011*, *Bled, Slovenia*, *10.–12. November 2011*. Ljubljana: Trojina, Institute for Applied Slovene Studies, 191–198.

Meyer, Peter/Eppinger, Mirjam (2018): fLexiCoGraph: Creating and Managing Curated Graph-Based Lexicographical Data. In: Čibej, Jaka et al. (eds.): *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts, Ljubljana*, *17–21 July 2018*. Ljubljana University Press, Faculty of Arts, 1017–1022.

Meyer, Christian M./Gurevych, Iryna (2016): Der lexikografische Prozess im deutschen Wiktionary. In: Hildenbrandt, Vera/Klosa, Annette (eds.): *Der lexikografische Prozess bei Internetwörterbüchern. 4. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*. Mannheim: Institut für Deutsche Sprache, 55–68.

Müller-Spitzer, Carolin (2013): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: *Lexicographica* 19, 140–168.

Niestadt, Jan (2019): De ANW-artikeleditor: software als strategie. In: Beijk, Egbert, et al. (eds.): *Fons verborum. Feestbundel voor prof. dr. A.F.M.J. (Fons) Moerdijk, aangeboden door vrienden en collega's bij zijn afscheid van het Instituut voor Nederlandse Lexicologie*. Leiden/Amsterdam: Instituut voor Nederlandse Lexicologie/Gopher BV, 215–222.

Riedel, Hans/Wille, Margit (1979): *Über die Erarbeitung von Lexika*. Leipzig: Bibliographisches Institut.

Rundell, Michael/Kilgarriff, Adam (2011): Automating the creation of dictionaries: where will it all end? In: Meunier, Fanny, et al. (eds.): *A Taste for Corpora. A tribute to Professor Sylviane Granger*. Amsterdam: Benjamins, 257–281.

Rundell, Michael (2023): Automating the creation of dictionaries: are we nearly there? In: *Proceedings of Asialex 2023*, *June 22–24*. Seoul: Yonsei University Seoul, 9–17.

Schaeder, Burkhard (1987): *Germanistische Lexikographie*. Tübingen: Narr.

Schröder, Martin (1997): Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projektes "Deutsches Dialektwörterbuch". In: *Zeitschrift für Dialektologie und Linguistik* 64/1, 57–65.

Storrer, Angelika (1998): Hypermedia-Wörterbücher: Perspektiven für eine neue Generation elektronischer Wörterbücher. In: Wiegand, Herbert Ernst (ed.): *Wörterbücher in der Diskussion III*. Tübingen: Niemeyer, 106–131.

Storrer, Angelika (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, Ingrid/Schröder, Bernd/Storrer, Angelika (eds.): *Chancen und Perspektiven computergestützter Lexikographie*. Tübingen: Niemeyer, 54–69.

Storrer, Angelika/Freese, Karin (1996): Wörterbücher im Internet. In: *Deutsche Sprache* 24, 97–136.

Svensén, Bo (2009): *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Tavast, Arvi, et al. (2018): Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. In: Čibej, Jaka, et al. (eds.): *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts, Ljubljana, 17–21 July 2018*. Ljubljana University Press, Faculty of Arts, 749−761.

Tiberius, Carole/Krek, Simon (2014): *Workflow of Corpus-Based Lexicography. Deliverable COST-ENeL-WG3 meeting*. https://www.elexicography.eu/wp-content/uploads/2015/04/LexicographicalWorkflow_DeliverableWG3BolzanoMeeting2014.pdf [last access: April 26, 2024].

Tiberius, Carole/Niestadt, Jan (2015): Dictionary Use: A Case Study of the ANW dictionary. In: Tiberius, Carole/Müller-Spitzer, Carolin (eds.): *Research into dictionary use/Wörterbuchbenutzungsforschung. 5. Arbeitsbericht des wissenschaftlichen Netzwerks "Internetlexikografie"*. Mannheim: Institut für Deutsche Sprache, 28–35.

