



New loan words in the Neologismenwörterbuch: corpus-based development of lexicographic information for an online dictionary of contemporary German

Maike Park¹

Received: 27 January 2020 / Accepted: 25 April 2020 / Published online: 18 May 2020
© The Author(s) 2020

Abstract

The majority of new words in dictionaries are included following a certain period of time during which they have become more frequent in use and established morphosyntactic and orthographic features consistent with the language system they are borrowed into. In case of borrowed new words, inclusion often takes place at a transitional state of assimilation to the language system, where delayed orthographic or phonetic change cannot be ruled out and the differentiation between standard-conforming and non-standard orthographic word forms of a lemma oftentimes depends on the proximity between the writing systems of the donor and the recipient language. Following a brief overview of loan words and their lexicographical description in the Neologismenwörterbuch, a specialized online dictionary for neologisms in contemporary German, this paper presents findings of an investigative case study on dictionary entries for a neologism borrowed from a logographic language system and discusses the potential of a corpus-based description of new loan words.

Keywords Loan words · Neologisms · Internet lexicography · Print lexicography · Specialized lexicography

1 Lexicographic approaches to new loan words

For a dictionary maker, the question as to when a word borrowed from another language can be considered fully lexicalized marks only the beginning of further thorough considerations and conclusions. Among words that enter the dictionary, borrowed new lexemes pose a special challenge for dictionary makers, who are challenged to give detailed information on lexicographic decisions. Is *Caffè Latte* Italian-based or an English loan word? Does *gugeln* need to be included as a

✉ Maike Park
park@ids-mannheim.de

¹ Leibniz-Institut für Deutsche Sprache, Mannheim, Germany

spelling variant or an additional form of *googeln*? And why is *gugeln* considered to be a standard-conforming adaptation to German whereas *googlen* is highlighted as a word form in transition? In the age of the internet, users of electronic dictionaries demand for quickly consumable information, challenging lexicographers to provide a coherent and detailed description of a loan word's origin, adaptation and usage, while offering a medium that is easy to navigate and search through without overusing features offered by electronic lexicography. For contemporary German, a variety of studies has been conducted on borrowing phenomena with a majority of works focusing on anglicisms (cf. Busse 2004; Onysko 2007), novel word formation elements (cf. Dargiewicz 2013) or issues concerning the terminology applied to different types of borrowing (cf. Kirkness 1976). In the case of borrowed neologisms, inclusion into the dictionary often takes place during a transitional state of assimilation to the language system, where the differentiation between standard-conforming and non-standard orthographic word forms of a lemma oftentimes depends on the proximity between the writing systems of the giving and the receiving language. Previous studies have not taken into account to what extent lexicographical descriptions of recently borrowed loan words can provide information on the oftentimes not yet completed integration process of new words, in a way that is beneficial towards the dictionary user. Following an overview of loan words and their lexicographical description in the *Neologismenwörterbuch* (NWB), a specialized online dictionary for neologisms in contemporary German, this paper discusses issues concerning the grapheme–phoneme-correspondence of orthographic and phonetic information given by common German print dictionaries and evaluates the corpus-based approach to the description of new loan words applied in the NWB.

2 The *Neologismenwörterbuch*

The *Neologismenwörterbuch* (NWB) belongs to a number of specialized dictionaries for Contemporary German (cf. Quasthoff's *Neologismenwörterbuch* (2007) and the regularly updated online word collection *Wortwarte* that describe usage, meaning and origin of neologism. Its first editions for neologisms from the 1990s and 2000s were published as print versions in 2004 and 2013 and by 2014 the NWB became integrated into the online dictionary portal OWID (Online-Wortschatz-Informationssystem Deutsch) at the Leibniz-Institute for German Language in Mannheim in 2014. As a so-called “Ausbauwörterbuch”, a dictionary in process (cf. Schröder 1997), the NWB is frequently worked on and updated. Aside from yearly additions to the ongoing decade, it is possible to add entries for past decades later on and to alter existing entries at any time. It currently comprises 2055 lemma entries, spanning almost three decades from the beginning of the 1990s up until today, that provide information about meaning and usage, spelling and pronunciation, grammatical features, frequent word formation patterns and collocations, and details regarding the etymology or connotation of a neologism.

Neologisms are new lexemes (*Unverpacktladen* (‘store for unpackaged groceries’), *Hygge*), new meanings (*Lichterkerze* (‘candle-lit demonstration’), *Alltagsbegleiter* (‘attendant for daily routines’)), multi-word expressions (*Generation Y*, *freie*

Trauung (‘free wedding ceremony’)) or elements of word formation ([...] *-alarmismus* (‘-alarmism’), *Cyber-[...]*), that have emerged in a language at a certain point in time, spread, and are being accepted and used as part of the Standard German vocabulary (cf. Herberg 2004a: 337–338; Lemnitzer 2010: 66–67).

