

---

Petra Storjohann

## THE PUBLIC AS LINGUISTIC AUTHORITY: WHY USERS TURN TO INTERNET FORUMS TO DIFFERENTIATE BETWEEN WORDS

**Abstract** This paper addresses the question of why we face unsatisfactory German dictionary entries when looking up and comparing two similar lexical terms that are loan words, new words, (near)-synonyms, or confusables. It explains how users are aware of existing reference works but still search or post on language forums, often after consulting a dictionary and experiencing a range of dictionary-based problems. Firstly, these dictionary-based difficulties will be scrutinised in more detail with respect to content, function, presentation, and the language of definitions. Entries documenting loan words and commonly confused pairs from different lexical reference resources serve as examples to show the shortcomings. Secondly, I will explain why learning about your target group involves studying discussion forums. Forums are a valuable source for detailed user studies, enabling the examination of different communicative needs, concrete linguistic questions, speakers' intuitions, and people's reactions to posts and comments. Thirdly, with the help of two examples I will describe how the study of chats and forums had a major impact on the development of a recently compiled German dictionary of confusables. Finally, that same problem-solving approach is applied to the idea of a future dictionary of neologisms and their synonyms.

**Keywords** Internet forums; synonyms; confusables; sense discrimination; problem-solving approach

### 1. Introduction

In any language, there are specific lexical terms which can cause confusion and uncertainties among native speakers and language learners. It is often loanwords, neologisms, synonyms, or paronyms (confusables) which can trigger doubts about their appropriate contextual use and their exact semantic differences because they have foreign origins (loanwords), new and unknown meanings (neologisms), semantic similarities (synonyms<sup>1</sup>), or commonalities in their lexical forms (paronyms) and because they can designate similar concepts. Whenever two words exist in a close semantic relationship or even in lexical competition with one another, they pose linguistic difficulties. In the past twenty years, language forums have established themselves as linguistic authorities which the public uses to judge instances of lexical uncertainty. Typically, a user posts a question with or without elucidating the contextual circumstances in which a lexical choice between two words is necessary. Typical replies include suggestions, intuitive responses, or copied dictionary entries, and these are further commented on by different users or referred to again by the initial user.

Examining online forums, we see an astonishing number of questions relating to language situations where someone is seeking advice on how to distinguish between two or more lexical items belonging to one of the aforementioned categories. Often, users consult forums after looking up words in a dictionary and experiencing various dictionary-based problems (Murphy 2013). These often concern insufficient information, lack of encyclopaedic knowledge, missing entries, specific emphasis on dominant meaning, or ignorance of language change. Hence, dictionaries are not always the most effective resources to solve problems of

---

<sup>1</sup> For a discussion on the notion of synonymy used in lexicography cf. Murphy (2013).

language production. This even holds true for native speakers who are sufficiently competent to identify and reflect on the information given in entries (Chon 2008). Although German has a long-established lexicographic tradition of describing loan words, synonyms, and neologisms, there are only a few monolingual contrastive reference works, such as a paronym dictionary, which allow users to look up two lexical items simultaneously in order to compare their meanings and usage.<sup>2</sup> Evidently, there is a genuine need for contrastive dictionaries explaining semantic nuances, equivalent terms, and relatedness. Despite the fact that user studies have uncovered a number of insights into dictionary behaviour, skills, and consultation habits and that these studies have identified strategies of dictionary use in interactions with existing online dictionaries (e.g. cf. Müller-Spitzer 2014, cf. Lew 2015), little research has been carried out to investigate actual communicative needs and the linguistic queries associated with them, together with their corresponding answers. In fact, this is the only way to truly understand the potential target group for a linguistic resource, to identify their skill levels, and to develop innovative tools to ensure appropriate and reliable use of the resource in specific situations (cf. Storrer 2013).

In this paper, I will show why some well-known dictionaries fail to address common user queries. At the same time, I will show how we can overcome unsuccessful lexicographic habits by studying users' enquiries carefully. Finally, I will demonstrate how central conceptual ideas for an online dictionary of confusables ("Paronyme – Dynamisch im Kontrast") were derived from forums and effectively implemented during its planning phase and how they could be applied to the development of a future dictionary of German neologisms, synonyms, and loan words.

## 2. Dictionary-based problems and forums

Today, popular options among resources for language consultation include search engines, user-generated collaborative formats like Wiktionary, digitised and new online dictionaries produced by publishing houses (e.g. Duden), academic reference guides like DWDS, and NLP-based lexical tools (e.g. WortschatzLeipzig). Generally, users are accustomed to these but are not aware of the differences between them in terms of their underlying data, editorial processes, or their compilers' qualifications. Most online resources are characterised by typical dictionary-based problems, and users face a variety of challenges, e.g. the exhaustiveness and reliability of lexicographic details, the relationship between linguistic and extralinguistic information, the lack of (corpus) examples, how up-to-dated the data is, and the use of appropriate description style.<sup>3</sup> In the worst cases, dictionaries ultimately confuse users and cause vocabulary problems instead of solving them. Modes of presentation are rarely subject to criticism by dictionary users in chats, unless they prevent them from locating relevant pieces of information.

Users searching for synonyms, for example, do so for different reasons. Chon (2008) refers to these as "competence deficit word problems", which occur when a word or specific aspects of it are unknown. Searching for contextually appropriate lexical substitutes in dictionaries

<sup>2</sup> A new contrastive tool is WikiUnterschied.com, which compares wiktionary entries in a table format.

<sup>3</sup> We know from user studies of German online dictionaries by Müller-Spitzer (2014), contents and reliability are most crucial to users.

or thesauruses is also a typical problem in situations of text production when native speakers and language learners are searching for lexical alternatives (Rundell 1999). In the context of English language learners and with respect to synonyms, Chon (2008, p. 24) points out that “successful language production depends considerably on the ability to make appropriate lexical choices in dictionary entries [...]”. Looking up synonyms is also essential in a situation of language reception when users are not familiar with a specific item, usually a loan word, technical term, or a new word. A typical query in forums might adopt the following style: *What is the difference between Grippe/Erkältung/Influenza (flu/cold/influenza)?* Besides learning about collocational and syntactic norms, getting a deeper understanding of differences means internalising semantic and encyclopaedic variations.

