

New Words for the *Duden*!

Kathrin Kunkel-Razum

Bibliographisches Institut GmbH

kathrin.kunkel-razum@duden.de

Abstract

A new printed edition of the *Rechtschreibduden* (*Duden's Orthographic Dictionary (Duden)*), the prestigious German language dictionary, is published every three or four years. The past five or six editions have boasted approximately 5,000 newly added lemmas each (the 27th edition, with 145,000 lemmas, is the most recent), and since 1996, the year of the “Rechtschreibreform” (national spelling reform), public response to each new edition has focused primarily on these new additions. Once a word is included in the *Rechtschreibduden*, it is considered to have become officialized. There are people who wonder whether words that are not included exist, although even its online version offers an additional 100,000 lemmas.

What then are the criteria applied by the *Duden*'s editorial staff when deciding which new words to include? Which sources are used? What is the editors' position in the ongoing discussion about the – arguably excessive – use of Anglicisms in the German language and the addition of terms and grammatical adaptations related to or dictated by political correctness? How about the ratio of new entries in the printed edition of the *Rechtschreibduden* as opposed to its online version, and what are the procedures for inclusion? And finally, on what grounds are words deleted from the dictionary?

In this paper I will refer to these issues and, with regard to future editions, I will also talk about which new sources the *Duden* will have to consider and work with to remain the predominant dictionary of the German (standard) language.

Keywords

German, orthographic dictionary, Anglicisms, print vs. online dictionary

Introduction

The commonly named *Rechtschreibduden* or *Duden 1* is Duden's orthographic dictionary and the most famous dictionary of the German language. The *Rechtschreibduden* is based on the official standard in Germany, which is the norm of the Council for German Orthography. The reason for its prominence and importance can be traced back to it being the first published work for a uniform German orthography, printed in 1880. This so-called *Urduden*, the ancient edition of the *Rechtschreibduden*, included 27,000 entries on 187 pages. Since then, *Duden*, which is the surname of the founder Konrad Duden, has established itself as the number one standard reference dictionary and thus as a leading authority for all issues concerning the German language and orthography.

When a word has been included in the *Rechtschreibduden*, it is considered to have become officialized and from that point onwards, it is "an official word". People often wonder whether words that are not included even exist (which cannot be the case, because how can a word "not exist"?).

In 1996 there was a huge reform of German orthography, the "Rechtschreibreform", which applied to Germany, Austria, Switzerland, Belgium, South Tyrol, Liechtenstein and Luxemburg. Since then, all new editions of the *Rechtschreibduden* have focused primarily on these rule changes and new additions.

The most recently published *Rechtschreibduden* is the 27th edition from 2017, which has 145,000 lemmas, with 5,000 added lexemes compared to the previous edition from 2013. The past five or six editions have each been extended by a similar number of new entries, mostly nouns. Here are examples of neologisms from the last three editions:

- 25th, 2009: *fremdschämen* ('the feeling of shame on someone else's behalf'), *twittern* ('to tweet, to use Twitter'), *Abwrackprämie* ('scrappage bonus')
- 26th, 2013: *Energiewende* ('transition in energy policy'), *Liebesschloss* ('love lock' (pad lock)), *QR-Code* ('QR code')

- 27th, 2017: *Ampelfrau* ('little traffic-light woman'), *hyggelig* ('having or feeling the quality of cosiness and comfortable conviviality'), *Späti* ('open-late store'), *postfaktisch* ('post-factual'), *Schmähgedicht* ('smear poem'), *Selfie*, *tindern* ('to use Tinder')

Criteria for Words to Enter the *Rechtschreibduden*

In order to be included in the *Rechtschreibduden*, a word should – in general terms – occur with a certain frequency in written language, over a longer period of time, preferably over several years. A word that is only used for a few days or weeks and which is very frequent in the media, a flash in the pan if you will, is one that we would wish to deliberately exclude.

In addition, the word should appear in various types of texts, in newspaper and magazine articles, as well as in novels, and possibly technical texts, etc. Therefore, it should not only be known in professional circles, but it should also have found its way into everyday language.

Our most important sources for the substance of the dictionary are the *Duden corpus* and frequency-based lists, which we review manually and evaluate lexicographically. The corpus currently contains 5.2 billion tokens, which mainly consist of newspaper texts, but also of fiction books, non-fiction texts and functional texts. Our corpus is continuously growing, with new material and data being added every day. For the process of a new edition of the *Rechtschreibduden*, new frequency-based lists are extracted from the corpus. Those lists are intended only for internal use and are not published. Computer linguists use certain criteria for creating those lists. The most important criterion, of course, is the frequency of the occurrence of a word in the corpus in a certain period of time. This period is defined as the time between the end of the manuscript work for the last edition and the present day. In addition, we use a filter for preposterous and flawed words. With the evaluation of those lists, different criteria interact. Initially, this is the dictionary type (e.g. hard-to-spell words for the *Rechtschreibduden*). For example, we don't usually include compounds with more than three parts in the *Rechtschreibduden*.

