

Internet lexicography at the Leibniz-Institute for the German Language

Stefan Engelberg, Annette Klosa-Kückelhaus and Carolin Müller-Spitzer

1. Introduction

Over ten years ago, Ilan Kernerman has invited us to write an article about lexicography at the Department of Lexical Studies at the Leibniz-Institute for the German Language in Mannheim (IDS). We wrote about the tenets of Internet lexicography at our institute, about the dictionaries we made, about the lexicographic processes, and about the demands with respect to IT competences and staff recruitment (Engelberg, Klosa and Müller-Spitzer 2009). Back then, we emphasized that the “ability to handle and analyse mass data and the need for Internet-adequate lexicographic concepts is changing the profession”, that Internet lexicography is driven by “new possibilities of data integration and crosslinking” and that “the demands made on the competences that have to be gathered within lexicographic projects, reaching from the development of corpus analysis methods to text technology and web technology”, were constantly increasing (Engelberg, Klosa and Müller-Spitzer 2009: 16). Admittedly, it did not take a prophet to predict the relevance of these parameters for future Internet lexicography. However, it should be interesting to see how – in more detail – these parameters shaped the lexicographic praxis in our institute within the last ten years.

In 2009, we had just implemented the *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online Lexical Information System on German) as “a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the Institute for German Language” (Engelberg, Klosa and Müller-Spitzer 2009: 16). OWID has proven to be a successful concept and it still constitutes the backbone of lexicography at our institute. In 2007, it was released with four lexicographic resources: a general dictionary on contemporary German, a dictionary of neologisms, a dictionary of idioms, and a discourse dictionary. Since then, six more dictionaries have been integrated into OWID (cf. section 2).



Stefan Engelberg

is head of the department ‘Lexik’ at the Leibniz-Institute for the German Language (IDS) in Mannheim, professor for German linguistics at the University of Mannheim, and honorary professor at the University of Tübingen.

engelberg@ids-mannheim.de

On re-reading the article from 2009, we found that we have followed the path pointed out by OWID quite consistently. The focus is now explicitly on specific-domain dictionaries, that cover areas of the lexicon that have so far been neglected by lexicography and lexicology. Despite its focus on resources for special vocabulary areas, OWID reaches users in many different countries around the world (cf. Figure 1).

However, since 2009, we have also conceptualized and implemented new types of lexicological-lexicographic platforms, and OWID and its dictionaries have developed new kinds of access structures and modes of presentation (cf. section 2). This was partly necessary because it turned out that a single portal with a particular concept of dictionary integration cannot serve all purposes, either because dictionaries develop new forms of presentation that cannot easily be integrated into OWID (such as the dictionary of paronyms, cf. section 2, which is only linked to OWID via its lemma list) or because particular uses of lexical data, especially those by scientific communities, require very different kinds of data aggregation, access to data, and lexicographic collaboration. These demands led to the development of new lexicographic portals (cf. sections 3 and 4).



Annette Klosa-Kückelhaus heads the 'Lexicography and Language Documentation' program area in the Department of Lexicology at the Leibniz-Institute for the German Language (IDS) in Mannheim, and is chief editor of its online dictionary of neologisms.
klosa@ids-mannheim.de

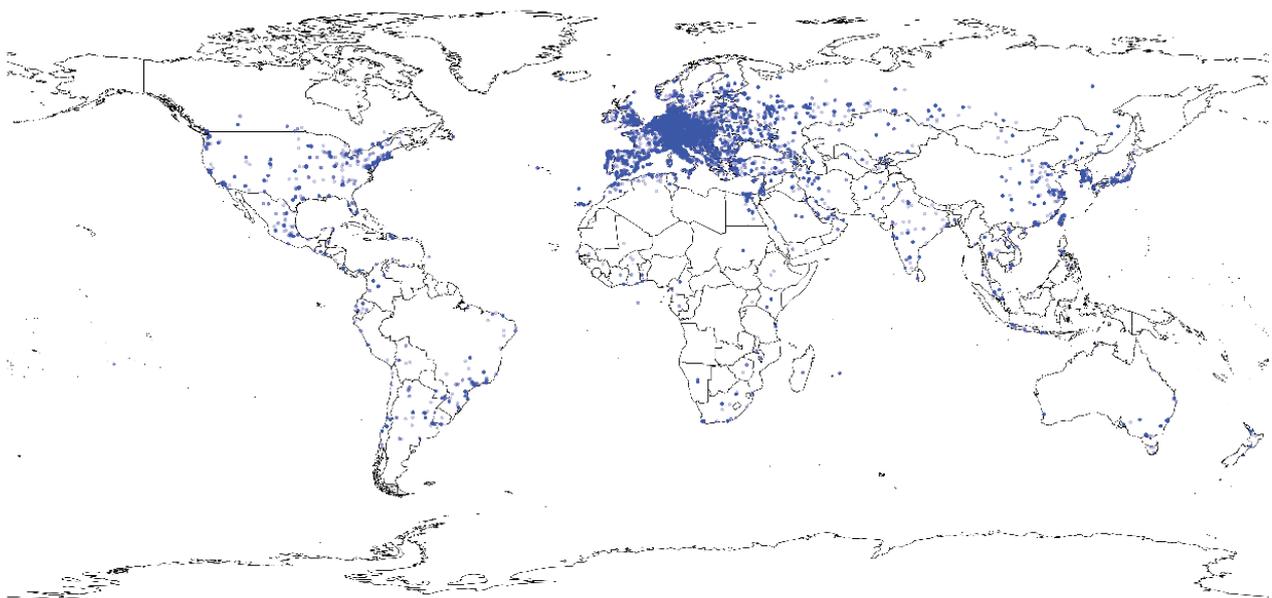


Figure 1. Geographically located accesses to OWID from the years 2018 and 2019

2. OWID

A. Dictionaries in OWID and OWID^{plus}

As of April 2020, eleven different lexicographic resources are offered in the OWID dictionary portal and in OWID^{plus}:

(1) *elexiko* – Online-Wörterbuch zur deutschen

Gegenwartssprache¹ (online since 2003, no further editing). *elexiko* is an online information system for the contemporary German language, which documents, explains and scientifically comments on the vocabulary based on current language data in individual modules. It comprises almost 1,900 systematically compiled, detailed entries for individual words, over 50 word group articles on meaning-relational groups (e.g. ‘Defizit – Mangel’ [deficit – deficiency], ‘Kindheit – Jugend – Alter’ [childhood – youth – old age]), thematic fields (e.g. ‘Beruf und Familie’ [job and family], ‘Jahreswechsel’ [turn of the year]) and word fields (e.g. ‘Getränke’ [drinks], ‘Speisen’ [food]) as well as over 250,000 entries that offer only automatically generated information on the headwords (orthographic information, corpus citations, frequency information).

(2) **Paronymwörterbuch**² (online since 2018, work in progress). This paronym dictionary documents easily confusable expressions in their current public usage. It contains expressions that, for example, have strong similarities in spelling, pronunciation and meaning, such as ‘farbig – farblich’ [colored], ‘kindlich – kindisch – kindhaft’ [childlike – childish], ‘universell – universal’ [universal]. The paronyms are presented together in new, contrasting dictionary articles. Their similarities and their differences can be seen at a glance, with the users deciding for themselves which sections or comparative views they want to get presented. The paronym dictionary clearly illustrates the focus of the dictionaries in OWID: on the one hand, to pick out certain vocabulary sections, for which lexicographic work has not been carried out to date, and on the other hand, to go new ways for the user interface. As such new visualizations are not easy to understand for users, a short video was created as a tutorial.³ This is also new for scholarly lexicography, but could certainly serve as a model.