3 Loan words in the Neologismenwörterbuch

In this paper, the term *loan word* is defined in the broadest sense¹ and refers to words, elements of word formation, and multi-word expressions that entered the lexicon of a language as the result of a borrowing process. Types of *borrowing* include native speakers “adopting elements from other languages into their recipient languages” (Haspelmath 2009: 36) and producing lexical innovations that only appear to have been borrowed from a donor language, e.g. in the case of pseudoanglicisms.

Just like native neologisms, borrowed neologisms included in the NWB are detected, evaluated and described based on frequency, distribution and degree of lexicalization.² Since the majority of borrowed neologisms is characterized by an overall slower lexicalization, which can be attributed to specific morphosyntactic features of German (e.g. assignment of grammatical gender, development of an inflectional system for verbs and adjectives) (cf. Lemnitzer 2010) and their orthographical and phonetical alignment, new loan words are assessed in terms of the degree of their integration into the German language system and not only the degree of their assimilation to standard orthography and pronunciation.

The NWB includes borrowed neologisms

- that have adopted the German declension paradigm and are aligned with orthographic and phonetic German standards,
- that occur in (pseudo-) loan creations with native material (so-called hybrids, cf. Dargiewicz 2013),
- that are exact translations “of all components of a word from another language” (cf. Klosa-Kückelhaus 2019: 109).

Correspondingly, its lemma list comprises loan words (*Morphsuit*, *Skyr* (a yoghurt dish from Iceland) or *Qigong*) and elements of word formation (*cyber-/Cyber-[...]*), loan translations (*Blutdiamant* (Blood diamond) or *Waldbaden* (Japanese ‘bathing in the woods’)), pseudoanglicisms (*Candystorm*, an analogy to shitstorm) and loan meanings (e.g. *episch*, ‘epic’).

¹ Kirkness (1976) discusses issues with theoretical approaches concerning the differentiation between *Fremdwort* and *Lehnwort* in Contemporary German, which cannot be resolved by applying the English umbrella term *loan word* without further specification.

² cf. Klosa and Lungen (2018) for details regarding corpus linguistic tools applied to the detection and description of neologisms in the NWB.

3.1 Loan words by domains

Most of the loan words included in the NWB fill semantic gaps, where a borrowed denotat was borrowed together with its reference (Schippa 1992). While the internet and new technologies have long been known to have contributed to the sudden surge of English loan words at the beginning of the twenty-first century, neologisms borrowed from English and other languages continue to emerge in various domains where inventions and novelties demand for lexical change. Aside from new technologies and innovations regarding the internet, English loan words dominate in the domains economy/trade, new media (including social media activities) and fashion. The majority of loan words and loan meanings from other donor languages are distributed among categories concerning

- food(s) like *Pu-Erh-Tee*, *Macaron* and *Ciabatta*
- well-being, e.g. *Lomi-Lomi* or *Qigong*
- leisure, e.g. *Ken-Ken* and *Scoubidou*
- culture, e.g. *Strickguerilla*, ‘s.o. decorating trees with knittings’ or *Nikab*.

3.2 Loan words by language

A look at the lemma list in the Neologismenwörterbuch confirms previous findings (cf. Onysko 2007; Yang 1990) that English contributed to a majority of the borrowing phenomena in contemporary German during the past two and the current decade. As of December 2019, the evaluation of borrowed lemmas compiled from the NWB³ yielded a total of 868 new words, word formation elements and multiword-expressions borrowed from English. Additionally, 68 headwords in the NWB were classified as pseudoanglicisms, i.e. lexical innovations formed with English words or word formation components, which accounted for 7% of all borrowed neologisms. In contrast, the other foreign languages identified as donor languages accounted for only a small number of loan words (4%) and borrowed components in German word formation (2%). Among foreign languages included in the NWB, the majority of loan words stem from Japanese, followed by Chinese, several European languages (e.g. Italian, French, Danish, Swedish), Arabic and isolated cases of a single borrowing. With consideration to differences between direct and indirect language contact, words borrowed through English as a transfer language (e.g. *Churro* or *Barista*) were analyzed separately and yielded another 35 lemmas.

³ Since loan words in the NWB are classified according to the last language they have been borrowed from, the data presented in this paper has been reclassified to distinguish direct from indirect borrowing.

