Lexicography – Introduction

Blockseminar "e-lexicography"
Univ. Potsdam
Geyken / Würzner

Blockseminar e-lexicography: overview

Day 1:

- 10-10.30: General Introduction (AG&KMW)
- 10.30-11.30: Brief History of (German) Lexicography
- 11.45 12.30 visit of the "Grimm Library" (Jörg Schröder)
- 13.00: Corpora, annotation (KMW)
- 14.30-16.00 Corpus Annotation with Python (1)

(KMW)

Day 2:

- 10-12.30: Corpus Annotation with Python (2) (KMW)
- 13.15: Components of DWDS system and Workflow (AG)
- 14.45: TEI-Lex0 : a new format for dictionaries (Axel Herold) – part 1

Blockseminar e-lexicography: overview

Day 3:

- 10.00: Crawling corpora on the web (Adrien Barbaresi)
- 13:00 TEI-Lex0 : a new format for dictionaries (Axel Herold) – part 2
- 15.00 Writing the entry:
 Definition style and templates
 (Sebastian Göttel)

Day 4:

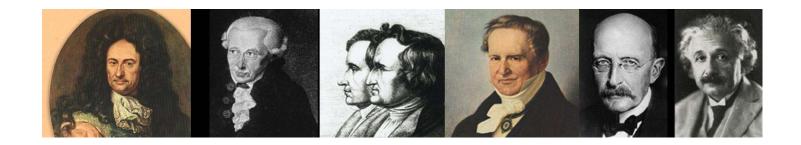
- 10:00: corpora and collocations: the linguistic perspective
- 11.00-12.00 Discussion of the Oxford Guide to Practical Lexicography
- 13:00: Q&A corpus crawling (Adrien Barbaresi)
- 14-16: Term paper preparation (AG, KMW, AH)

Berlin-Brandenburg Academy of Sciences and the Humanities



Predecessors of BBAW

- Founded in 1700 (Leibniz)
- Internationally renowned as Prussian Academy of Sciences
- Among its members were: Kant, Grimm, Humboldt, Planck, Einstein...



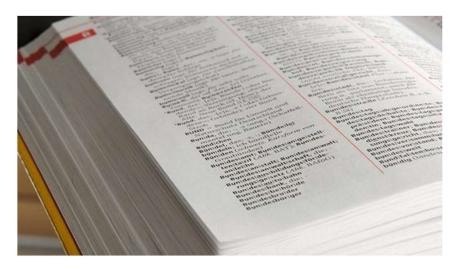
BBAW - today

- Largest extra universitarian research institute with a background in the humanities in the region of Berlin
 - ~200 scholars (members)
 - research departments: around 30 long term projects
 - organized in Clusters: Centre of language, Centre of the Old World
 - o dictionaries (Goethe, DWDS), editions (Marx-Engels, Kant, Leibniz)...
 - Since January 2019: BBAW is coordinator of the national "Centre for digital lexicography of the german language (Partners: 3 academies in Göttingen, Mainz, Leipzig, IDS, Goethe-Institut)

What is lexicography?

Lexicography vs. Lexicology

lexicon, short for lexikón biblíon [gr], nominalization of the adj. lexikós, "wrt a word, an idiom", and biblíon, "diminutiv of" zu bíblos, "papyrus [/pəˈpaɪrəs/], paper; book, writing". literally: "small book of words and idioms"



Origin: A borrowing from

Latin. **Etymon:** Latin *dictionarius*.

First appearance: 1480

Etymology: < post-classical

Latin *dictionarius* wordbook, collection of phrases (*c*1220 as the name of a textbook for learners of Latin; also *dictionarium* (neuter) as the name of an alphabetized encyclopedic guide to the Vulgate Bible (late 14th cent.); from 15th cent. in both forms in sense 'alphabetized wordbook') < classical Latin *diction-*, *dictio* <u>DICTION *n.*</u> + -ārius -ARY suffix1.