In addition, we also receive suggestions for new entries from our users and we closely follow the daily press and other trends and developments.

Our focus is on the written standard language. However, we also incorporate colloquial and youth language if these play an important role in the German language.

Following extensive discussions, we decided against including new lexemes in our dictionaries too quickly. This is because, with our printed dictionary, our aim is to focus on words that have established themselves over a longer period of time. We are of course expected to ensure that entries are up to date, but reliability and consistency are the most important factors. This decision is part of our concept, which also seems to be firmly established in the minds of our users. With each new edition of the *Rechtschreibduden*, journalists and language enthusiasts ask: Which words have “made it” into the *Duden*? This inclusion is sometimes mistakenly perceived as actual evidence of the existence of a word. Of course, such unprofessional views must be set straight. But our impression remains that our users understand the concept that a word can only find its way into the dictionary when it has been established in language use. In doing so, users naturally advocate the inclusion of certain words, ask questions, provide evidence for their use, etc.

To a large extent, the words we include reflect developments that have particularly shaped politics, society, economics, technology, sports or culture. They show how vocabulary permanently adapts to social, technical or cultural changes. New objects and phenomena must be named to ensure a functioning linguistic communication. Language always reflects social shifts and processes which we have to observe and understand in order to adapt our dictionaries accordingly. An example is the current sociopolitical discourse about gender and equality which has made us aware of the importance of listing all female word forms, for example of occupational titles (*Vorständin* and *Vorstand*, *Parkettlegerin* and *Parkettleger*, female and male forms respectively of ‘chairman’ and ‘parquet layer’).

But somebody needs to decide whether a word is a neologism or not. As I mentioned above, regarding the *Rechtschreibduden*, the editorial staff of the dictionary department take these decisions. Other dictionary publishers or institutions who want to keep the German lexicon up to date decide for themselves. Sometimes, a word labelled as “neologism” rarely appears in a corpus or even on the Web. There are institutions who immediately include these words to their lexicon or dictionary and some who refuse to act rashly. We at *Duden* try to only include words that represent steady and current language use.

Among these approximately 5,000 new entries per edition there are not just real neologisms, but many other – newly officialized – words. By “real” neologisms we mean:

- new lexical units (e.g. *prokrastinieren* ‘(to) procrastinate’ or *Emoji*)
- new meanings of words that are already listed (e.g. the verb *dampfen* in the context of using an e-cigarette ‘(to) vape’)
- new idioms

In order to be included in the *Rechtschreibduden*, the candidates must occur frequently over a longer period of time, be established and accepted in the language community, i.e. be part of language use, and be common and familiar. Other groups of new entries in the *Rechtschreibduden* are:

- lexemes, meanings and idioms that are older, but have not yet found their way into the *Rechtschreibduden* for various reasons. (This might occur following a methodical examination of certain groups of words. For example, an examination of the names of cities in Germany revealed that five cities – *Bergisch Gladbach*, *Hagen*, *Herne*, *Salzgitter* and *Siegen* – were missing, and these were then incorporated into the dictionary.)
- older words whose frequency has significantly increased due to certain events or developments in language usage (e.g. *futschikato* ‘down the drain’, *Kopfkin* ‘inner/mental cinema’, and *Namaste* (the (Indian) closing and greeting gesture))

- proper names, e.g. of places or people. (These are traditionally listed in the *Rechtschreibduden* but they are not captured systematically; in the last edition the names of Chancellors and Federal Presidents were completed.)
- new spelling variants, such as new approaches to words with the prefix *Ko-* / *Co-*. (For example, the 26th edition (2013) only had *Koautor* ‘coauthor’ but now the recommended spelling is *Co-Autor* or *Coautor*, although *Ko-Autor* and *Koautor* are also accepted.)

Moreover, there are numerous new additions of compounds with a relatively transparent meaning, such as *karrieregeil* ‘career-obsessed’ or *Nachtshopping* ‘late-night shopping’. They are included because the *Duden* aims to represent a mixture of vocabulary types and a solution for questions about spelling. Words such as *Mütterrente* ‘mother’s pension’ are transparent in their formation, but can only be understood with additional encyclopedic information.

When we speak of 5,000 new dictionary entries, we only mean the newly recorded lexemes. Unsurprisingly, these are mostly nouns, and there are far fewer adjectives, verbs and words from other grammatical categories.