Tiberius, Carole/Niestadt, Jan/Schoonheim, Tanneke (2014): The INL Dictionary Writing System. In: Kosem, Iztok/Rundell, Michael (eds.): *Slovenščina 2.0: Lexicography* 2:2, 72–93.

Váradi et al. (2022): Váradi, Tamás et al. (2022): *Lexonomy: Mastering the ELEXIS Dictionary Writing System. Version 1.1.0*. DARIAH-Campus [training module]. https://campus.dariah.eu/id/Qm_2SzS_rGB-Py9YTCHYm [last access: April 27, 2024].

Wiegand, Herbert Ernst (2018): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie. 1. Teilband*. Berlin/New York: De Gruyter.

Zgusta, Ladislav (1971): *Manual of Lexicography*. The Hague/Paris: Mouton.

Dictionaries

ANW = *Algemeen Nederlands Woordenboek*. Leiden: Instituut voor Nederlandse Lexicologie. https://anw.ivdnt.org/ [last access: April 27, 2024].

DRWB = *Deutsches Rechtswörterbuch*. Heidelberg: Heidelberger Akademie der Wissenschaften. http://drw-www.adw.uni-heidelberg.de/drw/ [last access: April 27, 2024].

DWB = Deutsches Wörterbuch von Jacob und Wilhelm Grimm online. In: *Wörterbuchnetz des Trier Center for Digital Humanities/Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier*. http://woerterbuchnetz.de/DWB/ [last access: April 27, 2024].

DWDS = *Das Digitale Wörterbuch der deutschen Sprache*. Berlin-Brandenburgische Akademie der Wissenschaften. http://www.dwds.de [last access: April 27, 2024].

ELEXIKO = Online-Wörterbuch zur deutschen Gegenwartssprache. In: *OWID – Online Wortschatz-Informationssystem Deutsch*. Mannheim: Institut für Deutsche Sprache. http://www.elexiko.de [last access: April 27, 2024].

ERLANGER LISTE = *Lexika & Wörterbücher*. http://www.erlangerliste.de/ressourc/lex.html [last access: April 27, 2024].

NOB = *Norsk Ordbok*. Oslo: Universitetet i Oslo, Institutt for lingvistike og nordiske studier. http://no2014.uio.no/l/ordbok/no2014.cgi [last access: April 27, 2024].

ODS = Ordbog over det Danske Sprog. In: *ordnet.dk*. *Dansk Sprog i Ordbøger og Korpus*. *Den Danske Sprog- og Litteraturselskab*. http://ordnet.dk/ods [last access: April 27, 2024].

OED = *Oxford English Dictionary online*. Oxford: Oxford University Press. http://dictionary.oed.com [last access: April 27, 2024].
OneLook = *OneLook Dictionary Search*. http://www.onelook.com [last access: April 27, 2024].
ordnet.dk = *Dansk Sprog i Ordbøger og Korpus. Den Danske Sprog- og Litteraturselskab*. http://ordnet.dk [last access: April 27, 2024].
OWID = *OWID Online-Wortschatz-Informationssystem Deutsch*. Mannheim: Institut für Deutsche Sprache. Online: http://www.owid.de/ [last access: April 27, 2024].
SAOB = *Svenska Akademiens Ordbok*. Stockholm: Svenska Akademien. http://g3.spraakdata.gu.se/saob/index.html [last access: April 27, 2024].
Trierer Wörterbuchnetz = *Wörterbuchnetz des Trier Center for Digital Humanities/Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier*. http://woerterbuchnetz.de/ [last access: April 27, 2024].
Wiktionary = *Wiktionary, das freie Wörterbuch*. http://de.wiktionary.org/wiki/Wiktionary:Hauptseite [last access: April 27, 2024].
WNT = *Woordenboek der Nederlandsche Taal*. Leiden: Instituut voor Nederlandse Lexicologie. https://gtb.ivdnt.org/ [last access: April 27, 2024].