Source: www.oed.com, entry dictionary

a. A book which explains or translates, usually in alphabetical order, the words of a language or languages (or of a particular category of vocabulary), giving for each word its typical spelling, an explanation of its meaning or meanings, and often other information, such as pronunciation, etymology, synonyms, equivalents in other languages, and illustrative examples. Cf. LEXICON *n.*, WORDBOOK *n.*

Source: www.oed.com, entry dictionary, definition 1a.

a. A book which explains or translates, usually in alphabetical order, the words of a language or languages (or of a particular category of vocabulary), giving for each word its typical spelling, an explanation of its meaning or meanings, and often other information, such as pronunciation, etymology, synonyms, equivalents in other languages, and illustrative examples. Cf. LEXICON *n.*, WORDBOOK *n.*

c1480 Medulla Gram. (Pepys) f.

32^v *Dixionari*[us],..an[gli]ce Dixionare.

1574 Accts. Treasurer Scotl. f. 348 [Ane] dictionar in latene and frence.

1580 Edinb. Test. VIII. f. 109, in Dict. Older Sc.

Tongue at *Dictionar(e* Certane bukes bund & vnbund, sik as ane dictionar in latyne & inglis.

1643 in J. Stuart Extracts Council Reg. Aberdeen (1872) II.

9 The paines takin be Mr John Row..for setting furth ane Hebrew dictionar, and dedicating the same to the counsell.

Source: www.oed.com, entry dictionary, definition 1a.

The earliest books to be referred to as dictionaries in English were those in which the meanings of the words of one language or dialect were given in another (or, in a polyglot dictionary, in two or more languages). Dictionaries (thus named) of this type began to appear in England during the 16th cent., initially of Latin, later of modern languages (see quots. 1538 at β ., 1547 at β . respectively), although of course such works had been compiled and disseminated under other names long before this (see etymology for information about cognate words in other European languages). During the 17th cent. dictionary came also to be used of works giving explanations in English of 'hard words', of which the earliest to be printed was Robert Cawdrey's *Table* Alphabeticall of 1604; the earliest to include the word dictionary in the title was Henry Cockeram's of 1623. Later dictionaries extended the range of words covered to include more of the common words of the language.

Source: www.oed.com, entry dictionary, section comment

The word has not been widely used of books in which the words are not arranged alphabetically, although one such book, an English-Latin vocabulary for schoolchildren in which the words were entered under subject headings, was issued by John Withals in 1553 under the title 'A shorte dictionarie for yonge begynners' and went through numerous editions during the 16th and 17th centuries. See also THESAURUS n. 2. For works dealing with the vocabulary of a particular text or dialect *glossary* is now more commonly used; cf. also VOCABULARY n. 1a.

collegiate, desk, learner's, rhyming dictionary, etc.:

Source: www.oed.com, entry dictionary, section comment

What is lexicography?

Lexicography vs. Lexicology

lexicon, lit.: "small book of words and idioms"



Types of dictionaries:

- Monolingual: orthography, pronunciation, grammar, syntax, semantics, pragmatics
- Learner dictionaries
- Dictionaries of technical domains
- Bilingual
- Historical vs. synchronic
- Print vs. electronic
- Dictionary vs. encyclopedia

History
17/18th century
19th century
21st century: the digital turn for dictionaries

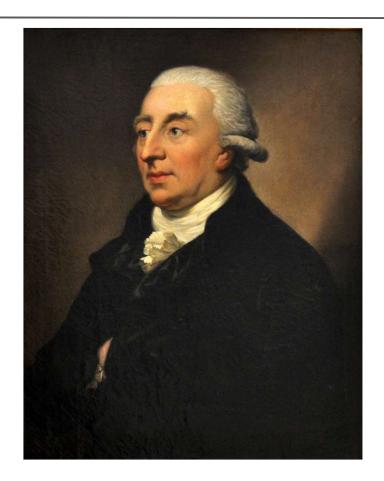
Brief history of German dictionaries

1781: Adelung *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart.*

1838-1961: Jacob und Wilhelm Grimm *Deutsches Wörterbuch* an.

1860 - 1865, Daniel Sanders: Wörterbuch der deutschen Sprache.