Carolin Müller-Spitzer

is head of the program area ‘Lexic Empirical and Digital’ in the Department of Lexicology at the Leibniz-Institute for the German Language (IDS) in Mannheim, and professor of German Linguistics at the University of Mannheim. Her research focuses on user research, empirical gender linguistics and online lexicography. mueller-spitzer@ids-mannheim.de

1 <https://www.owid.de/docs/elex/start.jsp>

2 <https://www.owid.de/parowb/>

3 <https://www.owid.de/parowb/docs/hilfe.html>

(3) Sprichwörterbuch⁴ (online since 2012, no further editing).

This proverb and slogan dictionary illustrates that proverbs are still alive and an object of constant variation (e.g. ‘Harte Schale, weicher Kern’ [hard shell, soft core] with the variants ‘Rauhe Schale, weicher Kern’ [rough shell, soft core] and ‘Harte Schale, kaputter/genialer/leckerer Kern’ [hard shell, broken/ingenious/delicious core]), how speakers use them and how new ones are created, for example in advertising (e.g. ‘Nicht immer, aber immer öfter’ [not always, but more and more often]). It is also the first empirically validated documentation of currently used fixed phrases in the German language based on criteria of scholarly lexicography (created as part of the multilingual EU project SprichWort. An Internet Platform for Language Learning⁵, 2008-2010).

(4) Kommunikationsverben⁶ (online since 2013, finished).

This dictionary is the electronic version of a handbook of German speech act verbs, the *Handbuch deutscher Kommunikationverben* (edited by G. Harras, S. Erb, K. Proost and E. Winkler. Berlin/New York: de Gruyter 2004-2007).

It contains 241 entries on German verbs that refer to communicative actions, focusing on speech act verbs (e.g. ‘schimpfen’ [to scold], ‘resümieren’ [to summarize]). A specialty of this dictionary is that the semantic description takes place on two levels: the conceptual one (represented by the reference situation types) and the lexical one (represented in the word articles).

(5) Kleines Wörterbuch der Verlaufsformen im Deutschen⁷

(online since 2013, finished). This small dictionary of aspectual forms in German presents German verbs with regard to their occurrence in three aspectual forms, the *am*-progressive, the absentive and the *beim*-progressive. The aim is to provide researchers, students and teachers with the largest and most easily searchable collection of evidence on over 900 verbs that illustrate the use of these forms. This dictionary will be moved to the OWID^{plus} platform in the future (cf. section 3).

4 <https://www.owid.de/wb/sprw/start.html>

5 <http://www.sprichwort-plattform.org/sp/Projekt>

6 <https://www.owid.de/docs/komvb/start.jsp>

7 <https://www.owid.de/wb/progdb/start.html>

(6) Deutsches Fremdwörterbuch – Neubearbeitung⁸

(letters A-H of the revision online since 2016, work in progress). This German dictionary of foreign words (DFWB) describes and documents the vocabulary of today's learned everyday language, both in its current use and in its historical development from the respective date of borrowing to date. It currently comprises around 1,700 entries (with approximately 25,000 main and secondary headwords as well as approximately 130,000 corpus citations), e.g. 'Charakter' [character] with its numerous composites ('Charakterkopf' [striking head] etc., as well as 'Nationalcharakter' [national character], etc.) and derivatives ('charakterlos' [unprincipled], etc.). The entire first edition (letters I-K edited by Hans Schulz, published 1913; letters L-Q edited by Otto Basler, published 1942; letters R-Z, edited by IDS, published 1977-1983⁹) has been retro-digitized and is fully available online. The DFWB is the most comprehensive dictionary in OWID. The new edition contains the above-mentioned number of entries from the more than 5,500 pages of the previously published volumes 1 to 7 (1995-2010). The data released in April 2019 contains 3,354 entries from approximately 3,400 pages of volumes 1 to 6 of the first edition (1913-1983).¹⁰

(7) **Neologismenwörterbuch**¹¹ (online since 2004, work in progress). This dictionary presents over 2,000 new words, new phrase units and new meanings of established words that were incorporated into the general part of the vocabulary of the German standard language between 1991 and today. The new vocabulary from the three decades – 1990s (e.g. 'Handy' [cellphone]), the first (e.g. 'skypen' [to skype]) and the second decade of the 21st century (e.g. 'Influencer' [influencer]) – can be searched using various access routes (by subject groups, via the advanced search) (cf. section B).

(8) **Schulddiskurs 1945-55**¹² (online since 2008, finished). This dictionary is the online version of Heidrun Kämper's printed reference work *Opfer – Täter – Nichttäter. Ein Wörterbuch zum Schulddiskurs 1945-1955* [Victim – Culprit – Non-Culprit. A

8 <https://www.owid.de/wb/dfwb/start.html>

9 For more detailed information see <https://pub.ids-mannheim.de/laufend/fremdwort/aufgabe1.html?loop=2>

10 For information on the different parts of „Deutsches Fremdwörterbuch“ see: <https://www.owid.de/wb/dfwb/uebersicht.html>

11 <https://www.owid.de/docs/neo/start.jsp>

12 <https://www.owid.de/wb/disk45/einleitung.html>

IDS

LEIBNIZ-INSTITUT FÜR
DEUTSCHE SPRACHE

The Leibniz Institute for the German Language (Leibniz-Institut für Deutsche Sprache, IDS)

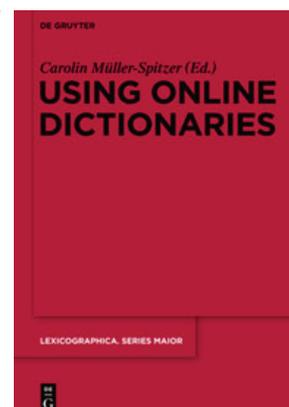
is the central non-university institution for the study and documentation of the contemporary usage and recent history of the German language. Together with more than 90 research and service institutions, it belongs to the Leibniz Association, one of the four major research organizations in Germany. At present, the IDS has 227 employees including 105 researchers. In the Department of Lexical Studies, more than 30 full and part time researchers and about 20 student researchers work in three so-called program areas on lexicography and language documentation (head: Annette Klosa-Kückelhaus), on syntagmatic aspects of the lexicon (head: Stefan Engelberg), and on empirical methods and the digital foundation of lexical studies (head: Carolin Müller-Spitzer). <https://ids-mannheim.de>

dictionary on the discourse of guilt 1945-1955] (Berlin & New York: de Gruyter 2007), which summarizes the lexical-semantic results of this investigation in the form of lexicographic entries. The entries describe those words according to lexicographic principles which represent the lexical framework of the discourse on guilt (having in the center the lexemes ‘Pflicht’ [duty] and ‘Schuld’ [guilt]).

(9) Protestdiskurs 1967/68¹³ (online since 2012, finished). This dictionary is the online reference work on Heidrun Kämper’s monograph *Aspekte des Demokratiediskurses der späten 1960er Jahre. Konstellationen – Kontexte – Konzepte* [Aspects of the Discourse on Democracy in the late 1960s. Constellations – Contexts – Concepts] (Berlin and Boston: de Gruyter 2012). Conceptually, it is a further development of the *Wörterbuch zum Schulddiskurs 1945-55*, containing around 90 articles on approximately 210 lemmas, each of which represent the discourse of left-wing students and intellectuals in a specific way.