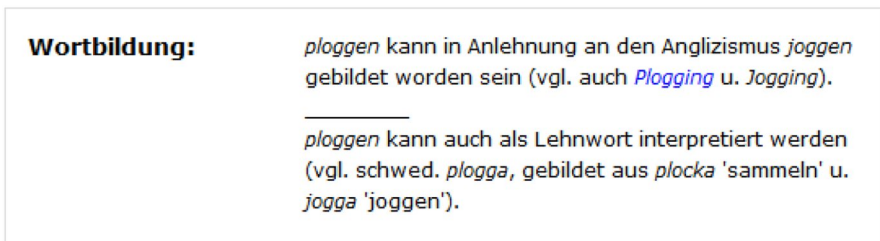


Fig. 1 Details on word formation in the dictionary entry for *ploggen* in the Neologismenwörterbuch (<https://www.owid.de/artikel/407888>)

4 Lexicographical description of new loan words in the Neologismenwörterbuch

Dictionary articles in the NWB contain lexicographic information on meaning, orthography, pronunciation, etymology, frequency, grammar, typical usage and word formation productivity of a lemma. Lexicographic descriptions are illustrated by examples from the *Deutsche Referenzkorpus (DeReKo)*, given for each year within a decade starting from first rise in frequency of occurrence, where the lexeme in question might have initially occurred in instances marked by textual distance markers (cf. Lemnitzer 2010: 69) or with later on missing morphosyntactic features. If available, encyclopedic information is linked to other online sources providing images or detailed explanations. The following sections introduce some of the features in the online edition of the NWB that provide dictionary users with additional lexicographic information, to clarify ambiguities regarding origin, usage or orthography of a loan word that might be associated with findings in the corpora or conceptual directives of the dictionary.

4.1 On origin and donor language

In the NWB, origin of a loan word is attributed to the direct donor language and further analyzed in the dictionary article section *Enzyklopädisches* (encyclopedic information). Words that have undergone further orthographic or phonetic change during one or potentially multiple transfers via their spreading through other languages, are analyzed in regard to (a) features contributed to a donor language (direct contact included borrowing), (b) alterations contributed to the adaptation into a recipient language and (c) the source word that might have served as a model for the borrowing (cf. Haspelmath 2009). Since newly borrowed words among neologisms in the dictionary of neologisms are considered to be fairly new at the time of their inclusion, where the borders between donor language and the actual source of a new word might blur, the dictionary does not aim at a singular explanation of a word's origin, but offers multiple lexicographic interpretations for users. Accordingly, the dictionary entry for *ploggen* with the meaning 'to gather up trash during a jog', that was included in the

NWB in 2019, offers two explanations for the word's potential origin in the section *Wortbildung* (word formation) in Fig. 1, with the word either having been derived in German from the loan word *Plogging* (noun describing the exercise that emerged earlier during the late 2010s) in analogy to the established English loan words *joggen* and *Jogging* or being borrowed directly from Swedish *plocka upp*.

The corresponding noun *Plogging* on the other hand, referring to the same recreational activity, spread globally through activities on the social media platform Instagram and was attributed to English as a donor language accordingly.

4.2 On grammar

In general, grammatical information given in the grammar entry of dictionary articles in the NWB comprises morphological and syntactical features of a word or multi-word expression. Borrowed nouns, for instance, have to be assigned a gender (fem., masc., neutrum) during their lexicalization process and the lexicographic information on gender is given in accordance with findings in the corpora sorted by the frequency of occurrences. Nouns with several genera and diverging declensions are listed for each gender respectively. Entries comprising up to three genera are complemented by optional, expandable commentary boxes. These boxes are marked by icons and contain further lexicographical description, e.g. details regarding genera that were confirmed by corpus data but not included in other common general dictionaries (illustrated by the first commentary in Fig. 2 with information on genitive singular declensions for the headword *Blog*) or examples from the corpora (second box in Fig. 2).

4.3 On spelling and pronunciation

Following orthographic assimilation rules, for borrowed compound nouns from English that consist of an adjective and a noun in the donor language both separated spelling (emphasis on both components) and conjoined spelling (emphasis on the first component of the compound noun) are considered for standard orthography. The word form of the headword in the lemma list of the NWB is the more common one of the two (cf. *Benutzerhinweise* regarding additional standard-conforming spelling, <https://www.owid.de/extras/neo/html-info/benutzerhinweise.html>). The entry for orthography and pronunciation in Fig. 3 illustrates an example of a borrowed lexeme with a standard-conforming spelling variation that depends on its respective intonation. *High Heel*, a neologism of the 1990s, is included as a headword in conjoined spelling. Additionally, a lexicographic commentary follows next to the additional standard-conforming spelling (*Weitere normgerechte Schreibung*) in a visually raised comment box, containing the lexicographical description and phonetic information pointing out emphasis on the first component.