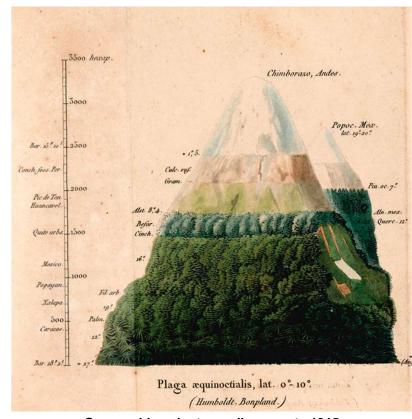
1880 Konrad Duden: Orthographisches Wörterbuch.



19th century: empiricism, evolution of sciences and large national dictionaries

Based on the scientific ideal of positivism the surveying and systematization of the world becomes a task of the sciences:

- Supported by technical progress and significant improvements in navigation and timing
- Collection, sketching, surveying geographical and biological information
- Research trips and major expeditions (Charles Darwin, Alexander von Humboldt)
- Dictionaries as an instance of surveying efforts at the linguistic level



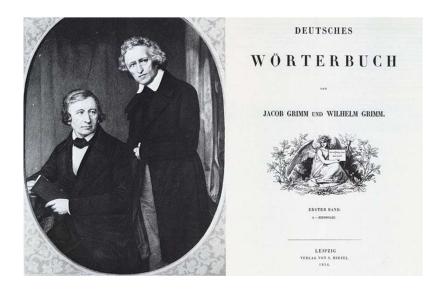
Geographiae plantarum lineamenta 1815

Lexikography in the 19th century

Unlike today, lexicography was not yet a discipline in its own right.

In principle, lexicographers were "universal scholars" who dealt with the most diverse fields of knowledge.

With increasing national consciousness and the need for a nation state, the need for a national dictionary arose.



Lexikography in the 19th century

Source corpus:

mostly "exemplary" writers: Goethe, Schiller, Uhland

due to "historically oriented" linguistics but also older sources

from Luther to his contemporary Goethe

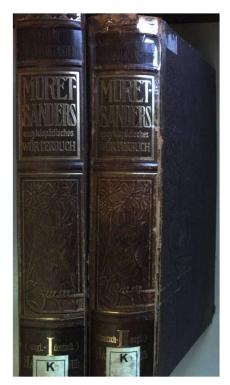
...and how did Daniel Sanders work?

Who?

Daniel Hendel Sanders (1819 – 1897)

- Study of Mathematics and Philology in Berlin (1839-1842)
- Doctorate in mathematics (1842)
- Director of the Jewish free school in Strelitz (1842-1852)
- Criticism of Grimm's dictionary
- Offer from the publisher O. Wigand to publish an independent dictionary
- Publication of the Dictionary of the German Language (1860-1865)
- freelance lexicographer, writer, translator, pedagogue, poet
- Participation in the First Orthographic Conference, among others together with Konrad Duden
- numerous publications on the German language, last work: Muret-Sanders





(1901)



(2013)

Lexikography in the 19th century

Corpus:

for Sanders mostly New High German, therefore contemporary, but also supra-regional and interdisciplinary

dein Handwerkszeug u. s. w. Du hast zu deinem Werk offenbar mancherlei Stoff gebraucht. Wer hat ihn dir geliefert oder woher hast du ihn genommen?

Sanders, Daniel: Aus der Werkstatt eines Wörterbuchschreibers. Plaudereien. Berlin, 1889, S. 23.

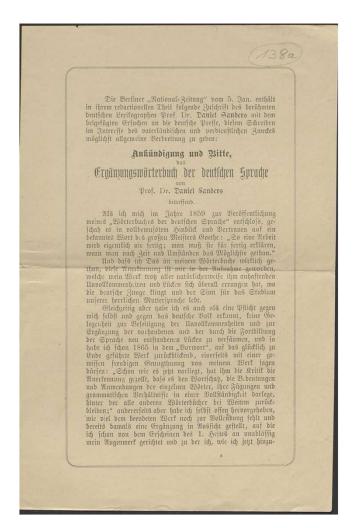
Lexikography in the 19th century

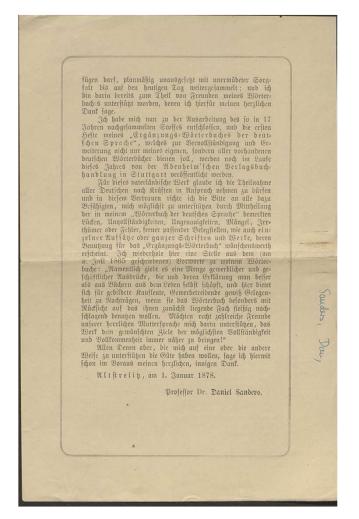
Quellenkorpus:

```
Tärche, f.; -n; -n=, Lärch=: eine Art Fichte, Pinus larix (mhd. larche, lerchboum). Börne 1, 193; freiligrath H. 93; G. 23, 14; Oken 3, 349; Stumpf 605b 2c. — -n, a.: von Lärchenholz.
```