All these situations, at least to some degree, also apply to searching for easily misused words. Paronyms are similar to one another in their lexical form and often, to some extent, in meaning. They share a morphological root and typically differ with respect to prefixes or suffixes. A large number of paronyms are in fact loan words, such as *anarchisch/anarchistisch* (*anarchic*), *fiktiv/fictional* (*fictitious/fictional*), and some of them denote identical concepts and exist in well-established synonym relationships (e.g. *patriarchalisch/patriarchal/patriarchisch*, (*patriarchal*)). Using loan words, in particular, can cause misunderstandings, as they are stylistically marked and exhibit a certain degree of education. There are also terms with indigenous roots such as *farbig/farblich* (*coloured, colourful, in/concerning colour*) or *lesbar/leserlich* (*readable/legible*) which can cause problems. Again, these competence-deficit word problems often relate to both insufficient semantic and extra-linguistic knowledge. Speakers have different or only vague and subjective intuitions and show a lack of knowledge as to the precise contextual circumstances in which the terms should be used. In forums, questions like *What does autoritativ (authoritative) mean and how does it differ from autoritär (authoritarian)?* or *Is there is difference between fremdsprachig/fremdsprachlich (in terms of a foreign language?)* are a source of debate and controversy. With new words (coinings or new loan words) uncertainties differ. The element of novelty uncovers deficits in specific knowledge about a phenomenon. *What do the terms Covid/Corona/SarsCov-2 mean exactly?* is a question arising from new and simultaneous information and lexical input about similar or related phenomena.

In what follows (2.1 and 2.2), I will pick out common failings and pitfalls typically encountered when searching for lexical pairs with an explicit need to identify of a precisely drawn spectrum of meaning. I will look at their treatment in popular German dictionaries with regard to three aspects: lexicographic information, degree of detail, and defining style.

## 2.1 Depth and presentation of lexicographic information

In German, the use of *formal*, *formell*, or *förmlich* poses difficulties in various contexts. These loan words, adopted in the late 15<sup>th</sup> century from Latin *formalis*, are also paronyms and used synonymously in contemporary German in some of their contexts. Looking them up in a German dictionary is a confusing experience. In an example taken from the Leo-language forum (Fig. 1), a dictionary-based problem is reported by a user.

Übersicht > Sprachlabor > formal - formell 5 Antworten ↓

Betrifft	<b>formal - formell</b>
Kommentar	Hallo, grüße gerade über die Bedeutungsunterschiede zwischen formal und formell. Gibt es welche? Es ist ein Unterschied, ob ich die formalen Voraussetzungen erfülle, um Mitglied zu werden (eine Führerscheinprüfung abzulegen etc.), oder die formellen? Dank und Gruss
Verfasser	mawa 13 Sep. 06, 19:43
Kommentar	Eine Daumenregel: Das Gegenstück zu 'formell' ist 'informell' Das Gegenstück zu 'formal' ist 'formlos' ... Vergleiche aber auch noch LEOs Antworten auf 'formal' und 'formell' ...
#1 Verfasser	Daddy 13 Sep. 06, 20:08
Kommentar	Ich hatte in Leo nachgeschaut und Wortschatz Uni Leipzig und war zu keinem abschliessenden Ergebnis gekommen. Bei <a href="http://wortschatz.uni-leipzig.de/">http://wortschatz.uni-leipzig.de/</a> gibt es folgende Beispielsätze: "Unbekannt ist indes, welche Institution formal für die Verwaltung des Projekts verantwortlich ist. (Quelle: Neues Deutschland 2003) Die USA haben begonnen, auch formal die letzten Voraussetzungen für einen Krieg gegen Irak zu schaffen. (Quelle: Neues Deutschland 2003)" und "Jener war Abwehrchef der formell noch gesamt-jugoslawischen Armee, bis Milosevic im Frühjahr 1992 das serbisch-montenegrinische Rumpf-Jugoslawien bildete. (Quelle: Süddeutsche Online) Drei Milliarden überweist BP in bar an die AAR, sobald der Gemeinschaftskonzern formell gegründet ist." (Quelle: Süddeutsche Online) Wären formal und formell hier austauschbar?

Fig. 1: Question about the difference between *formal*-*formell*

His or her question referring to the distinction between *formal* and *formell* is put into a specific linguistic context where someone needs to fulfil official requirements in order to become a member or, for instance, in order to obtain a permit. In a second note the user adds "I checked Leo and Wortschatz Uni Leipzig and couldn't come to a conclusive result"<sup>4</sup>. Then some examples are copied from the second resource and a further question follows "Can one mutually substitute *formal* with *formell*?" Figure 2 shows both entries in WortschatzLeipzig. Speakers will not successfully resolve their problems by using either entry in the NLP-tool, as they do not encounter a definition or any semantic information that can be used without further linguistic interpretation. Both items are polysemous and exhibit a range of semantic commonalities and differences. The entries, however, neither include senses and their differences nor correlate with any information about distinct usages. The examples lose their illustrative value when given as a block for a headword with many different options for contextual usage. Users cannot relate their existing knowledge and their specific query to this kind of entry without prior fine-grained disambiguation. The problem of assigning words to context is further increased by cross-referencing the headwords as synonyms and by referring to identical meaning equivalents (cf. Chon 2009, p. 28).

<sup>4</sup> Wortschatz Uni Leipzig is officially known as WortschatzLeipzig.

Wort: **formal** Anzahl: 2.358 Rang: 15.297 Häufigkeitsklasse: 13

Siehe auch: **Formal**  
 Wortart: Adjektiv  
 Grundform von: **formaler, formales, formalen, Formale, Formal, formale**  
 Synonym: unpersönlich, äußerlich, vorschriftsmäßig, bürokratisch, formell

▲ Dornseiff-Bedeutungsgruppen

11.10 Unterscheiden **formal, formalistisch, kritisch, spitzfindig**

▼ Formen mit ähnlichem Satzkontext: **formell** | aufgelöst | Flügel | offiziell | faktisch

▲ Beispiele

- Es sei aber richtig, dass Rassismus oft dort vorkomme, wo **formal** keine Augenhöhe bestehe. ([www.fr.de](#), gesammelt am 02.05.2021)
- Damit sind den Gewerkschaften **formal** die Hände gebunden. ([linkzeitung.de](#), gesammelt am 13.11.2021)
- Allerdings haben die Kommunen **formal** kein Vetorecht gegen das Projekt. ([www.die-glocke.de](#), gesammelt am 26.02.2021)
- Tatsächlich muss ich mich aber auch **formal** gar nicht unterordnen. ([www.wivo.de](#), gesammelt am 19.10.2021)
- Rein **formal** hat die Stadt mit dem Antragsverfahren erst einmal nichts zu tun. ([www.come-on.de](#), gesammelt am 15.09.2021)
- Es wäre demnach eine Blamage, **formal** an der Fünf-Prozent-Hürde zu scheitern. ([www.merkur.de](#), gesammelt am 17.09.2021)
- Der habe dargelegt, dass das Schreiben rein **formal** keine Beeinflussung der Wahl darstelle. ([www.fr.de](#), gesammelt am 17.03.2021)

Wort: **formell** Anzahl: 1.307 Rang: 24.516 Häufigkeitsklasse: 13

Siehe auch: **Formell**  
 Wortart: Adjektiv  
 Grundform von: **formellen, formelles, formeller, Formelle, Formell, formelle**  
 Synonym: **formal, geschäftlich, zeremoniell, unpersönlich, steif, äußerlich, formlich, konventionell, offiziell**  
 Antonym: **informell**