(10) Schlüsselwörter der Wendezeit 1989/90¹⁴ (online since 2014, finished). This dictionary is the online version of the reference book by Dieter Herberg, Doris Steffens and Elke Tellenbach *Schlüsselwörter der Wendezeit. Wörter-Buch zum öffentlichen Sprachgebrauch 1989/90* [Keywords of the time of the German reunification. Wordbook of words in public language 1989/90] (Berlin and New York: de Gruyter 2007). It represents the public language around 1989-1990, is consistently corpus-based, and was prepared for the online version on the occasion of the 25th anniversary of the German reunification. It gives information on the more than 1,000 words and phrases that are relevant to that time, ordered according to 150 keywords (e.g. ‘Akte’ [file] and ‘Mauer’ [wall]) and thematic groups (e.g. ‘Die Deutschen vor und nach der Wiedervereinigung’ [The Germans before and after reunification], ‘Die politischen Veränderungen in der DDR und deren Vorboten’ [The political changes in the GDR and its heralds]).

(11) LeGeDe – Lexik des gesprochenen Deutsch¹⁵ (online since 2019, finished). This prototype of a dictionary of spoken German (conceptualized and implemented in a third-party funded, three year research project) offers a limited number of entries on nouns, adverbs, adjectives, verbs, etc. as well as multiword expressions which are characteristic for discourse in spoken language, e.g. the adjective



Research into dictionary use at the IDS

The IDS has a strong focus on research into dictionary use. Besides numerous journal papers, the team also published the first book-length work on empirical research on *Using Online Dictionaries*. A detailed review is: Lew, Robert. 2015. Research into the use of online dictionaries. *International Journal of Lexicography* 28.2, 232–253.

13 <https://www.owid.de/wb/disk68/start.html>

14 <https://www.owid.de/wb/swwz/start.html>

15 <https://www.owid.de/legede/>

‘gut’ [good], that can indicate the closure of a communicative task, or the expression ‘keine Ahnung’ [no idea], which may be used as a manifestation of uncertainty. For the first time in German lexicography, the lexicographic focus was not only placed exclusively on spoken language, but the dictionary entries are the first ones based entirely on corpora of spoken language, showing and explaining examples with (aligned) transcripts and audio files. Lexicographic information comprises details on the function of each word or expression in discourse, its function for signaling turns, its prosodic integration, etc., but also on how the use specifically for spoken language is connected to the respective word in written language (e.g. the expression ‘keine Ahnung’ is connected to the meaning ‘knowledge’ of the noun ‘Ahnung’ [idea, knowledge]). Our work on lexicographic information on spoken German is continued in a new project, which will cooperate with the new online research platform for spoken varieties of German outside the closed German language area in Central Europa (cf. section 4.B).

To summarize, in addition to retro-digitized online dictionaries (e.g., *Kommunikationsverben*), there are also dictionaries in OWID that were developed directly for online publication, e.g. *Sprichwörterbuch*. Besides completed dictionaries (e.g. *Schlüsselwörter der Wendezeit 1989/90*), there are some that are constantly worked on and are published dynamically (e.g. *Paronymwörterbuch*), and there are diachronic (e.g. *Deutsches Fremdwörterbuch*) as well as synchronic dictionaries

Recent journal articles on dictionary use

Wolfer, Sascha; Kosem, Iztok; Lew, Robert; Müller-Spitzer, Carolin and Ribeiro Silveira, Maria. 2018. Web-based exploration of results from a large European survey on dictionary use and culture: ESDexplorer. *Lexikos* 28, 440-447.

Müller-Spitzer, Carolin; Nied Curcio, Martina; Domínguez Vázquez, María José; Silva Dias, Idalete Maria and Wolfer, Sascha. 2018. Correct hypotheses and careful reading are essential: Results of an observational study on learners using online language resources. *Lexikos* 28, 287-315.

Kosem, Iztok; Lew, Robert; Müller-Spitzer, Carolin; Ribeiro Silveira, Maria and Wolfer, Sascha. 2018. The image of the monolingual

dictionary across Europe. Results of the European survey of dictionary use and culture. *International Journal of Lexicography*, 32.1, 92-114.

Wolfer, Sascha; Bartz, Thomas; Weber, Tassja; Abel, Andrea; Meyer, Christian M.; Müller-Spitzer, Carolin and Storrer, Angelika. 2018. The effectiveness of lexicographic tools for optimising written L1 texts. *International Journal of Lexicography* 31.1, 1-28.

Wolfer, Sascha and Müller-Spitzer, Carolin. 2016. How many people constitute a crowd and what do they do? Quantitative analyses of revisions in the English and German Wiktionary editions 26/16. *Lexikos* 26, 347-371.

(e.g. *Neologismenwörterbuch*). A special feature of OWID are the two discourse dictionaries *Schulddiskurs 1945-55* and *Protestdiskurs 1967/68*, a type of dictionary that was developed at the IDS.

B. Access structures: OWID and *Neologismenwörterbuch*

The main function of OWID is to provide a common access structure in the form of search options across the individual dictionaries. This is the typical function of lexicographic portals (e.g. YourDictionary.com, Wörterbuch-Portal) (cf. Müller-Spitzer and Engelberg 2013).

In OWID, there is a clear distinction between the level of the portal and the level of an individual dictionary. The search box of the portal is always accessible on the top of the webpage, while for each of the dictionaries, specific access structures are offered, which are shown once a user selects a certain dictionary by clicking on the dictionary button. With this distinction, we address two different user needs: firstly, searching for one word in any dictionary or, secondly, searching within one specific dictionary only.

Some of the OWID dictionaries offer advanced search options. An outstanding feature of OWID is that we try to develop appropriate advanced searches for each dictionary and use very diverse technologies to do so. For the German dictionary of foreign words (DFWB), for example, a rather narrative dictionary, we have invested primarily in developing effective full-text search. Since the lexicographic information in this dictionary is not granularly structured, a good full-text search is the most interesting way to access it. The full-text search can be narrowed down according to different text levels (keywords, article text, examples). On the other hand, in the neologism dictionary the situation is quite different, as the entries are structured in a fine-grained way. Therefore, we have applied another search technology and developed a user interface that aims to provide exactly these fine-grained structures to the end user with very distinguished search options. The advanced search¹⁶ here enables users to find all keywords that have a common feature (e.g. all new lexemes that were borrowed from languages other than English in the 1990s, as shown in Figure 2).

In addition, users can select the page 'Wortartikel'¹⁷ [entries] as a starting point for further exploration of the dictionary content.

Different lists group all entries according to different criteria:

Type of headword (single word entries, multiword units, new elements

¹⁶ <https://www.owid.de/docs/neo/suche/index.jsp>

¹⁷ <https://www.owid.de/docs/neo/wortartikel.jsp>

Erweiterte Suche im Neologismenwörterbuch

Stichwort

Neologismtyp

Neulexem
 Neulexem (nur Wörter)
 Neulexem (nur Phraseologismen)
 Neubedeutung

Informationen sichtbar

Aufkommen und Herkunft

Aufkommen

90er Jahre
 Anfang der 90er Jahre
 Mitte der 90er Jahre
 Ende der 90er Jahre
 Nullerjahre
 Anfang der Nullerjahre
 Mitte der Nullerjahre
 Ende der Nullerjahre
 Zehnerjahre
 Anfang der Zehnerjahre
 Mitte der Zehnerjahre
 Ende der Zehnerjahre
 Genau datierbar
 Kurzzeitwort

Informationen sichtbar

Herkunft (Sprache)

aus Englisch
 aus anderer Fremdsprache

Informationen sichtbar

Neologismtyp

Aufkommen

Herkunft (Sprache)

< zurück 1-11 von 11 vor >

Ciabatta	Neulexem (W.) E-90er andere
Fengshui	Neulexem (W.) A-90er andere
Hygge	Neulexem (W.) M-90er andere
hyggelig	Neulexem (W.) M-90er andere
Karaoke	Neulexem (W.) A-90er andere
Latte macchiato	Neulexem (W.) E-90er andere
Manga	Neulexem (W.) M-90er andere
Qigong	Neulexem (W.) M-90er andere
Tamagotchi	Neulexem (W.) M-90er, Kurzzeitw., dabierb. andere
Taupe	Neulexem (W.) E-90er andere
umami	Neulexem (W.) E-90er andere

Figure 2. Advanced search in *Neologismenwörterbuch* showing new lexemes from the 1990s borrowed from languages other than English

of word formation plus neologisms which are not headword but are contained in entries, e.g. lesser-used synonyms, derivations and compounds with the lemma).