"Deiner Wurzeln gib, o Tamaract! Deiner Wurzelfasern, Lärche! Meinen Kahn damit zu binden, Seine Enden so zu binden,

Freiligrath, Ferdinand: Der Sang von Hiawatha. Stuttgart, 1857, S. 93.





Aufruf Daniel Sanders' in der Berliner National-Zeitung vom 5ten Januar 1878

Ankündigung und Bitte,

bas

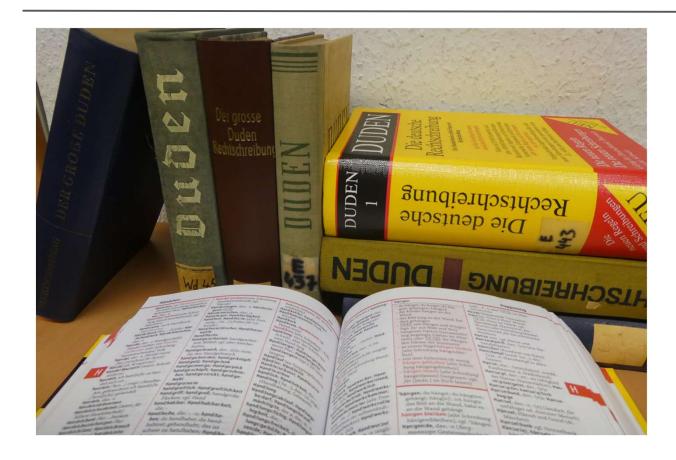
Ergänzungswörterbuch der deutschen Sprache

non

Prof. Dr. Daniel Sanders

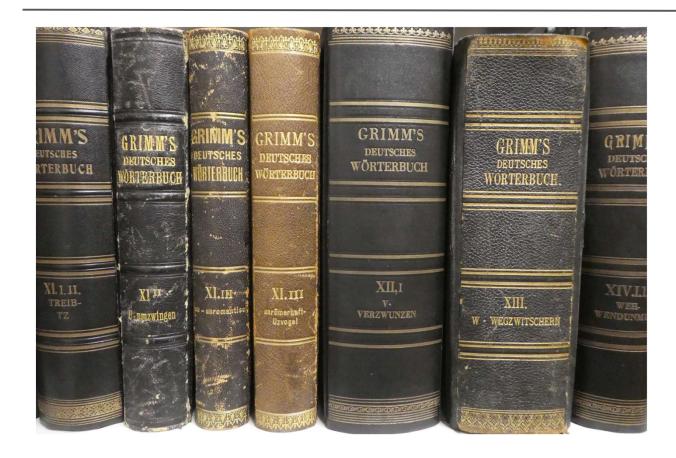
Für dieses vaterländische Werk glaube ich die Theilnahme aller Deutschen nach Kräften in Anspruch nehmen zu dürfen und in diesem Vertrauen richte ich die Vitte an alle dazu Vefähigten, mich möglichst zu unterstützen durch Wittheilung der in meinem "Wörterbuch der deutschen Sprache" bemerkten Lücken, Unvollständigkeiten, Ungenauigkeiten, Vängel, Krrsthümer oder Fehler, ferner passender Velegstellen, wie auch einzelner Aufsähe oder ganzer Schriften und Werke, deren Venutung für das "Ergänzungs-Wörterbuch" wünschenswerth erscheint. Ich wiederhole hier eine Stelle aus dem (am S. Juli 1865 geschriebenen) Vorworte zu meinem Wörterbuche: "Namentlich giebt es eine Wenge gewerblicher und gesschäftlicher Ausdrücke, die und deren Erklärung man besser als aus Büchern aus dem Leben selbst schöpft, und hier bietet