▲ Dornseiff-Bedeutungsgruppen

9.78 Umweg **brieflich, bürokratisch, formalistisch, formell, formlich, indirekt, mittelbar, umständlich, weitschweifig**  
 12.28 Behaupten, bejahen **apodiktisch, ausdrücklic, bejahend, endgültig, entschieden, ernstlich, feierlich, formell, kategorisch, klar, offiziell, positiv, unfehlbar, untrüglich**  
 16.30 Höflichkeit, Gruß **artig, aufmerksam, charmant, devot, diskret, ehrerbietig, feinsinnig, formell, galant, gebildet, geschillien, gesüß, höflich, korrekt, kultiviert, lakvoll, verleinet, weifmannisch, wohlherzogen**  
 16.66 Schaustellung **abgemessen, formell, steif**

▼ Formen mit ähnlichem Satzkontext: **formal** | offiziell | Repräsentantenhaus | endgültig | Kongress

▲ Beispiele

- Rein **formell** bräuchte man die Zustimmung der beiden Landeschauptleute nicht. ([www.sn.at](#), gesammelt am 10.11.2021)
- In den letzten Wochen hat die Partei diese Position ihrem Sicherheitsdirektor **formell** mitgeteilt. ([www.tagesanzeiger.ch](#), gesammelt am 18.06.2021)
- Sie ist **formell** dem Bundesgesundheitsministerium unterstellt. ([www.tagesspiegel.de](#), gesammelt am 13.04.2021)
- Eher ein bisschen zu **formell**. ([www.welt.de](#), gesammelt am 27.04.2021)
- Al-Kaida soll er sich **formell** erst später angeschlossen haben. ([www.vienna.at](#), gesammelt am 11.09.2021)

Fig. 2: Entries *formal* and *formell* in WortschatzLeipzig

Broadly speaking, the resource gives the impression that both items are almost identical, apart from the obscure fact that *formell* is embedded contextually in more thematic domains, as this lexeme is listed in four different meaning sets taken from the onomasiological dictionary DORNSEIFF whereas *formal* is only documented in one thematic group.

WortschatzLeipzig is a computer-generated tool and it appears to be used in chats as a source for common language queries. Identifying particular lexical environments and domains is a prerequisite to decide whether two terms are contextually interchangeable. As information is not adequately differentiated and presented and is not entirely reliable without underlying editorial procedures, the resource must be deemed unsuitable to answer the initial question.

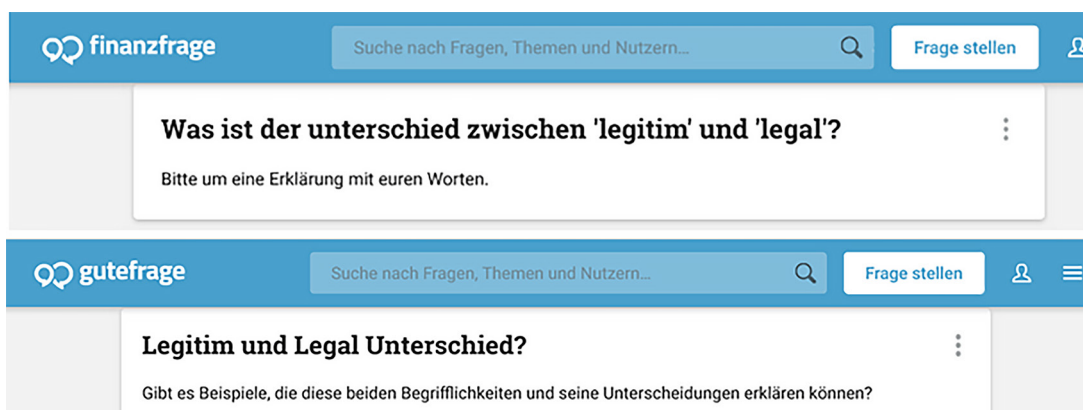
## 2.2 Language of description and examples

One main criticism addressed in discussion forums is the language used in reference guides. Although user-friendliness and usability in terms of descriptive style has long been addressed in meta-lexicography, we still face some old problems.<sup>5</sup> Three difficulties can be observed. Firstly, abbreviations serving as usage notes to indicate register or variation are not always familiar to everyone or are difficult to decode. Secondly, we need to question the comprehensibility of a concise, logical, and structuralist style of definition which follows a strict formula of *genus proximum* and *differentiae specifica*. Such definitions do not correspond to everyday language use and were originally established for print dictionaries. They can cause a situation where looking up one term requires an endless series of additional terms to be looked up (cf. Antor 1994, pp 78 f.). This style of entry has often been adopted as an inherent dictionary style even for online resources where the lack of space is irrelevant. Alternatively, some dictionaries use single synonyms to paraphrase the term without further explanations of syntagmatic restrictions. This tradition goes back to the notion of the referential substitutability of words. Cobuild's dictionary for language learners was the first project with an informal and discursive style of definition that used basic vocabulary (cf. Hanks 1987). Only a few dictionaries (e.g. *lexiko*) have reflected carefully on adopting a different definitional style (cf. Storjohann 2005), avoiding abbreviations altogether within the German context. Thirdly, most definitions lack extra-linguistic information, which is

<sup>5</sup> For an overview see Rothe (2001).

not a discrete category from linguistic knowledge about a word. As a result, a word’s meaning cannot be fully explored when no allowance is made for its designated discourse properties or referential domains.

Some traditional definitions found today appear antiquated and awkward or are difficult to understand, so that the description provided fails to be received meaningfully by the user. As a result, users request explanations of meanings in “your own words” or concrete language examples to show words in context (see Fig. 3).



**Fig. 3:** Request for clarification of the difference between *legitim* and *legal* “in your own words”

The first request implies a discernible difference between *legitim* (*legitimate*) and *legal* (*legal*). The second request involves the search for examples to clarify an assumed difference between the two items. A closer look at the definitions and examples of the headwords in the DWDS, as summarised in Table 1, show both words are defined by synonyms which themselves contain the paronyms *gesetzmäßig* and *gesetzlich*.