Decades (neologisms from 1991-2000, 2001-2010, and 2011-2020).

Different types of entries such as the latest full entries (published always at the end of a calendar year), groups of short entries (published monthly, each related to a single topic, e.g. Europe, education), a list of neologisms that might enter the dictionary at some point and which are still monitored, or the most recent list of neologisms relating to the coronavirus pandemic (as shown in Figure 3)

Subject groups, for example sports, media, health and wellbeing, economics.

OWID

G

- Gabenzaun
- Geistermeister
- Geistersitzung
- gelockdownt
- Generation Corona
- Gesichtsschild
- Gesichtsschirm

H

- Handytracking
- Heimquarantäne
- Herdenimmunität

I

- Immunitätsausweis
- Immunitätsnachweis
- Immunitätspass

K

- Kontaktnachverfolgung
- Kontaktperson
- Kreativsemester
- Krisenkanzlerin

L

- Lockdown
- lokaler Lockdown

Home-Work-out

zu Hause durchgeführtes Fitnesstraining

Pamela hat es gestern vor lauter PR nicht ins Fitnessstudio geschafft, deshalb musste sie nachts um eins noch ein 45-minütiges **"Homeworkout"** einschieben. (Die Zeit, 13.07.2017)

Wie am laufenden Band produzieren Fitness-Influencer momentan neuen Content, um ihren Fans frische Inspiration für **Home-Workouts** zu liefern. (www.fitforfun.de; datiert vom 01.04.2020)

Erfasst: April 2020

Figure 3.

Extract from the list of monitored neologisms relating to the coronavirus pandemic

The interactive listing of entries according to their subject group¹⁸ is combined with the presentation of all relevant entries according to decades, thus enabling users to gain information on the social, cultural, technical, economical, etc., developments over the last 30 years. In addition, the *Neologismenwörterbuch* offers various possibilities for accessing the dictionary content on its homepage¹⁹ (e.g. direct links to lists of latest additions, the advanced search and entries in subject groups plus a link to suggest a new word), where also a very short introduction into the content is given.

Over all, users find many different options to explore the *Neologismenwörterbuch*. Once an entry has been selected and is presented on screen, exploration inside and outside of the dictionary

¹⁸ <https://www.owid.de/docs/neo/gruppen.jsp>

¹⁹ <https://www.owid.de/docs/neo/start.jsp>

may continue by following one or the other of numerous (dictionary internal and external) hyperlinks contained in the entries.

C. Information types: Examples from different dictionaries in OWID

All dictionaries in OWID analyse and describe the entries on the basis of extensive empirical, mostly corpus-derived, linguistic data. These dictionaries are products of scholarly lexicography and are the result of lexicological-lexicographic and metalexicographic research. Although most of them focus on specific parts of the vocabulary and not the general language, through their connection in OWID they offer fascinating insights into the German vocabulary, as shown in the following examples.

A search for 'smart' in OWID yields several results (cf. Figure 4) from three different dictionaries, namely *Deutsches Fremdwörterbuch*, *lexiko* and *Neologismenwörterbuch*. While *Deutsches Fremdwörterbuch* focuses on etymology and historical sense development of the adjective and noun derivations 'Smartness' and 'Smartheit' [being smart] as sub-lemmas, in *lexiko* three senses are disambiguated (clever, chic, technically highly developed) and explained from a synchronic view, focusing especially on judgmental usages specific for the first two senses. The third sense is the one found in compounds like 'Smart-TV' or 'Smartboard', neologisms of this millennium which are recorded in *Neologismenwörterbuch*. Here, users learn, among other things, that both loanwords stem from English and are productive in the formation of new

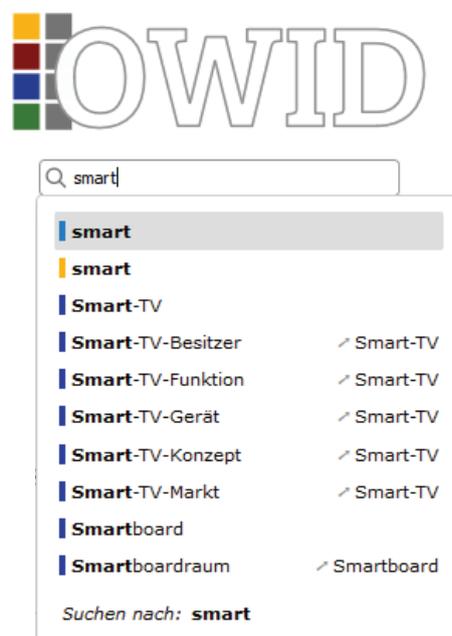


Figure 4.
Search results for
'smart' in OWID

The screenshot shows the Paronyme online dictionary interface. At the top, it says 'Paronyme Dynamisch im Kontrast'. Below this, there are two tabs: 'Vergleich' (Comparison) and 'Zusammenfassung' (Summary). The 'Vergleich' tab is active, showing a comparison between 'Evaluation' and 'Evaluierung'. The interface is divided into two columns, each representing one of the terms. Each column contains a definition, usage information, and several questions with answers. At the bottom of each column, there are expandable sections for 'Verwendungsbeispiele' (Usage examples) and 'Sinnverwandte Wörter' (Synonyms).

Evaluation, Bewertung	Evaluierung, Bewertung
bezeichnet eine (oft sachbezogene) Beurteilung oder Leistungsbewertung eines Prozesses, eines Sachverhaltes, einer Institution oder einer Person(engruppe)	bezeichnet eine (oft sachbezogene) Beurteilung oder Leistungsbewertung eines Prozesses, eines Sachverhaltes, einer Institution oder einer Person(engruppe)
oft in WISSENSCHAFT, BILDUNG	oft in WISSENSCHAFT, BILDUNG
Was oder wer erfährt eine Evaluation? Lehre, Studium, Forschung, Praxis, Lehrveranstaltung, Hochschule, Schule, Professoren, Standort, Lehrqualität	Was oder wer erfährt eine Evaluierung? Gesetz, Lehre, Forschung, Programme, Projekte, Maßnahmen, Institute, Arbeit, Lehrer
Was gibt es in Bezug auf eine Evaluation? Ergebnisse, Methoden, Konsequenzen	Was gibt es in Bezug auf eine Evaluierung? Ergebnisse, Kriterien
Was macht man bzw. was geschieht in Bezug auf eine Evaluation? (sich) unterziehen, durchführen, erfolgen	Was macht man bzw. was geschieht in Bezug auf eine Evaluierung? (sich) unterziehen, vornehmen, durchführen, beauftragen
Wie ist eine Evaluation? extern, intern, wissenschaftlich, regelmäßig, systematisch, umfassend, seriös	Wie ist eine Evaluierung? regelmäßig, unabhängig, wissenschaftlich, extern, intern, umfassend, systematisch, international
Verwendungsbeispiele	Verwendungsbeispiele
Sinnverwandte Wörter	Sinnverwandte Wörter

compounds such as ‘Smartboardraum’ [room with a smart board] or ‘Smart-TV-Konzept’ [smart TV concept].