Lexicography in the 20th century: the example of Duden



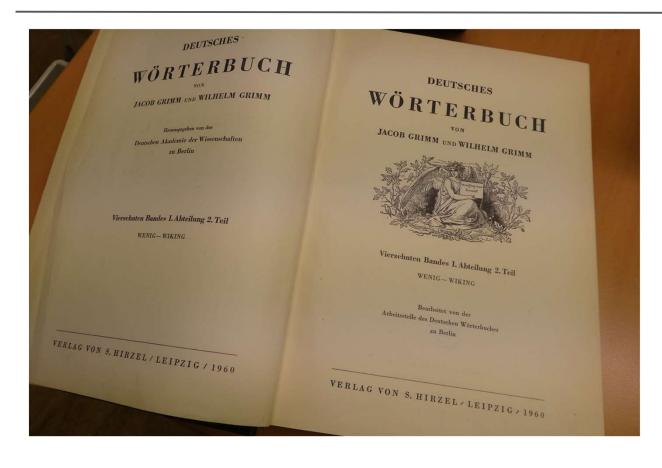
Im 20. Jahrhundert explodiert die Produktion von Nachschlagewerken. Erfolgreiche Wörterbücher wie der Duden, deren Anfänge im 19. Jahrhundert liegen, werden fortgeführt und prägen Generationen von Nutzern.

Lexicography in the 20th century: Deutsches Wörterbuch



Zu den großen Werken, die im 20. Jahrhundert fortgeführt werden, gehört das von Jacob und Wilhelm Grimm begonnene Deutsche Wörterbuch (DWB).

Lexikographie im 20. Jahrhundert: Deutsches Wörterbuch



The work begun by the Brothers Grimm was not finished until 1961, after 123 years of eventful history.

Lexikographie im 20. Jahrhundert: Deutsches Wörterbuch, Neubearbeitung

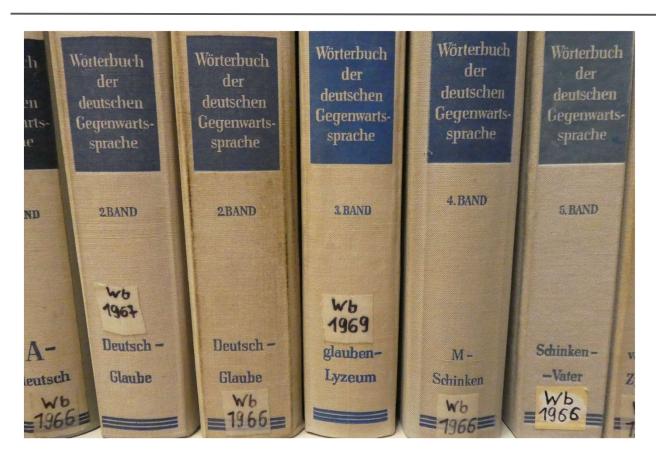


The first letters of the work, which was already largely obsolete when it was completed, were reworked in Berlin and Göttingen in 1957.

The Berlin office was located in what is now the Berlin-Brandenburg Academy of Sciences and Humanities.

After the publication of the last planned delivery in 2018, the dictionary will no longer be continued.

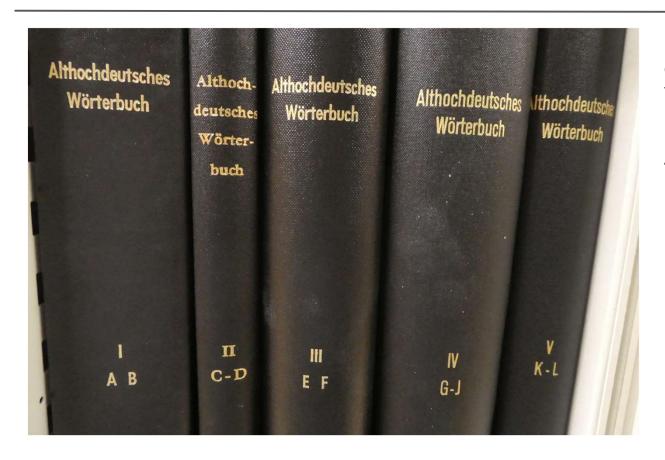
Lexicography in the 20th century: Wörterbuch der deutschen Gegenwartssprache



Neben das Deutsche Wörterbuch als historisches
Bedeutungswörterbuch tritt in der DDR das Wörterbuch der deutschen Gegenwartssprache (WDG), das mit seiner modernen Konzeption Vorbild auch für andere Wörterbuchprojekte geworden ist.