<b>legitim</b>	<b>legal</b>
gesetzmäßig, rechtmäßig <u>Beispiele:</u> <i>eine legitime Macht, Regierung mit legitimen Mitteln arbeiten</i> [jemand] der keinerlei Ansprüche stellte auf legitime Zuzugsgenehmigung [Kasack, Stadt, 579]	dem Gesetz entsprechend, gesetzlich <u>Beispiele:</u> <i>eine legale Regierung, Partei etw. auf legalem Wege tun</i> <i>legal handeln</i> <i>Devisen legal erwerben, umtauschen</i>
ehelich <u>Beispiele:</u> <i>ein legitimer Nachkomme</i> <i>ein Kind für legitim erklären</i>	
<b>Synonyms</b> berechtigt · dem Recht entsprechend · erlaubt · gesetzeskonform · gesetzlich · legal · legitim · nach Recht und Gesetz · nach dem Gesetz · rechtens · rechtlich einwandfrei · rechtmäßig · statthaft · zugelassen · zulässig ● rechtssicher	<b>Synonyms</b> berechtigt · dem Recht entsprechend · erlaubt · gesetzeskonform · gesetzlich · legal · legitim · nach Recht und Gesetz · nach dem Gesetz · rechtens · rechtlich einwandfrei · rechtmäßig · statthaft · zugelassen · zulässig ● rechtssicher

**Table 1:** Dictionary information for *legitim* and *legal* in DWDS

As such, the user now actually needs to know the precise difference between two confusables. Synonyms paraphrasing a headword can be useful, but they need to be chosen carefully or further substantiated with additional lexicographic data. The example, e. g. *eine legitime Macht/Regierung* vs. *eine legale Regierung/Partei*, are quite similar, and the identical synonym groups at the end suggest a meaning overlap in at least one “shared” sense. The only difference recognisable is one additional sense (‘ehelich’ (‘in wedlock’)) for *legitim* when referring to humans. As will be shown in 4.2, *legitim* and *legal* are, in fact, not meaning equivalents at all. The definitional style and the examples used in the DWDS create an inadequate impression about their use.

Although the examples given in 2.1 and 2.2 refer to forums where native speakers exchange their thoughts, comparable questions are found in forums designed for language learners who address the difficulties they encounter when faced with learner dictionaries. Members of the general public participating in discussion forums recommend specific dictionaries and explain why they should be used or avoided. The answers also provide insight into speakers’ intuitions, their linguistic and encyclopaedic knowledge, and their beliefs as well as their reactions to vague or strictly prescriptive suggestions. The best chats reveal the final decision on the lexical choice (and the reasons for it) based on different comments left in the forum.

### 3. Impetus for a new paronym dictionary

A few years ago, the Leibniz-Institut für Deutsche Sprache initiated a dictionary of confusables, the first corpus-assisted online guide to German paronyms. As far as German lexicography is concerned, it was the first time a dictionary project had based its lexicographic contents, design, and functionality on users’ interests and expectations as derived from forums and by examining reports on individual instances of dictionary consultation (Storjohann 2016).<sup>6</sup> In the planning process, the project was interested in the target users, their linguistic competence, expectations, and experience with lexicographic data, and any conflicts with their own intuition etc. Through more cognitive-oriented studies of users we were able to include in the dictionary what users specifically demanded in their chats. Over 200 discussions on paronyms, including questions and reactions, were subject to examination. Specifically, our interest focussed on who showed uncertainties in their use of confusables, what the communicative contexts were in which difficulties occurred, and where users looked the words up. Once we learned about general dictionary skills, we analysed how satisfied the users were with the information in traditional entries and whether they differed from their own introspections. In addition, particular attention was paid to what skills are used to draw upon different types of knowledge and how users expressed a wish for more encyclopaedic information. Another fundamental question raised in the project was how users react to both vague and prescriptive answers and what choices they make when they receive a number of divergent responses.

<sup>6</sup> The results of this study were only used for design purposes during the development of the new resource. The project still holds all data (chats written between 2002 and 2016) from this investigation.

The insights obtained played a central part in the planning process and led to new ideas and alternative lexicographic principles. One of the aims was to create a reliable and user-friendly tool by applying contrastive corpus-linguistic methods and by realising the demands of cognitive lexicography (e. g. Ostermann 2015). Another objective was to overcome some of the major dictionary-based problems by integrating innovative modes of presentation and by exploiting new technological possibilities. Sections 3.1 to 3.2 will provide a link to the challenges explained in 2.1 to 2.2 and show some lexicographic solutions to the lexical pairs mentioned in the forums. These mainly concern: how to quickly identify similarities and differences, how to combine sufficient linguistic and extra-linguistic knowledge, how to use new means of presentation, how to involve the user with interactive, adaptive functionality, how to choose a more accessible definitional style, and how to select examples best suited to illustrate context (and synonymy).

### 3.1 Depth and presentation of lexicographic information

The objective of producing a reliable source implies addressing contextual information in terms of ontological reference, collocability, and thematic domains in different contexts.<sup>7</sup> Overlaps and differences need to be clearly accessible and understood at first sight. Besides quickly accessing information, some users require further information which needs to be selective, customisable, and generated on demand. These prerequisites were put to the test a number of times in the initial stages of the dictionary. As a result, we created a two-level entry consisting of a contrastive overview and a more detailed level. Both levels contain interwoven lexical, semantic, and world knowledge about words, their senses, and conventions. Senses are understood to segment the overall meaning potential into meaningful units perceived as typical pattern choices from corpus analysis. In addition to developing ideas about contexts and depth of information and modifying the style of description, it was also essential to assess the technological options for presentation as well as the (visual) functionalities which assist in the design of the resource. In fact, forms of presentation and intelligent modes or functions allow for an efficient and intuitive navigational structure. They also support the explanations of the headword in many different ways, for example, by providing interactive guidance and user-adaptive choices and by changing the linguistic perspective.

The focal point of the contrastive overview are the headwords and their contextual uses (each in a tile) encompassing the full semantic spectrum of the word and signalling its context-boundedness to users as detected in the underlying corpus. The slots/positions and colour marking of the tiles help to identify the relationships between the senses of the corresponding partner term(s) (cf. Fig. 4).

<sup>7</sup> For a detailed account of the German paronym dictionary, see Storjohann (2018).



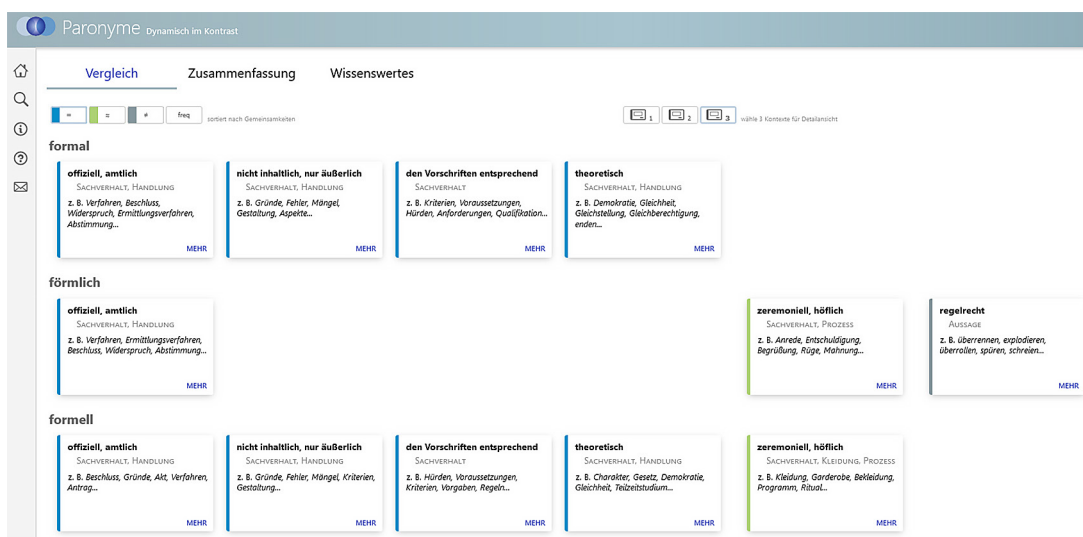


Fig. 4: Overview of senses of *formal/förmlich/formell* in the paronym dictionary