A search for the string ‘evalu’ yields the following results (among others): the entry ‘Evaluation – Evaluierung’ [evaluation] from *Paronymwörterbuch* and the verb ‘evaluieren’ (to evaluate) in *Kleines Wörterbuch der Verlaufsformen im Deutschen*. Here, users find the information that ‘evaluieren’ is used in the progressive form with ‘am’ (as in ‘Die Forscher sind noch am evaluieren’ [The researchers are still evaluating]), which is illustrated in many corpus citations. In *Paronymwörterbuch*, the two nouns derived from this verb are compared with regard to similarities or differences in their meaning and use. In a highly innovative way of presenting the data distributed on different tiles in a horizontal order, users are able to understand that Evaluation and Evaluierung are used synonymously in most contexts, especially in scholarly language (cf. Figure 5).

When looking for the string ‘leben’, OWID yields (among others) the entry ‘Leben’ [life] in *Schulddiskurs 1945-55* and a number of proverbs from *Sprichwörterbuch*: ‘Leben und leben lassen’ [live and let live], ‘Ordnung ist das halbe Leben’ [a tidy house, a tidy mind], ‘Totgesagte leben länger’ [there’s life in the old dog yet], and ‘Wer zu spät kommt, den bestraft das Leben’ [life punishes the latecomer]. For each proverb, not only are the meaning and usage explained and many corpus examples given, but also information on variability, for example ‘X ist das halbe Leben’ [X is half the life] and ‘X und X lassen’ [live and let X live], are provided along with many examples

Figure 5.

The entry ‘Evaluation – Evaluierung’ from *Paronymwörterbuch* presenting information on both words simultaneously, enabling users to compare them

interest for publication now extends far beyond digital dictionaries. From a data perspective, digital dictionaries are resources prepared for end users. As a general rule, they are presented in such a way that they can be used without prior knowledge. Therefore, resources that are more likely to appeal to a specialist audience fit less into a general dictionary portal. In addition, it is essential for a dictionary portal such as OWID that all the dictionaries included have at least a few central similarities. Only in this way, the portal is able to offer a uniform user concept and common access structures. With OWID, these lowest common denominators are the access by words or by word units and the restriction to the German language as the dictionary matter.

In the course of our research work and the contact with external colleagues, however, it became increasingly clear that internal and external research projects often produce data sets that are not prepared for end users, but which are too valuable for the professional community not to be published. In OWID^{plus}, we provide space for a wide variety of resources, also multilingual ones. The individual resources are modularly implemented as independent interactive applications. Whether it might be useful in the long run to create a common index for all resources in OWID^{plus} is still open at the moment. Currently we are working on a common faceted search option of OWID and OWID^{plus}, because with the growing number of resources, the existing interface of OWID^{plus} becomes too heterogenous.

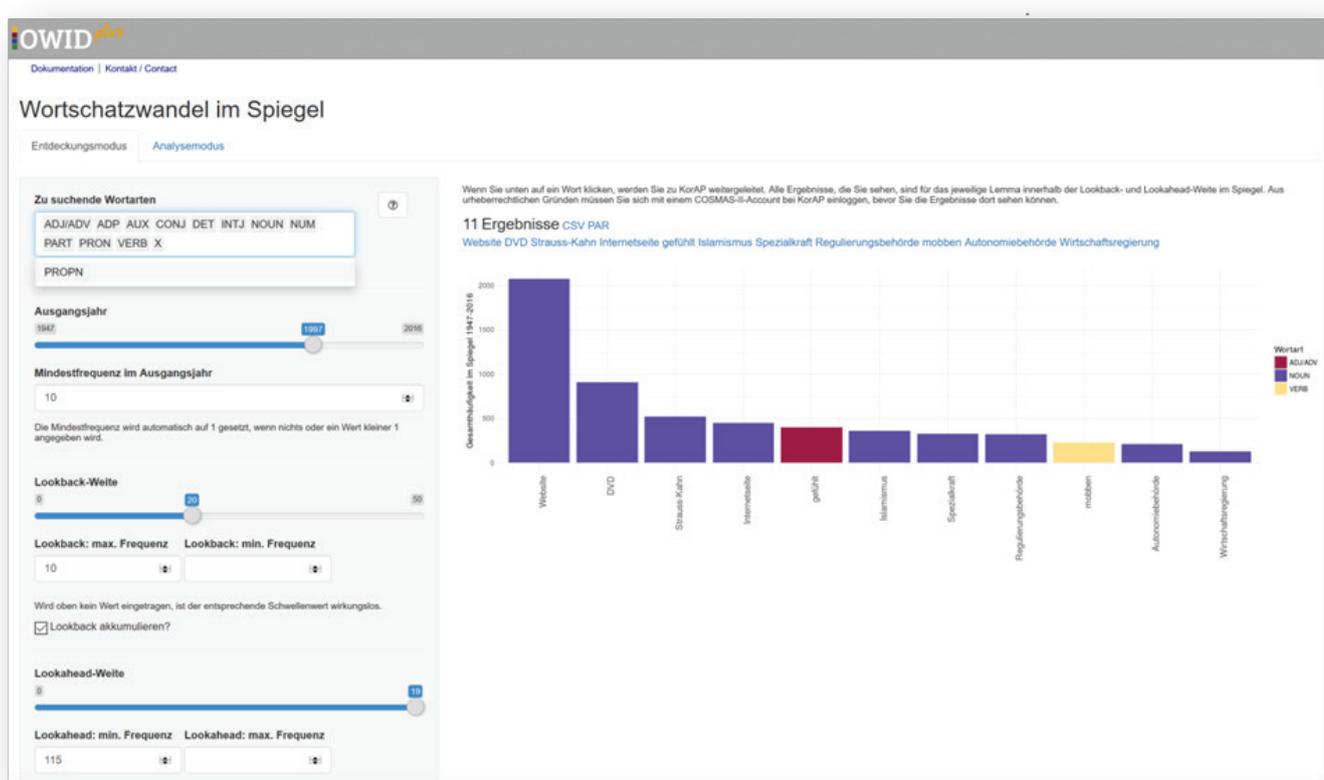
At present, 15 different resources can be found in OWID^{plus}, for example (i) the *Lexical Explorer*²⁰, with which quantitative corpus data on spoken German can be researched on the basis of frequency tables regarding the distribution over word forms, co-occurrences and metadata, (ii) the *ZAS Database of Clause-Embedding Predicates*²¹ of the Leibniz-Centre for General Linguistics (Leibniz-Zentrum Allgemeine Sprachwissenschaft) in Berlin, which contains clausal complementation patterns of lexical predicates in several languages, including multiple historical stages of German, (iii) a resource for log file statistics of six Wiktionary language editions²², and (iv) various visualizations of lexical change, for example Lexical change in *Der Spiegel*²³. With the last resource we would like to encourage users to move from passive consumption of linguistic knowledge to active exploration of a limited textual basis (all texts of the news magazine *Der Spiegel* from 1947 to 2016) and a limited phenomenon (here,