Der Inhalt des WDG gehört später zum Kernbestand des heute die lexikographische Arbeit an der Berliner Akademie fortführenden Digitalen Wörterbuchs der deutschen Sprache (DWDS).

Lexicography in the 20th century: from OHG to modern German



In the 20th century, intensive efforts were devoted to the thorough lexicographical recording of the older language levels of German:
Althochdeutschen (750 - 1050), Mittelhochdeutschen (1050 - 1350) und Frühneuhochdeutschen (1350 - 1650).

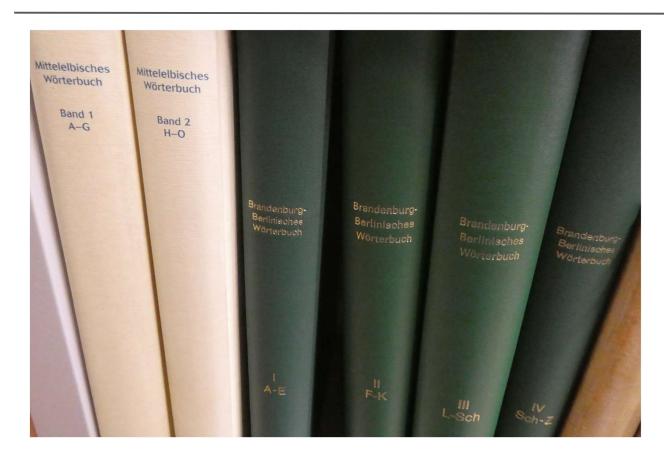
Lexicography in the 20th century: dialect dictionaries

Sudmeißn., Chtz (3 Bel.). °D; Nischel m. derb '/Kopf' 1. als Körperteil 'jm des Menschen, allg.; du krischst glei eene Zure vor'n Nischel Wurz, ähnl. verstr.; wie de wie-°ZW der im a (um den) Nischel gihst, a richtger Strubbelkupp! Görl; mir brummt der Nischel mol 'ich habe Kopfschmerzen' verstr.; in Spottan strophen auf Einwohner benachbarter Ortben, schaften: in Wolkenstaa (Wolkenstein) hom °Zsc se kromme Baa, hom se gruße Nischeln wie Mei de Reisigbischeln °Zschop Ehr, ähnl. verstr. (Bäi Westerzg.; aff den sann Nischel ka mer Reisich hacken (von einem Dickkopf) Rei very ähnl. verstr. Vgtld.: d (der

regional dialect dictionaries are continued or begun (here an excerpt from the dictionary of the Upper Saxon dialects).

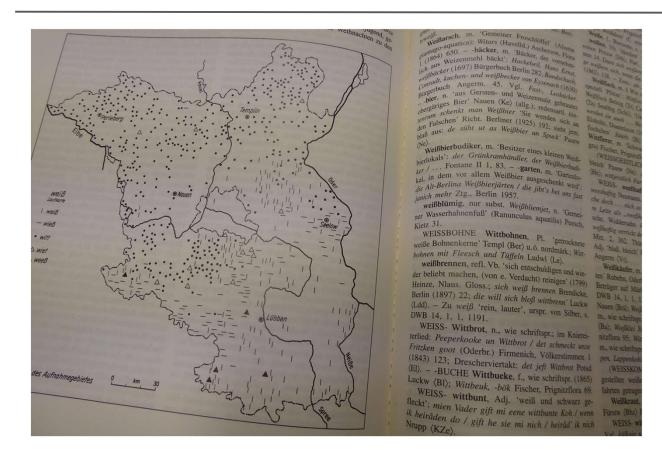
Most of these collections can no longer be made up due to the loss of many local dialects