Internet sources

AGD = *Archiv für Gesprochenes Deutsch*. Mannheim: Institut für Deutsche Sprache. http://agd.ids-mannheim.de/index.shtml.
DeReKo = Deutsches Referenzkorpus. Mannheim: Institut für Deutsche Sprache. Online: http://www.ids-mannheim.de/kl/projekte/korpora.html 8last access: April 27, 2024].
DMLex = *OASIS Lexicographic Infrastructure Data Model and API (LEXIDMA)*. https://groups.oasis-open.org/communities/tc-community-home2?CommunityKey=0fd41fbb-72be-4771-8faf-018dc7d3f419 [last access: April 27, 2024].
Double Tree JS = *Double TreeJS: A compact view or words in context*. http://linguistics.chrisculy.net/lx/software/DoubleTreeJS/index.html [last access: April 27, 2024].
Elexis = *European lexicographic infrastructure*. https://elex.is/ [last access: April 27, 2024].
European Dictionary Portal = *European Dictionary Portal*. http://www.dictionaryportal.eu [last access: April 27, 2024].
Git = *Git Tool zur Versionenkontrolle*. https://git-scm.com [last access: April 27, 2024].
ISO LMF = *ISO Lexical Markup Framework*. https://www.iso.org/standard/68516.html [last access: April 27, 2024].
ISO 1951:2007 = *ISO 1951:2007. Presentation/representation of entries in dictionaries*. https://www.iso.org/standard/36609.html [last access: April 27, 2024].
Lexonymy = *Lexonymy – online, open-source platform for writing and publishing dictionaries*. https://www.lexonomy.eu/ [last access: April 27, 2024].
MySQL = *MySQL Open-Source-Datenbank*. https://www.mysql.de/ [last access: April 27, 2024].
Ontolex-Lemon = *Lexicon Model for Ontologies*. https://www.w3.org/2016/05/ontolex/ [last access: April 27, 2024].
ORACLE = *ORACLE Datenbank*. https://www.oracle.com/de/index.html [last access: April 27, 2024].
oXygen = *oXygen XML Editor*. https://www.oxygenxml.com/ [last access: April 27, 2024].
Pixelio = *pixelio.de – Deine kostenlose Bilddatenbank für lizenzfreie Fotos*. http://www.pixelio.de [last access: April 27, 2024].
PostgreSQL = *PostgreSQL Datenbankmanagementsystem*. http://www.postgresql.org/ [last access: April 27, 2024].

Sketch Engine = *Sketch Engine*. https://www.sketchengine.co.uk/ [last access: April 27, 2024].
TEI Lex-0 = *TEI Lex-0. A baseline encoding for lexicographic data*. https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html [last access: April 27, 2024].
TLex = *TLex Suite: Dictionary Compilation Software*. https://tshwanedje.com/tshwanelex/ [last access: April 27, 2024].
Wikimedia = *Wikimedia Commons, das freie Medienarchiv*. http://commons.wikimedia.org [last access: April 27, 2024].

Images

Fig. 3.1 "Robotereinsatz am Fließband": KUKA Systems GmbH [CC BY-SA 3.0-de (http://creativecommons.org/licenses/by-sa/3.0/de/deed.elegen.)], via Wikimedia Commons (http://commons.wikimedia.org/wiki/Category:KUKA_robots?uselang=de).

Fig. 3.1 "Werkzeugschlosser bei der Arbeit": Deutsche Fotothek [CC-BY-SA-3.0-de (http://creativecommons.org/licenses/by-sa/3.0/de/deed.en)], via Wikimedia Commons (http://upload.wikimedia.org/wikipedia/commons/7/75/Fotothek_df_roe-neg_0006486_016_Portr%C3%A4t_eines_Arbeiters_an_einer_Bohrmaschine%2C_W.jpg).