20 <https://www.owid.de/lexex/>

21 <https://www.owid.de/plus/zasembed/>

22 <https://www.owid.de/plus/wikivi2015/index.html>

23 <https://www.owid.de/plus/wvspiegel2018/>



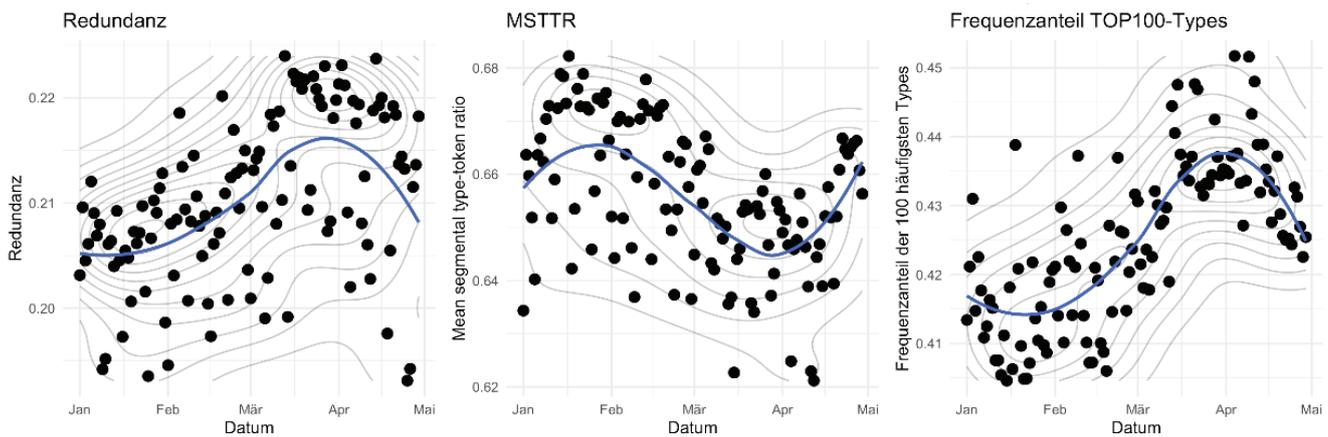
lexical innovations and archaisms). With a few settings, lists and frequency diagrams of emerging or dying word forms can be created. For example, it is possible to see that the word forms 'website', 'DVD', 'internet page' and 'mobbing' appeared in 1997 and continue to be frequently used since then (cf. Figure 7). The tool can be usefully applied in teaching to demonstrate the quantitative research of lexical innovation, or generally used to pursue one's own linguistic urge to explore more information.

The architecture of OWID^{plus} enables us to publish new resources very quickly. We have recently used this opportunity to upload two brand new resources on the occasion of the corona pandemic: cOWID^{plus} Analysis²⁴ and cOWID^{plus} Viewer²⁵. The motivation for these resources was that all around the globe, the coronavirus pandemic has been affecting almost every part of public life. Consequently, the pandemic is the subject of discussion not only in private face-to-face conversation (whenever this kind of talk has been possible), but also in the news. With lots of daily life activities like sports and cultural events coming to a stop, corresponding newspaper desks might very well run out of events to report on and shift their focus to pandemic-related topics. This gives rise to the assumption that the vocabulary used in articles,

Figure 7.
Results from
'Wortschatzwandel im
Spiegel' from 1997

24 <https://www.owid.de/plus/cowidplus2020/>

25 <https://www.owid.de/plus/cowidplusviewer2020/>



not only in printed but also in online media, is changing. To be precise, we would assume a concentration of (content word) vocabulary on concepts that are associated with the coronavirus pandemic. This does not necessarily mean that fewer word types are used. Rather, it suggests a shift in frequency distribution over types in a way that such distributions are more heavily skewed towards (temporarily) important types. Such a change is detectable by quantitative measures. In cOWID^{plus} Analysis we can indeed see a measurable narrowing of topics (and hence of the vocabulary) during the corona pandemic in selected (online) news outlets in German language, especially in mid-March 2020 (cf. Figure 8).

The cOWID^{plus} Viewer enables users to explore the data of cOWID^{plus} Analysis. An easy-to-use interface enables visualisation of the frequency curves of word forms. Data and illustrations are also available for download and are updated weekly. It is especially the topicality of the data that allows us to quickly discover and follow interesting lexical trends of the corona crisis. Figure 9 gives an example: at the beginning of the corona crisis, panic shopping (in German 'Hamsterkäufe', literally "hamster purchases") were a big issue. All over the world different things were "hoarded", in Germany it was mainly toilet paper ('Klopapier') and flour ('Mehl'). This is apparent in the news feeds as well: mid-March was the climax of the "hamster topic".

All resources in OWID^{plus} are primarily aimed at the linguistic community. Accordingly, we are always interested in receiving feedback on the existing offers as well as in offers to provide further data.

Figure 8.

Development of various lexical diversity measures from January to May 2020 (Redundanz = redundancy, MSTTR = Mean segmental type-token ratio, Frequenzanteil Top100-Typen = frequency proportion of Top100 Types)

4. Two Internet portals for the lexicography of language contact

A. The Loanword Portal for German

The Loanword Portal for German ('Lehnwortportal Deutsch') documents words that have been borrowed from German into other languages. It is mostly not based on original etymological research but integrates already-existing loanword dictionaries or ongoing dictionary projects. Methodologically, it differs from OWID in



Published within the Loanword Portal for German are the following lexical resources:

Belarusian, Russian, and Ukrainian. Hentschel, Gerd et al. *Dictionary of German loanwords in the East Slavic languages with a parallel in Written and Standard Polish*. [compiled for the portal]

Czech and Slovak. Newerkla, Stefan Michael (2nd ed., 2011): *Sprachkontakte Deutsch – Tschechisch – Slowakisch. Wörterbuch der deutschen Lehnwörter im Tschechischen und Slowakischen*. Frankfurt am Main: Peter Lang. [digitized from a PDF text file]

Dutch. Veen, Pieter Arie Ferdinand van and van der Sijs, Nicoline (1997): *Etymologisch woordenboek: de herkomst van onze woorden*. Utrecht: Van Dale Lexicografie. [retro-digitalized, excerpts]

English. Pfeffer, J. Alan and Cannon, Garland (1994): *German Loanwords in English. A Historical Dictionary*. Cambridge, New York, Melbourne: Cambridge University Press. [retro-digitalized]

French. (1) *Le Trésor de la Langue Française Informatisé*, online:

<http://atilf.atilf.fr> [relevant excerpts converted from another digital format]; (2) Sarcher, Walburga (2001): *Das deutsche Lehnwort im Französischen als Zeugnis für den Wissenstransfer im 20. Jahrhundert*. Hamburg: Verlag Dr. Kovač [retro-digitalized]

Hebrew. Adiv, Uriel and Mendel, Jakob (2015ff.), *Deutsche Lehnwörter im Hebräischen*. [compiled for the portal]

Hungarian. Benkő, Loránd and Büky, Béla (1993-1997): *Etymologisches Wörterbuch des Ungarischen*. Budapest: Akadémiai Kiadó. 3 vols. [retro-digitalized, excerpts]

Polish. de Vincenz, Andrzej and Hentschel, Gerd (2010), *Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache*. [converted from another digital format]

Portuguese. Schmidt-Radefeldt, Jürgen and Schurig, Dorothea (1997): *Dicionário dos Anglicismos e Germanismos na Língua*