Lexicography in the 20th century: Brandenburg-Berlinisches Wörterbuch



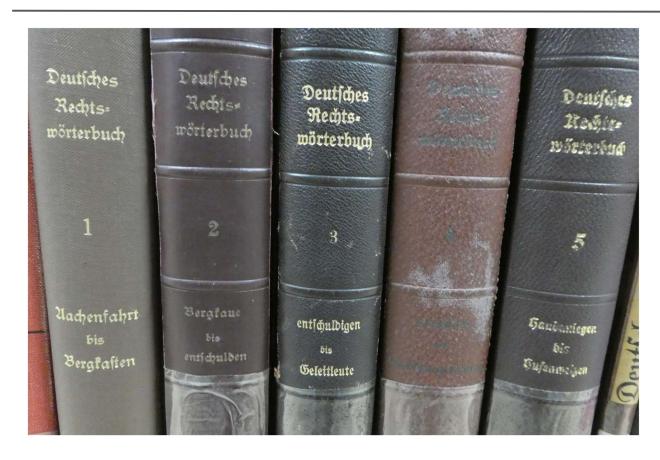
For example, the Brandenburg-Berlin Dictionary was created at the Berlin-Brandenburg Academy of Sciences and Humanities (and its predecessor, the Academy of Sciences of the GDR).

Lexicography in the 20th century: Brandenburg-Berlinisches Wörterbuch



The Brandenburg-Berlin
Dictionary covers the vocabulary
of the former province of
Brandenburg.

Lexicography in the 20th century: Deutsches Rechtswörterbuch



Special dictionaries document the various areas of vocabulary. The German Legal Dictionary, whose first deliveries appeared at the beginning of the 20th century, collects and comments, for example, on the historical German legal vocabulary.

Lexicography in the 20th century: file cards



The typical working basis of all dictionaries up to the beginning of automation and digitisation at the end of the 20th century was a corpus of paper which, in the case of large dictionaries such as the German Dictionary, could contain many millions of slips of paper on which individual vouchers for a word were noted.

Cf. visit of the "Grimm library"

Lexicography in the 21st century

Surveying and documentation of language(s) via very large digital collections of mostly textual data (e.g. text corpora, corpus-based lexicography)

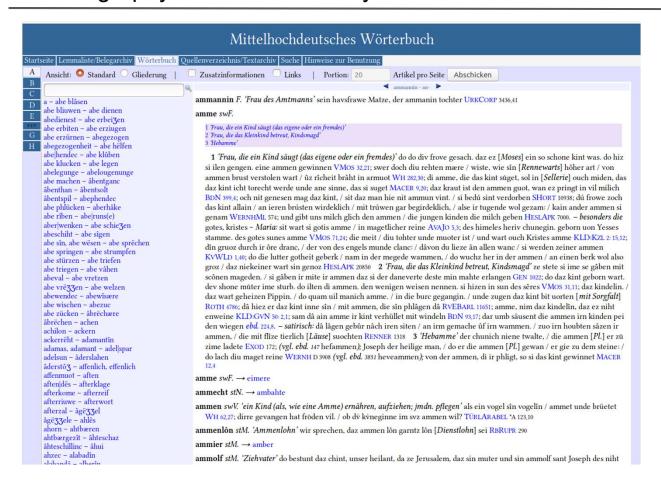
Tools of computational linguistics help lexicographers to find examples of linguistic phenomena to be described by them ("the needle in the haystack").

The commitment and involvement of users has become closer in the times of the Internet (own research in the dictionary base, quick reaction to reporting errors and gaps).

Presentation of the data on the Internet as largely decoupled from the "physical" printed work:

- Individual work (e.g. Middle High German dictionary)
- Dictionary portal (networking and querying many dictionaries, e.g. Trier dictionary network, OWID)
- Lexical information system, access to lexical knowledge in various forms (dictionary, corpora, statistical evaluations, e.g. DWDS)