By adding lexicographic commentaries next to general lexicographic information as illustrated in Figs. 2 and 3, ambiguities concerning dependencies between


Grammatische Angaben


Wortart: **Nomen**

Genus: **Maskulinum (der *Blog*)**

Deklination

Genitiv Singular: **Blogs, Blog** 

 Genitiv Singular *Blog* in den IDS-Textkorpora der Nullerjahre belegt, nicht jedoch in den Wörterbüchern der Nullerjahre gebucht

 Zahlreiche Blogger benutzen jetzt BlogVoices, die es den Lesern erlauben, die Einträge eines **Blog** zu kommentieren und sich mit den Autoren in Verbindung zu setzen. (Berliner Zeitung, 06.02.2001)

Nominativ Plural: **Blogs**

Genus: **Neutrum (das *Blog*)**

Deklination

Genitiv Singular: **Blogs, Blog**


Nominativ Plural: **Blogs** 

Fig. 2 Grammatical information in the dictionary entry for *Blog* in the Neologismenwörterbuch (<http://www.owid.de/artikel/316388?module=neo&pos=6>)

Schreibung und Aussprache

Weitere normgerechte

Schreibung: **Highheel**

 Die Zusammenschreibung *Highheel* ist normgerecht, wenn die Betonung nur auf der ersten Konstituente liegt: ['haihi:]

Nichtnormgerechte

Schreibung: **High heel**

Fig. 3 Information on orthography and pronunciation in the entry for *High Heel* in the Neologismwörterbuch (<https://www.owid.de/artikel/315741?module=neo&pos=5>)

orthography and pronunciation are resolved in a way that is beneficial towards the dictionary user, who can compare orthographic variations without having to consult the dictionary's user manual.

In contrast, the lexicographical description of ambiguities concerning the correspondence between graphematic and phonetic adaptation of a loan word seemingly differs regardless of the conceptual approach followed by different dictionary types. The qualitative case study presented in the following chapter explored to what degree graphematic features of the giving and the receiving language interfere with the stability of a loan word's lexicalization.

5 Orthography and pronunciation of loan words from different writing systems: a case study on *Qigong*

Orthographic and phonetic variations of loan words can have a considerable impact on lexicographical decisions, because the question as to when a new word can be considered to be an actual neologism is usually answered by the degree of its integration into the receiving languages writing system. Interestingly, only a few of the new loan words included as words borrowed from languages other than English originated from languages with writing systems that are not based on the Latin alphabet. Albeit some of these cases, such as *Hidschab*, *Namaste* or *Shawarma*, have spread through different languages and might not always be traced back to a single source word (cf. McConvell 2009), a majority of them are found to diffuse through the spread of cultural items or customs (cf. Haynie et al. 2014), whose distribution progresses much more rapidly in comparison to cases of older Wanderwörter like *Tee* ('tea') or *Hängematte* ('hammock') due to various types of off- and online interchange in today's highly interconnected global society. With state-of-the-art-tools for the compilation and analysis of large amounts of data at hand, lexicographers and researchers ought to aim to reconstruct a new word's way into the language system by investigating actual instances of the targeted language in use. For neologisms from the 1990s to 2010s, the NWB opts for the description of orthographic, phonetic and grammatical features that are consistent with the German language system and assigns the potential source word serving as a potential role model for its adaptation, which might not always relate to features of the same word in a donor language: The norm-conforming spelling variant of the German neologism *Hidschab* ('traditional covering for the head and neck that is worn by Muslim women'), for instance, differs from orthographic variants in other Latin-based languages that it might have been borrowed through (such as English: *hijab* or French: *hijab*, *hidjab*), which lean towards the transliteration of the Arabic source word *ḥiǧāb* (following the DIN-norm transliteration), but exhibits a higher degree of grapheme–phoneme-correspondence in German and correspondingly higher degree of lexicalization in the receiving language.

To explore correlations between the stability of a loan word's assimilation to a recipient language and the degree of grapheme–phoneme-correspondence of its transcription, an investigative case study was conducted on *Qigong* (氣功 in traditional Chinese, 气功 in simplified Chinese), a loan word in German that originated from Mandarin Chinese and was included in the NWB as a neologism that had

become part of Standard German in the 1990s.⁴ Romanization systems (e.g. (Hanyu) Pinyin, Wade-Giles or the ALA-LC Romanization rules⁵) used for the transcription of words from the Chinese logographic writing system into the Latin alphabet vary widely in regard to their graphematic representation of Chinese characters and phonemes—possibly influencing the lexicalization of loan words in recipient languages with different types of writing systems. As a loan word originating from a logographic (ideographic) writing system, *Qigong* was anticipated to remain orthographically inconsistent for a longer period of time following its first inclusion as a headword in a dictionary.