*Formal/förmlich/formell* are presented beneath each other with the most frequent term at the top of the entry. Their senses are in line horizontally and placed directly beneath each other when semantically identical or similar, but are offset when different. In cases where no equivalent sense exists, the allocated slots remain empty. The colour scheme further indicates whether senses are classified as being identical (blue), similar with slight semantic nuances (green), or distinct (grey). As such, the type of senses can be identified, arranged, and set into relationships with others. The semantic spectrum of the items is concisely depicted, and one can instantly grasp that the adjectives are polysemous with contextual overlaps and differences between them. A short description is provided for each sense, and the tiles also serve as points of contextual navigation to further detail. Looking at a tile more closely, it reveals the following information:

**formal**

- (1) 'offiziell, amtlich'
- (2) Sachverhalt, Handlung
- (3) z.B. *Verfahren, Beschluss, Widerspruch, Ermittlungsverfahren, Abstimmung*

A synonym (or two) serves as a label for an identified context, while general reference is provided by ontological categories (e.g. STATE OF AFFAIRS, PERSON, PROCESS, SITUATION), and these are exemplified by collocates illustrating lexical realisations of the more abstract reference. Its content can be read as follows: *Formal* means 'official' (1) and it refers to STATE OF AFFAIRS and PROCESSES (2) such as a *procedure, decision, objection, investigation, or voting* (3). Together these discriminate sufficiently the contextual uses from each other. Via a menu, the senses can be rearranged flexibly depending on whether the focus is on commonalities, differences, or frequency. Knowing the precise circumstances under which both expressions (better: their senses) are interchangeable can require more detail. This can then be selected individually at the detail level, where information is more extensive and where users can study that detail side by side (Fig. 5).

formal, offiziell, amtlich (3)	förmlich, offiziell, amtlich (2)	formell, offiziell, amtlich (1)
beschreibt einen Sachverhalt oder eine Handlung dahingehend, offiziell, amtlich und regelkonform bzw. gesetzesgemäß zu sein	beschreibt einen Sachverhalt oder eine Handlung dahingehend, offiziell, amtlich und regelkonform bzw. gesetzesgemäß zu sein	beschreibt einen Sachverhalt oder eine Handlung dahingehend, offiziell, amtlich und regelkonform bzw. gesetzesgemäß zu sein
<b>Was ist formal?</b> Verfahren, Beschluss, Widerspruch, Ermittlungsverfahren, Abstimmung, Anhörung, Ermittlungen	<b>Was ist förmlich?</b> Verfahren, Ermittlungsverfahren, Beschluss, Widerspruch, Abstimmung, Ermittlungen, Anhörung, Bauabnahme, Festlegung, Umweltverträglichkeitsprüfung	<b>Was ist formell?</b> Beschluss, Gründe, Akt, Verfahren, Antrag, Entschuldigung, Entscheidung, Beschwerde, Vertrag
<b>Was macht man formal?</b> einleiten, beschließen, ablehnen, entscheiden, besiegeln, beenden, anklagen, abstimmen, zustimmen	<b>Was macht man förmlich?</b> beschließen, beenden, einleiten, anklagen, entscheiden, besiegeln	<b>Was macht man formell?</b> einleiten, beschließen, beenden, anklagen, besiegeln, entscheiden, ablehnen, abstimmen
<b>Verwendungsbeispiele</b>	<b>Kontextmuster</b>	<b>Kontextmuster</b>
Kontextmuster – ein formales [Verfahren] einleiten – einen formalen Widerspruch gegen [...] einlegen	– ein förmliches Ermittlungsverfahren gegen [...] einleiten – einen förmlichen Beschluss fassen – die Einleitung eines förmlichen [Verfahrens / Ermittlungsverfahrens] – einen förmlichen Widerspruch gegen [...] einlegen	– aus formellen Gründen ablehnen – ein formelles Ermittlungsverfahren einleiten – die Aufnahme formeller (Koalitions-)Verhandlungen – rein formell betrachtet
<b>Belege</b> Wie geht die Echa dann vor? Wir suchen den Dialog, gehen auf Unternehmen zu und geben ihnen Gelegenheit nachzubessern. Geschichte das nicht, müssen wir jedoch ein aufwendiges formales Verfahren vorbereiten. (VDI nachrichten, 27.05.2011, S. 7, „Stichproben zeigen, dass manche Firma noch nachbessern muss“.)	<b>Belege</b> Die Räte Wesel und Xanten können gemeinsam mit ihren Verwaltungen ein förmliches Verfahren in die Wege leiten und an Deichverband und Bezirksregierung einen Antrag auf Befreiung vom Verbot der Nutzung der Deichkrone als Radwanderweg stellen. (Rheinische Post, 03.07.2003, Schilkanen.)	<b>Belege</b> Stadtsprecherin Ilka Richter erklärt: „Für die Namensänderung ist ein formelles Verfahren notwendig. Der Stadtrat muss dazu einen Beschluss fassen. Dieser wird derzeit im Fachbereich Schulen und Soziales vorbereitet.“ Zur letzten Stadtratssitzung vor der Sommerpause soll er dann endlich eingereicht werden. (Leipziger Volkszeitung 21.01.2005 S. A1.)

Fig. 5: Details of the sense ‘official’ shared by *formal/förmlich/formell*

The paraphrase here is longer, with the reference categories embedded into further relevant contextual information. More collocates open further contextual options, and these are classified according to word class to show their syntagmatic role (similar to semantic frame organisation). These also help the user to avoid violating conventional collocational patterns. Together, they create an interplay of lexical and non-lexical information. Corpus examples, typical construction patterns, and synonyms/antonyms allow for further comparison and illustration. As looking up paronyms often occurs in situations of text production, locating diverse and comprehensive information on a specific word is essential. For such activities, Lew (2015, p. 9) remarks:

The lexicographic treatment should be more detailed than for text reception, allowing the dictionary user to construct natural phrases and sentences with the headword. To that end, the user will typically need guidance on syntactic patterns into which the headword enters, as well as collocates, preferably with examples of use to serve as a model for production.