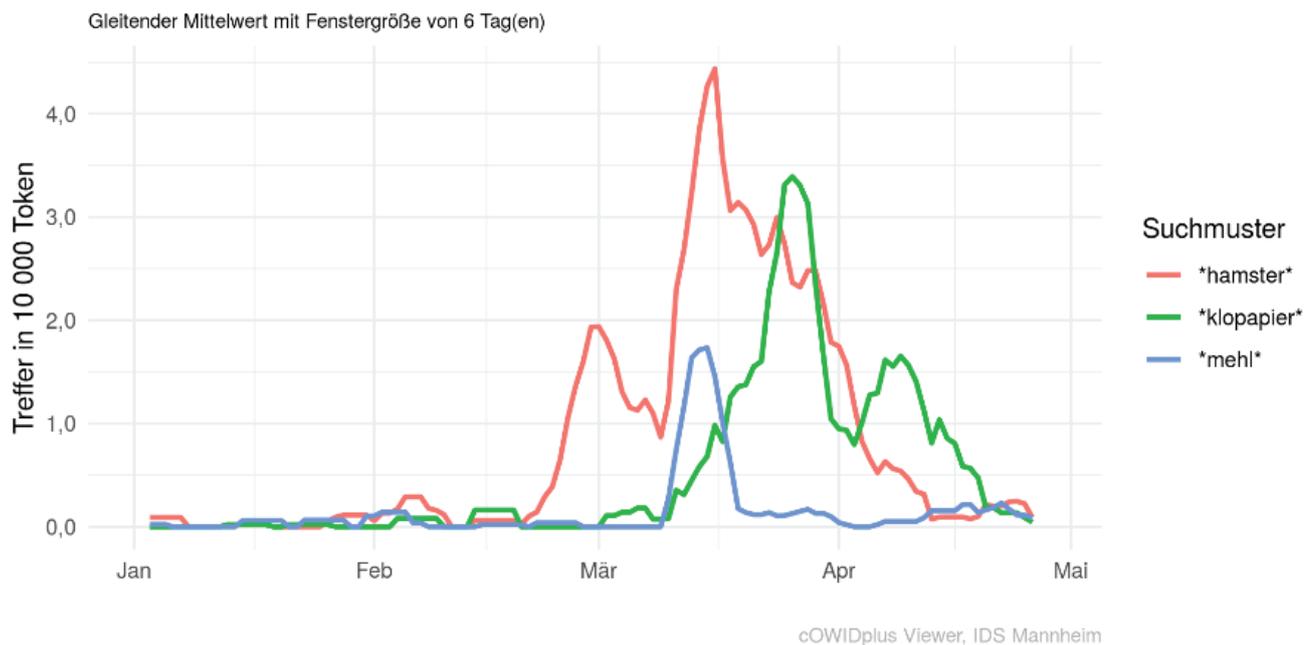
Portuguesa. Frankfurt/M.: Teo Ferrer de Mesquita. [retro-digitalized, entries on German loanwords]

Slovene. Striedter-Temps, Hildegard (1963), *Deutsche Lehnwörter im Slovenischen*. [retro-digitalized]

Teschen dialect of Polish. Menzel, Thomas and Hentschel, Gerd (2005), *Wörterbuch der deutschen Lehnwörter im Teschener Dialekt des Polnischen*. [converted from another digital format]

Tok Pisin. Engelberg, Stefan; Möhrs, Christine and Stolberg, Doris (2017ff.), *Wortschatz deutschen Ursprungs im Tok Pisin*. [compiled for the portal]
The following resources are going to be added in 2021 (some licenses pending):

Uzbek. Jumanioyov, Atabay and Vohidova, Nofiza (in preparation): *Dictionary of German loanwords in Uzbek* [retro-digitalized, edited and translated version of an Uzbek monograph]



two respects, as it does not primarily include dictionaries that were compiled in-house and (therefore) does not have a thorough corpus-lexicographic basis. The integrated dictionaries, which usually have lemmatized words in the target language of the borrowing process, can be consulted as individual works; however, in the database of the portal all information is additionally represented as a complex, cross-dictionary network of loanwords and words of origin. Thus, the portal can be used as a “reverse loanword dictionary”: in the dictionary of words of origin, generated from a German meta-lemma list, it is possible to search for related loanwords in other languages.

The loanword portal in its current version contains dictionaries of German loanwords in Hebrew, Polish, Slovene, the Teschen dialect of Polish, and Tok Pisin. A new, entirely revised, version, to be released in mid-2021, will also include dictionaries of German loans in Czech, Dutch, the East Slavic languages Belarusian, Russian, and Ukrainian (restricted to cases with a corresponding loan in Polish), English, French, Hungarian, Portuguese, Slovak, and Uzbek.

The portal stands out due to its complex search options that enable expert searches of lexical items in the database network based on arbitrary combinations of search features (target language, time of borrowing, dialect of source item, semantic components, grammatical category, relationship to other words in the network, etc.).

B. A language variety platform for German languages enclaves

A new project at our institute is in its early stages of development. Its aim is the implementation of a lexicological-lexicographic

Figure 9.
hamster, *klopapier*
and *mehl* in cOWID^{plus}
Viewer

The screenshot shows the website 'Lehnwortportal Deutsch' with a search bar containing 'Adresse'. The page displays the following content:

- Header:** IDS LEHNWORTPORTAL FÜR DEUTSCHE SPRACHE, Lehnwortportal Deutsch, Startseite | Erweiterte Suche | Dokumentation | Bibliografie, Suche nach deutschen Herkunftswörtern, Suche
- Navigation:** Wörterbuch der deutschen Herkunftswörter des Portals – umgekehrtes Lehnwörterbuch (Institut für Deutsche Sprache), A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
- Search Results:**
 - Adresse** (with a dropdown menu listing: abstechen, Abstrich, absurd, Abt, abuh, Abzug, ach, Achse, Achsel, Achselband, Achtel, ächten, Achtung, Ader)
 - Lehnwörterbucheinträge zu diesem Herkunftswort**
 - Slovenisch (Striedter-Temps 1963)**
 - **adresa**
 - Herkunftswort Adresse
 - Tok Pisin (Engelberg/Möhrs/Stolberg 2017ff.)**
 - **adres**
 - Herkunftswort Adresse
- Sidebars:**
 - Herkunftswörterbuch:** Deutsch (Erweiterte Suche, Graphensuche, Dokumentation)
 - Lehnwörterbücher:** Polnisch, Teschener Polnisch, Slovenisch, Hebräisch, Tok Pisin

Figure 10. Search for ‘Adresse’ [address] in the German meta-lemma list of the *Lehnwortportal Deutsch*

online research platform for spoken varieties of German outside the closed German language area in Central Europe. Varieties of this sort are spoken on all five continents and include (i) varieties that show a large degree of dialect levelling and therefore often exhibit a considerable overlap with the lexicon of standard German, for example Barossa German in Australia, Namibian German, or Paraguayan Laguna German, (ii) varieties that retain strong features of German dialects and therefore show a considerable difference from the standard German lexicon, for example Puhoi Bohemian German in New Zealand, or Siberian Mennonite East Low German, and (iii) German-based pidgin and creole languages such as Kiche Duits in Namibia or Unserdeutsch from New Guinea.

We plan to develop a modular online platform for variational and contact lexicology that dissolves the distinction between dictionary and corpus and accommodates the needs and interests of lexicographers, linguists and non-expert speakers of the included varieties. The platform builds on the increasing number of digitally searchable corpora of the spoken language of these varieties that adhere to current standards of transcription and annotation. The envisaged system will allow lexicologists to supplement standard metadata and token annotations of such corpora by custom annotation layers for specific research questions and additional variational parameters. Researchers and end users will be provided with advanced tools and presentation options for lexicographic documentation, quantitative analyses and information extraction beyond standard corpus query technology. The integrated toolset will make it possible

Angaben zum deutschen Herkunftswort

Herkunftswort +

Bedeutungserläuterung enthält +

sprachliche/raumzeitliche Einordnung +

grammatisches Merkmal +

Angaben zum Lehnwort

Lehnwort +

Lehnwort (2) +

Bedeutungserläuterung enthält +

Sprache +

grammatisches Merkmal +

Suchen

Suchergebnisse

- Pariser >>>
- Pucher >>>
- Springer >>>
- Weidwatschker >>>
- Wider

Zur Suchanfrage passende
Paare Herkunftswort → Lehnwort

bidra

Wider → bider
Wider → bidra

Figure 11.