Lexicography in the 21st century: Mittelhochdeutsches Wörterbuch, Akademie Mainz



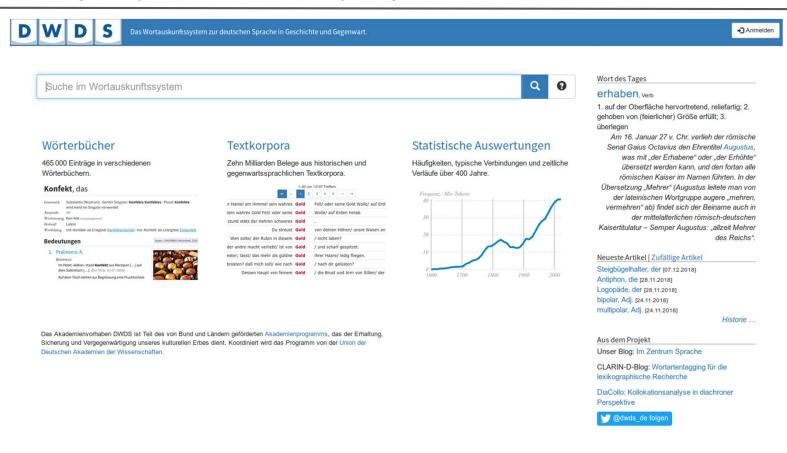
Lexicography in the 21st century: Wörterbuchnetz, Universität Trier



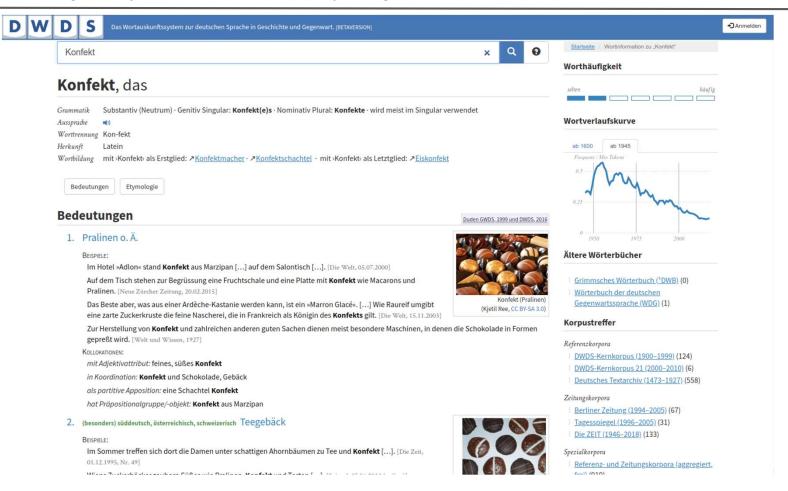
Lexicography in the 21st century: OWID am Leibniz-Institut für Deutsche Sprache



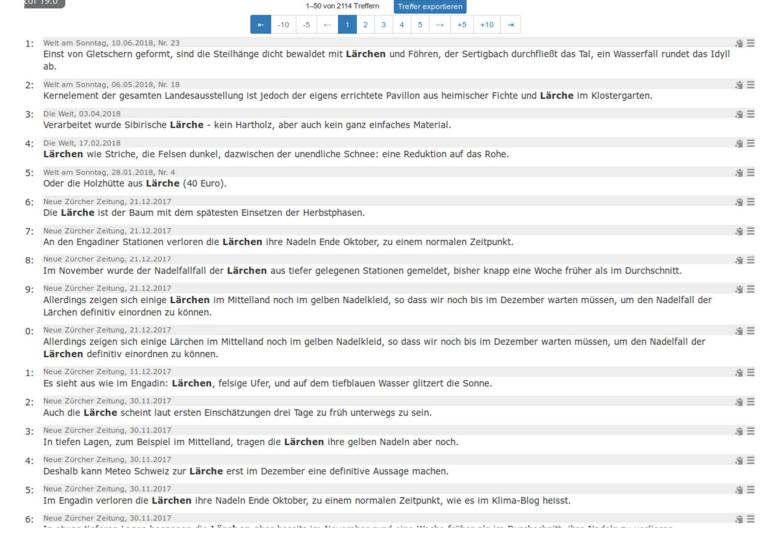
Lexicography in the 21st century: Digitales Wörterbuch der deutschen Sprache, BBAW



Lexicography in the 21st century: Digitales Wörterbuch der deutschen Sprache, BBAW







Lexicography in the 21st century: - Dictionary Writing System