Deciding what the essential type and the necessary depth of detail are, as well as where to present information and how to integrate sections generated on demand, has turned out to be highly complex also with respect to editorial practice. The editorial process includes the analysis and interpretation of corpus data, the discrimination of senses, the allocation of data to each sense, and the assigning of uses to headwords and to their relevant senses of the paronym by coordinating information in a specific way. As a result, linguistic and extra-linguistic information is more explicit, interlinked, and consistently illustrated, and all entries are harmonised. The four major display elements suitable for contrastive entries are: colour, positioning, sorting principles, and user-generated selection options. They support users in identifying, comparing, and setting new parameters, in changing perspective, and in choosing the relevant parts that are expandable. These functions and modes of presentation and visualisation are not superficial gimmicks, but rather they add valuable information to the descriptions.

### 3.2 Language of description and examples

Although the two polysemous terms *legal/legitim* both refer to the concept of law (see 2.2), they are not used synonymously, as we can see by analysing actual instances of real language use in corpora. Their individual and distinct contextual uses are therefore placed offset from one another in order to indicate that they are not in a relationship of similarity. Labels that are different enough justify the plausibility of the distinction between senses. Again, the combination of headword, synonym, reference category, and illustrative collocates specifies the contextual environments (Fig. 6).

The screenshot shows the 'Paronyme' dictionary interface. The main heading is 'Paronyme Dynamisch im Kontrast'. There are three tabs: 'Vergleich', 'Zusammenfassung', and 'Wissenswertes'. Under 'Vergleich', there are three comparison options: '=', '≈', and '≠', along with a 'freq' button and the text 'sortiert nach Gemeinsamkeiten'. To the right, there are three icons labeled '1', '2', and '3' with the text 'wähle ein'. Below this, the word 'legal' is shown with two entries: 'gesetzeskonform' (SACHVERHALT, GEGENSTAND, HANDLUNG) and 'erlaubt' (PERSON). Under 'legitim', three entries are shown: 'berechtigt' (SACHVERHALT), 'rechtmäßig' (PERSON), and 'ehelich geboren' (PERSON). Each entry includes a list of collocates and a 'MEHR' link.

Fig. 6: Entry *legal/legitim* in the paronym dictionary

Different lexical information and extra-linguistic details are incorporated into the meaning explanations. These confirm the distinction where two senses of the expressions have been selected (here, two that appear similar at first sight, since they both refer to PEOPLE). The long paraphrase contains a certain amount of entrenched world knowledge. The term *legal* characterises a person in such a way that he/she possesses an official permit or that he/she has proof of a certain official status (e.g. residence permit) or that he/she can prove to be allowed to own specific objects (e.g. weapons). By containing these facts about the real world, the description does not remain abstract but becomes concrete and illustrative. Both adjectives also occur in different thematic domains, which are given beneath the definition (here LAW vs. SOCIETY/CULTURE). As has been pointed out in 3.1, the collocates further illustrate typical syntagmatic constructions as well as realisations of the conceptual reference. People who have been characterised by *legal* can be *Einwanderer* (*migrants*) or *Waffenbesitzer* (*owners of weapons*). Entities which are modified as being *legitim* (*legitimate*) are *Nachfolger* (*successors*), *Erben* (*heirs*), and the *Regierung* (*government*).

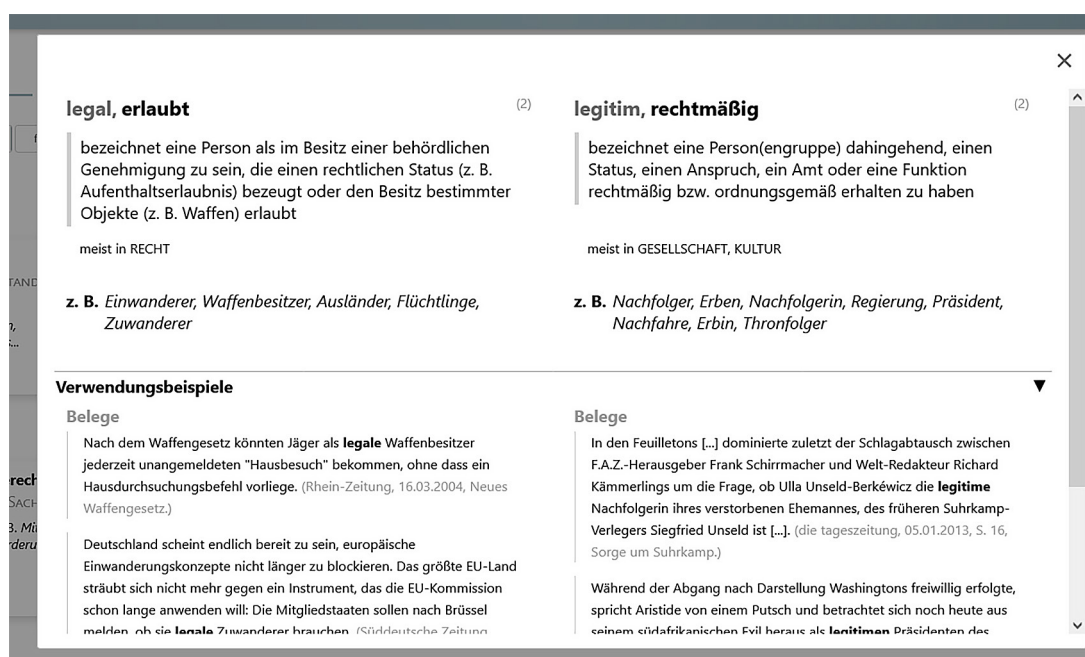


Fig. 7: Two distinct senses in the contrastive detail view

Choosing the right corpus examples entails following a number of different criteria. One of them is to have a context where the headword co-occurs together with some of the collocates given above. In the case of equivalent contexts between two usages, the examples must also contain identical patterns (see Fig. 5 *formales/förmliches/formelles Verfahren*). This practice is most effective in providing evidence of collocability, grammatical features, and context-bound near-equivalence with corresponding headwords. By choosing longer definitions with a style closer to everyday language and by avoiding abbreviations altogether, necessary information can be expressed in a more comprehensible manner. The language of the description is more extensive and includes details illustrating and referring to elements of the definition. This approach guarantees a more descriptive and coherent depiction of lexical facts combined with the necessary real-world knowledge.

#### 4. A new dictionary of neologisms (and their synonyms)

The architecture developed for the paronym dictionary is transferable to the description of synonyms in large measure because (near-)synonyms can also cause difficulties as far as their precise differences are concerned. As a next step, we will develop new resources describing German neologisms, including new synonyms such as *Lockdown/Shutdown*, *Corona/Covid/SarsCoV-2* or new loan words with their indigenous counterparts (e. g. *Prank/Streich (prank/prank)*). In addition to questions which typically arise for neologistic synonyms, there are similar questions concerning how or whether to distinguish between them. The core feature of neologisms is being new, and therefore they have the potential to be unfamiliar and not yet established in a speaker's mental lexicon. Their assimilation into German might be an ongoing process. Hence, changes as to their adoption of gender, inflectional paradigms, connotations, or even reference are still possible. In Figure 8, a user has a query asking for the difference between the nouns *Covid-19* and *Corona*. S/he provides additional information on an underlying situation that involves seeking details on reference and context: here, the use of both items in terms of a person affected by the disease.