Deleted Words

Dictionaries do not and cannot list all the words which have ever been used in all periods of history. We decided, for example, that the *Rechtschreibduden* should always consist of one volume only, to ensure that it remains manageable and convenient, and to prevent it from becoming too unwieldy, as it is already part of a 12-volume series. This means having 1,300 pages at most, so lemmas must be deleted frequently to make room for new words. Thus, over the years, many words have disappeared from the *Rechtschreibduden*¹, because they are hardly used anymore due to political progress or technical and sociological change. For

¹ However, we do not delete words from *Duden online*.

example, due to the downfall of the Third Reich, after World War II, and the end of the Nazi regime, Nazi-related terminology was no longer used in general language. There were also long-term social developments related to the division of Germany into East and West which left traces in German vocabulary and how this was recorded in dictionaries published in East and West Germany before reunification. In addition, there are many words that are no longer common and have therefore ceased to “exist”, although they are aesthetically pleasing to some people.

Here are some examples of lemmas which have been deleted from the *Rechtschreibduden*:

- 10e/1929: *naszieren* (‘to be born’)
- 12e/1941: *Nachmittagsruhe* (‘siesta in the afternoon’ – compared with *Mittagsruhe* ‘siesta at noon’)
- 25e/2009: *verschimpfieren* (‘to revile/insult’)

We also have many English words which have become obsolete:

- 10e/1929: *Penny-a-liner*, *Maidenspeech*, *Lord-Lieutenant*, *Logotype*
- 12e/1941: *schampuen* (‘to shampoo’)
- 16e/1961: *Centweight*
- 19e/1986 (w²): *Native* (was deleted, but has now returned as part of *Digital Native*)

Examples of deleted Nazi-related terminology:

- 13e/1947: *Volksschädling* (‘nation’s wrecker/varmint’), *fremdrassig* (‘belonging to a foreign race’), *Blutschutzgesetz* (‘law to protect the blood’), *Schutzstaffel* (‘SS’), *Hitlergruß* (‘Hitler salute’). (The term *Hitlergruß* has not disappeared from spoken and written language and is still included in larger dictionaries, but in 1947 the radical decision was made to delete it from the *Rechtschreibduden*, along with many other such words.)

² w = West German *Duden*.

- 18e/1980 (w), 14e/1951 (e³): *Rassenhygiene* ('race hygiene')

Furthermore, separate East and West versions of the *Duden* were published independently between 1951 and 1991. There is also specific GDR⁴ vocabulary, such as *Zeitkino* (a kind of cinema), *nachnutzen* ('to exploit'), *Sprachkabinett* ('language lab') and *Kaderakte* ('personnel records'), which has been deleted in more recent editions of *Rechtschreibduden*. But, of course, we still have historical and political words from that era, such as *Stasi*, *Stasi-Akten* ('state security files') etc. in our dictionaries.

Rechtschreibduden Compared to Duden Online

As specified above, currently about 145,000 lemmas are covered in the *Rechtschreibduden*. In addition to that, there is an online dictionary, *Duden online*, which includes many more entries. With about 237,000 lemmas, *Duden online* is the largest lexicographic database of contemporary German in existence. It is updated on a regular basis (3-4 times a year), with extensive technical and computational linguistic support. The content of *Duden online* is a mixture of all the entries of *Rechtschreibduden*, *Duden-Universalwörterbuch* (a comprehensive monolingual general dictionary of German), *Duden-Fremdwörterbuch* (a dictionary of German loan words), and some other words that are considered to be helpful for users who want to quickly check the spelling or meaning of a word online. Entries compiled only for *Duden online* include, for instance, *Hundenahrung* ('dog food'), *Außerirdischer* ('alien/extraterrestrial') and *fünftes* ('fifth'). As a result, more neologisms are listed in our online version as part of its additional 100,000 keywords compared to the *Rechtschreibduden*. Besides neologisms, the words added to *Duden online* include older words which have not previously been recorded in the online dictionary. All in all, the process of adding entries to or deleting them from *Duden online* is not easier (technically), but it is quicker than with the print dictionaries – which, it is worth mentioning, are also available in a digital version, called

³ e = East German *Duden*.

⁴ GDR = East Germany

Duden-Bibliothek (library). Nevertheless, entries cannot always be completely up to the minute.