Complex search in the *Lehnwortportal Deutsch*: nouns ending in ‘-er’ (an instrumental suffix) borrowed from the Styrian dialect of German into Slovene

to systematically address questions of intra-variety linguistic variation in domains such as register-specific collocations, phenomena at the interface between lexicon and grammar, lexical specifics of spoken language, or lexical contact phenomena. The first test case will probably be Namibian German, and as soon as the number of varieties integrated into the platform increases, the portal will also allow for comparative studies, both between these post-migration varieties and in relation to spoken varieties of German as a majority language in Germany. The details of the development of the platform will depend on both funding and the interest of linguistic colleagues in cooperating with us.

5. Competences and workflow

A: Consequences of changing lexicographic processes

When we began to develop OWID, the overall Internet lexicographic workflow was fairly straightforward: modelling XML structures, writing entries in an XML editor, storing entries in an Oracle database, creating stylesheets for HTML-based data presentation, and programming access structures. With more diversity in Internet dictionaries and portals and an increasing number of external cooperation partners, the workflow becomes more complex. Standardizing processes and structures is a strong demand on the one hand, while on the other hand, we want to develop new forms of lexicographic data presentation, new lexicographic methods, and new steps within lexicographic processes adapted to these methods. In the following, we sketch some developments in our department which determine work processes and staff recruitment.

Corpus analysis. The Leibniz-Institute of the German Language has a department for research infrastructure and corpus linguistics. The corpora and the corpus analysis systems developed in this department are used fairly intensively by our lexicographers. However, with our specific lexicological research questions and lexicographic projects, we often need certain corpora and methods of analysis that cannot all be provided by our in-house corpus linguistics resources. Therefore, we have to rely more and more on expertise on corpus linguistics within our own department and a larger part of the staff budget has to be shifted to meet this demand.

Quantitative methods. Quantitative analysis of corpora has become more important for lexicological and lexicographic purposes, which becomes particularly obvious in some of the resources in OWID^{plus}. Ten years ago, quantitative linguistics hardly played a role in the department; meanwhile, we have hired several persons who work mostly or partly with quantitative methods.

Collaboration. Lexicography is a labor-intensive process. Sooner or later it turns out that you cannot achieve everything on your own. During the last decade we have established a number of lexicographic co-operations, in particular with universities in Germany and several European countries. This pertains partly to OWID^{plus} and moreover to the loanword portal and the emerging variation platform, where the lexicographic content is almost exclusively produced with our external partners. These co-operations, in turn, require the development of specific lexicological and lexicographic tools.

New platforms. Our new Internet platforms require technical solutions

that go beyond what we had implemented for OWID. Besides the relational database (Oracle) used for OWID, we now employ also a full-text search engine (Elasticsearch) in OWID and NoSQL storage and retrieval (Neo4J for graphs, BaseX for XML) in the loanword portal and projects therein. The collaboration with lexicographers outside the institute led to developments affecting the lexicographic process. While many of our in-house lexicographers compile dictionary entries using the commercial XML editor Oxygen, the special needs of our collaborative projects required the programming of dedicated online editorial tools based on open-source XML and rich text editing components. In general, web development has become much more frontend-centered in recent years, increasingly replacing server-side generation of static browser content by interactive presentations and visualisations driven by modern reactive technology that runs in the browser.

Sustainability. The more lexicographic resources we produce, and the more diverse they are, the more work we have to allocate to measures of maintenance and sustainability. This includes trying to standardize procedures and formats where it is possible to develop software of a more generic sort. In spite of these efforts, the number of lexicographic products and services that have to be accompanied through the change of technical surroundings is growing, which creates a rising demand for computational lexicographic support.

B. Staff recruitment

The last ten years at our department have seen lexicological studies and lexicographic practice growing closer together. Many of our employees now conceive and compile dictionaries on the one hand and publish on empirical lexicological research on the other. In our paper from 2009, we estimated that the work of about 10-12 full-time equivalents of staff distributed over 18 members was allocated to Internet lexicography. While the overall time devoted to lexicography has not changed much (ca. 12 FTE), almost everybody in the Department of Lexical Studies is now involved part-time in one way or the other in lexicographic activities.

More than we did ten years ago, we advertise vacancies that combine expertise from various fields: in linguistics and lexicography, in corpus linguistics and computational lexicography, in lexicography and text technology, etc. We still do not hire IT specialists without a thorough background in the humanities, since the close collaboration among these domains in our department necessitates a common background in the conception and implementation of linguistic and lexicographic projects. A particular challenge arises from the

fact that IT expertise is in high demand in the private sector. As a publicly-funded institute, being restricted to standard wages, we have to offer a particularly interesting and inspiring work environment to attract highly qualified and motivated applicants with competences in these domains.

6. Outlook

We have shown how lexicography has changed at the Institute for the German Language within the last ten years. We have further developed our main dictionary portal OWID into a dictionary platform for specific-domain dictionaries in areas that have not been covered adequately by lexicological research and lexicography. Beyond that, new platforms have been designed and implemented: OWID^{plus} as an experimental platform for different kinds of lexical resources and tools, the Loanword Portal for German as a dictionary network for the publication of resources on German loanwords in other languages, and – forthcoming – a lexicological and lexicographic platform for spoken varieties of German outside the closed German language area in Central Europe. With these developments in progress, there is still an increasing demand for expertise in corpus analysis methods, text technology and web design but also for competences in the creation of new lexicographic formats and the lexicographic integration of current lexicological research.

The lexicographic resources at our department now reflect two different basic types of lexicography: communication-oriented dictionaries are directed more towards lay people and serve to solve communication problems or support language acquisition, whereas knowledge-oriented dictionaries mainly address a linguistically informed audience and document linguistic knowledge about the lexical system of a language (cf. for similar distinctions Tarp 2008; Engelberg 2014; Engelberg, Klosa-Kückelhaus and Müller-Spitzer 2019). However, there are of course over-arching principles of scholarly lexicography that both types of dictionaries have to adhere to: a thorough empirical basis of all lexicographic information, a concept of the nature and breadth of lexical knowledge, and a user-orientation with respect to access structures and information presentation.

References

- Engelberg, S. 2014.** Gegenwart und Zukunft der Abteilung Lexik am IDS: Plädoyer für eine Lexikographie der Sprachdynamik. In *Institut für Deutsche Sprache (Hg.): Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*. Mannheim: Institut für Deutsche Sprache, 243-253.
- Engelberg, S., Klosa, A., and Müller-Spitzer, M. 2009.** Challenges to Internet lexicography: The Internet dictionary portal at the Institute for German Language. *Kernerman Dictionary News*, 17: 16-25.
https://kdictionaries.com/kdn/kdn17_2009.pdf
- Engelberg, S., Klosa-Kückelhaus, A. and Müller-Spitzer, C. 2019.** Lexikographie zwischen Grimm und Google? *Sprachreport* 35.2, 30-34.
- Engelberg, S. and Müller-Spitzer, C. 2013.** Dictionary portals. In Gouws, R., Heid, U., Schweickard, W. and Wiegand, H.E. (eds.), *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin and Boston: de Gruyter, 1023-1035.
- Tarp, S. 2008.** *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Niemeyer.