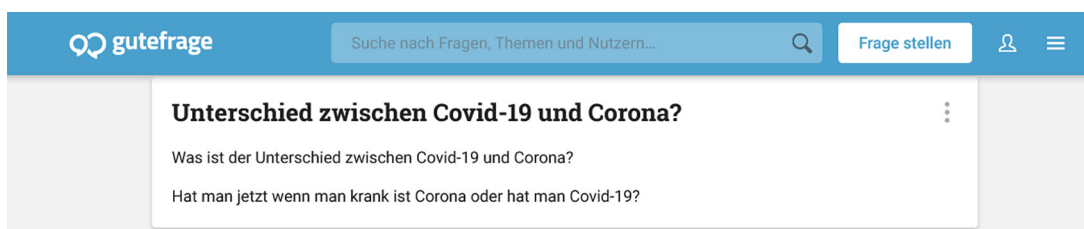


Fig. 8: User asking for the difference between *Covid-19* and *Corona*

In Wiktionary<sup>8</sup> they are both paraphrased as a “disease caused by an infection with Sars-Cov-2”, suggesting semantic identity. Fundamental details on the specific reference of the terms are missing, e.g. who is exposed to it, what type of disease it is, and what medical indications or symptoms typically occur. With regard to medical terms, encyclopaedic knowledge is an important part of their semantics and is often sought in queries.

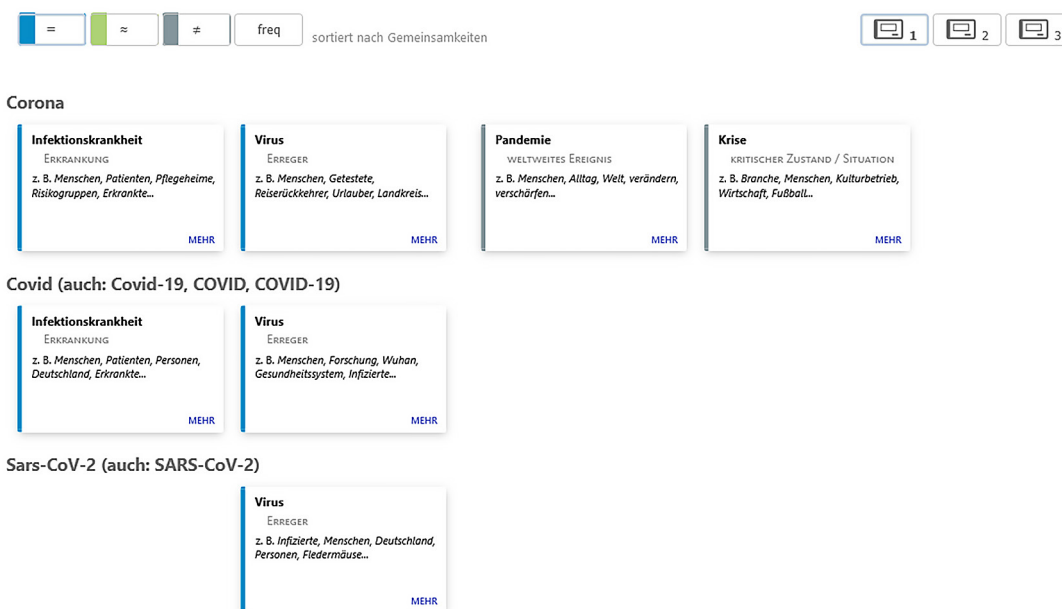


Fig. 9: Neologistic synonyms and collocations

Serving as an example, a fictitious entry including an overview and detailed documentation has been created (Fig. 9) to show the value of a contrastive entry on the basis of the existing paronym dictionary, also illustrating synonymous contexts in everyday language. The information is provided in a similar way, sufficiently disambiguating the senses for each headword. In order to recognise a contextual use, the synonym label (‘Infektionskrankheit’ or ‘Virus’) serves as a usage identifier to refer to information relevant for a specific communicative setting. In this case, for *Corona* one context referring to a specific virus and another referring to the infectious disease, a pandemic and a social crisis can instantly be identified. As outlined previously, additional information in both the overview and the detailed view pinpoint particularities concerning who is affected, any activities caused by the virus or

<sup>8</sup> Wiki has a new resource WikiUnterschied.com, published in 2021, which creates comparative entries on the basis of its wiktionary entries (hence similar to table 2). So far, there is no mention of this resource in forums.

alternatively by the disease), and what is typically associated with it as expressed lexically by collocates.

Corona, Infektionskrankheit <sup>(2)</sup>	Covid (auch: Covid-19, COVID, COVID-19), Infektionskrankheit <sup>(1)</sup>
bezeichnet eine durch das Virus SARS-CoV-2 ausgelöste Infektionskrankheit, die oft grippeähnliche Symptome verursacht (wie z. B. Fieber, Husten, Schnupfen), bei schwerem Verlauf besonders die Lunge oder das zentrale Nervensystem angreift	bezeichnet eine durch das Virus SARS-CoV-2 ausgelöste Infektionskrankheit, die oft grippeähnliche Symptome verursacht (wie z. B. Fieber, Husten, Schnupfen), bei schwerem Verlauf besonders die Lunge oder das zentrale Nervensystem angreift
<b>Wer oder was ist von Corona betroffen?</b> <i>Menschen, Patienten, Pflegeheime, Risikogruppen, Erkrankte, Genesene</i>	<b>Wer oder was ist von Covid betroffen?</b> <i>Menschen, Patienten, Personen, Deutschland, Erkrankte, Verstorbene, Bewohner, Wuhan</i>
<b>Was macht bzw. was macht man in Bezug auf Corona?</b> <i>erkranken, sterben, ausbrechen, versterben, verharmlosen, überstehen, genesen, impfen, (sich) anstecken</i>	<b>Was macht bzw. was macht man in Bezug auf Covid?</b> <i>erkranken, sterben, kämpfen, behandeln, auslösen, (sich) anstecken, impfen, genesen</i>
<b>Was hat jemand mit Corona?</b> <i>Symptome, Lungenentzündung, Verlauf, Herzmuskelschäden, Langzeitfolgen, Spätfolgen, Gedächtnisprobleme, Immunreaktion, Folgen</i>	<b>Was hat jemand mit Covid?</b> <i>Langzeitfolgen, Symptome, Spätfolgen, Gedächtnisprobleme, Immunreaktion, Folgen</i>
<b>Wie ist Corona?</b> <i>gefährlich, schlimm, schrecklich</i>	<b>Wie ist Covid?</b> <i>neuartig, gefährlich, grippeartig, tückisch, tödlich, real</i>
<b>Was wird im Zusammenhang mit Corona thematisiert?</b> <i>Quarantäne, Lockdown, Einschränkungen, Infektion, Lungenkrankheit, Krankheit, Impfung, Intensivstation, Fälle, Ausbruch</i>	<b>Was wird im Zusammenhang mit Covid thematisiert?</b> <i>Lungenkrankheit, Ausbreitung, Impfstoff, Coronavirus, Todesfälle, Symptome, Weltgesundheitsorganisation, Grippe, Pandemie</i>