Editorial Policy on Anglicisms

Of course, many new German words stem from English or consist of English-German compounds. In principle, there is no limit to the number of Anglicisms in our dictionaries, as is the case with other (purely German) words. Calculated on the basis of the total lemma stock of the *Rechtschreibduden*, the proportion of Anglicisms in the most recent editions remained relatively stable at about 3.7%, and a little higher in *Duden online*. The proportion is higher in *Duden online* because in the last two years numerous Anglicisms have evolved, and we can react more quickly if we want to add a new word to the online version.

In the past, German borrowed a lot of words from Latin, Greek and French, many of which have disappeared over time, whereas today more and more English words find their way into daily usage. This is mostly due to technological processes, such as social media, and globalization led by English. We do not see this as a disadvantage, but many language purists have fought and continue to fight Anglicisms.

The language user seems to know when it is appropriate to “Germanize” a borrowed word from another language or when it is perfectly acceptable to just adapt the original form/word/lemma. *Duden*’s job here is to make the new words compatible with the official German orthography rules. A related question is how strictly we should follow these rules, especially as we are regarded as an authority, and when to decide that it is time for a new rule or a change. This does not necessarily have to be anything like the German Spelling Reform, but could occur in any instance in which the common rule has become rather old-fashioned or is not adaptable to all cases. An example would be the debate on spelling compositions with a hyphen between the different components or writing them as one word, because one is the German way and the other is not (e.g. *Golfplatz* vs *Golf-Platz* ‘golf course’, *Schweineschnitzel* vs *Schweine-Schnitzel* ‘pork schnitzel’).

Acceptance by the Community, Anglicisms and Political Correctness

There will always be conservative language purists who fight Anglicisms. Does this mean that the German language is developing in a non-desirable way? Who decides? Who insists on this view? We do not, and we are glad that more and more people understand that there is no such thing as (qualitative) language decline. On the contrary, language works perfectly well because these developments are all part of (diachronically and synchronically recordable) language change. The language user speaks (or rather writes), and we record.

As editors for the German language, we aim to consider current developments, and we follow discussions in the media and monitor not only language use but also possible changes which might be important to our society. Language reflects contemporary (and past) events, which explains our current thinking about the sociopolitical discourse regarding gender and equality mentioned above. We try – at least in *Duden online* – to record all female word forms, delete inappropriate or outdated definitions or make reference to the fact that certain usages are discriminatory. Thereby, we always try to stay descriptive and, of course, consider the official norms.

But the discussion about our language does not end with Anglicisms or gender-neutral language (which is of course not an issue for all languages). There are always those who would rather fight against all new possible (language-related) developments, which could mean more work and inconvenience for the average (language) citizen in their opinion. This is a constant topic in the media. The *Duden* editorial staff always look at language descriptively. This does not mean that we have no opinion on important and relevant issues. As a commercial publisher, we can represent and convey a descriptive way of looking at language and at the same time use our public voice to make people think about the (German) language in a carefully considered, differentiated and open-minded way.

New Sources and the World Wide Web, or New Challenges of (Digital) Lexicography

So, all in all, detecting neologisms is an everyday matter for us at work, which can sometimes be very challenging. With the help of corpora, editorial evaluation of media, and users, we are able to add important words to new print editions and our online dictionary. When considering the future and digitalization, we need to discuss the scope of opportunities for new sources and corpora.

This gives rise to a number of questions. For instance, in which direction shall we expand our corpus and how do we obtain new sources? What does it mean to “document the contemporary (German) language”? *Where* is this language being used? In which contexts? Only in newspapers? In Wikipedia articles? What about chatrooms? Shall we focus on documenting only the *written* contemporary (German) language rather than spoken language?

If we adhere to the goal of consistently checking new additions to our *Duden corpus*, we can only track the frequency of certain words being used – as a result – delayed in time. For all neologism candidates for our dictionaries, their frequency (development) in the *Duden corpus* is checked. The number of times we review those words is determined by the type of each dictionary, but there are no absolute figures. We must observe words for three to four years (the period between two editions) and weigh up the pros and cons of including them. For example, *Coolpack* ‘cold pad’ is not very frequent but its spelling was problematic and there were many questions from users about how to write it correctly, so we had to decide on the spelling and add it to the *Rechtschreibduden*. Using Internet texts and faster-growing corpora would mean more data, but this would also enable us to check the distribution in real time, so to speak, bearing in mind that the Internet as a corpus has its own difficulties.

There are also questions surrounding which mechanisms could be used to check whether words “with neologism potential” actually last for a longer period of time. And what should happen with “short-term words” when they soon fall out of use again? Will they be canceled immediately or filed in a separate folder? Are the emergence and disappearance of words documented in the dictionary entry? Should there be more specific information on the *Duden*

online website? Could it be confusing for the user if there were short-term words right in the middle of vocabulary with established lexemes? There are a lot of questions to discuss in the future.