Fig. 10: Contrastive sets of collocates between neologistic synonyms

Certainly, the treatment of neologisms needs the kind of detail that typically characterises these terms. Essentially, these relate to the origin, etymology, and morphology of lexemes with foreign elements or the appropriate grammatical use of nouns (e. g. gender, genitive, and plural forms). Inflection paradigms are also important issues for adjectives and verbs. A large number of neologisms also require more discourse-based information (*Shutdown* vs. *Lockdown*) and information about where they first appeared. Nonetheless, the solutions found for the paronym dictionary still seem to serve some needs as far as the comparative aspect of (neologistic) meaning equivalents is concerned (and of synonyms where one term is a loan word). Currently, more studies are being performed looking at different linguistic situations for users for neologisms and loan words and their specific language-related requests, in order to acquire a more complete picture of the new target group.

## 5. Summary

Language-related deficiencies together with users' dictionary-based problems have not been studied thoroughly in order to improve and design new dictionaries. It is suggested that research on dictionary usage be combined with studies on actual instances of language use. Making adequate distinctions or finding the right word in a specific context when there is more than one option is a frequent subject in chats and blogs. Studying those offers an unprecedented wealth of information about language users, the challenges they face with various dictionaries, and their confusion with paronyms, (near) synonyms, loan words, and neologisms. User studies have assisted dictionary makers in learning about their users, a decisive step forward in building user-friendly resources. However, the insights gained

from the investigation of chats had an essential effect on the development of the paronym dictionary. Specifically, they have influenced the contents, presentation, functionality, and style of description. The design and solution-based approach applied in the paronym project centred on gaining a deeper understanding of the target users for whom we actually compile a dictionary (cf. Lew 2015).

Having recognised that users turn to discussion forums, one might wonder whether we still need dictionaries. The answer is “yes” because online forums also tell us about the community’s competence, their different intuitions, and their urgent search for reliable reference tools. Personal suggestions vary: often they are limited to prototypical or primary senses, or they are prescriptive, following old educational norms once learned or prevalent in traditional dictionaries. When we scrutinise the target user and his/her linguistic questions before we develop a new product and best combine it with studies of dictionary behaviour and when we redefine lexicographic boundaries and search for new possibilities, we are able to build new dictionaries that are reliable and educational while also being enjoyable to browse through.

## References

- Antor, H. (1994): Strategien der Benutzerfreundlichkeit im modernen EFL Wörterbuch. In: Henrici, G./Zöfgen, E. (eds.): Fremdsprachen lehren und lernen. Wörterbücher und ihre Benutzer. Tübingen, pp. 65–83.
- Chon, Y. V. (2008): The electronic dictionary for writing: a solution or a problem? In: International Journal of Lexicography 22 (1), pp. 23–54.
- Dornseiff (2020): Der deutsche Wortschatz nach Sachgruppen. 9th edition. Berlin/Boston.
- Duden: [www.duden.de](http://www.duden.de) (last access: 14-02-2022).
- DWDS: Digitales Wörterbuch der deutschen Sprache. <https://www.dwds.de/> (last access: 14-02-2022).
- lexiko*: Online-Wörterbuch zur deutschen Gegenwartssprache. <https://www.owid.de/docs/elex/start.jsp> (last access: 14-02-2022).
- Gutefrage.net. <https://www.gutefrage.net/> (last access: 14-02-2022)
- Hanks, P. (1987): Definitions and explanations. In: Sinclair, J. (ed.): Looking up – an account of the Cobuild project in lexical computing. London/Glasgow, pp. 116–136.
- Leo-Forum (Sprachlabor): <https://dict.leo.org/forum/viewGeneraldiscussion.php?idThread=3225&idForum=4&lp=ende&lang=de> (last access: 14-02-2022).
- Lew, R. (2015): Dictionaries and their users. In: Hanks, P./de Schryver, G.-M. (eds.), International handbook of modern lexis and lexicography. Berlin/Heidelberg, pp. 1–11.
- Müller-Spitzer, C. (ed.) (2014): Using online dictionaries. Berlin/Boston.
- Murphy, L. (2013): What we talk about when we talk about synonyms (and what it can tell us about thesauruses). In: International Journal of Lexicography 26 (3), pp. 279–304.
- Ostermann, C. (2015): Cognitive lexicography. Berlin/Boston.
- Paronyme – Dynamisch im Kontrast. <https://www.owid.de/parowb/> (last access: 14-02-2022).
- Rothe, U. (2014): Das einsprachige Wörterbuch in seinem soziokulturellen Kontext: Gesellschaftliche und sprachwissenschaftliche Aspekte in der Lexikographie des Englischen und des Französischen, Berlin/Boston, pp. 85–113.

- Rundell, M. (1999): Dictionary use in production. In: *International Journal of Lexicography* 12 (1), pp. 35–53.
- Storjohann, P. (2018): Commonly confused words in contrastive and dynamic dictionary entries. In: Čibej, J./Gorjanc, V./Kosem, I./Krek, S. (eds.): *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*. Ljubljana, pp. 187–197.
- Storjohann, P. (2016): Vom Interesse am Gebrauch von Paronymen zur Notwendigkeit eines dynamischen Wörterbuchs. In: *Sprachreport* 4/2016, pp. 32–43.
- Storjohann, P. (2005): Semantische Paraphrasen und Kurzetikettierungen. In: Haß, U. (ed.): *Grundfragen der elektronischen Lexikographie. *ellexiko* – das Online-Informationssystem zum deutschen Wortschatz*. Berlin/New York, pp. 182–203.
- Storror, A. (2013): Representing dictionaries in hypertextual form. In: Gouws, R./Schweickard, W./Wiegand, H. E. (eds.): *Dictionaries. An international encyclopedia of lexicography. Suppl. Vol.: Recent developments with focus on electronic and computational lexicography*. Boston/Berlin, pp. 1244–1253.
- WikiUnterschied.com. <https://wikiunterschied.com/> (last access: 14-02-2022).
- WortschatzLeipzig. <https://corpora.uni-leipzig.de/de> (last access: 14-02-2022).

## Contact information

### **Petra Storjohann**

Leibniz-Institut für Deutsche Sprache  
storjohann@ids-mannheim.de