5.1 Method

Information on orthography and pronunciation given in dictionary entries for the lemma *Qigong* was compiled from nine general and specialized German dictionaries to compare changes pertaining to the orthographic and phonetic standardization of the loan word across time. The different types of dictionaries of German that were investigated for this case study to ensure the consideration of different lexicographical requirements and user needs are listed below.

- three dictionaries of foreign words: *Das große Fremdwörterbuch* (Duden), *Das Fremdwörterbuch* (Duden), and *Fremdwörterlexikon* (Wahrig),
- three dictionaries for German orthography: *Die deutsche Rechtschreibung* (Bertelsmann), *Die Deutsche Rechtschreibung* (Duden), *Die deutsche Rechtschreibung* (Wahrig),
- two general dictionaries of German: *Deutsches Wörterbuch* (Wahrig), *Deutsches Universalwörterbuch* (Duden).
- one specialized dictionary for German pronunciation: *Das Aussprachewörterbuch* (Duden).

Lexicographic information on orthography and pronunciation of the headword was compared over the course of the past two and the current decade, starting from the word's first inclusion during the 1990s and two following editions for the 2000s and 2010s respectively.

5.2 Results

Table 1 contains orthographic and phonetic information given for the lemma in at least one and up to three editions of a dictionary, sorted by the year of the lemma's first inclusion in the headword list of a dictionary. *Qigong* was first included in *Das*

⁴ Earlier occurrences of domain-specific usage in German newspapers in DeReKo can be dated back to late 1980s.

⁵ cf. ALA-LC Romanization Table for Chinese published by the Library of Congress for differences between ALA-LC, Pinyin and the Wade-Giles romanizations (<http://www.loc.gov/catdir/cpsa/romanization/chinese.pdf>, last viewed on 15.01.2019).

Table 1 Information on orthography and pronunciation of *Qigong* in general and specialized dictionaries of for German as of December 2019

Dictionary / dictionary type	Orthographic and phonetic standard given in the dictionary article by decade		
	1990s	2000s	2010s (until 2019)
Duden – <i>Das große Fremdwörterbuch</i> / (exhaustive) dictionary of foreign words	Qi gong / [tʃiˈkɔŋ] (1994)	Qi gong / [tʃiˈgɔŋ] (2000)	--- / [---]
Duden – <i>Das Fremdwörterbuch</i> / dictionary of foreign words	Qi gong / [tʃiˈgɔŋ] (1997)	Qi gong / [tʃiˈgɔŋ] (2007)	Qi gong / [tʃiˈgɔŋ] (2010)
Wahrig – <i>Fremdwörterlexikon</i> / lexicon of foreign words	Qi gong / [tʃiˈgɔŋ] (1999)	Qi gong / [tʃiˈgɔŋ] (2004)	Qi gong / [tʃiˈgɔŋ] (2011)
Bertelsmann – <i>Die deutsche Rechtschreibung</i> / dictionary of German orthography	Qi Gong / [dʒiˈgɔŋ] (1999)	--- / [---]	--- / [---]
Duden – <i>Die Deutsche Rechtschreibung</i> / dictionary of German orthography	--- / [---]	Qigong / [tʃiˈgɔŋ] (2000)	Qi gong / [tʃiˈgɔŋ] (2017)
Wahrig – <i>Deutsches Wörterbuch</i> / general dictionary of German	--- / [---]	Qi gong / [tʃiˈgɔŋ] (2000)	Qi gong / [tʃiˈgɔŋ] (2011)
Duden – <i>Das Aussprachewörterbuch</i> / dictionary for pronunciation	--- / [---]	Qigong / [tʃiˈgɔŋ] (2000)	Qigong / [tʃiˈgɔŋ, tʃiˈgɔŋ] (2015)
Wahrig – <i>Die deutsche Rechtschreibung</i> / dictionary for German orthography	--- / [---]	Qi gong / [dʒiˈgɔŋ] (2002)	Qi gong / [tʃiˈgɔŋ] (2011)
Duden – <i>Deutsches Universalwörterbuch</i> / general dictionary of German	--- / [---]	Qi gong / [tʃiˈgɔŋ] (2007)	Qi gong / [tʃiˈgɔŋ] (2015)

große Fremdwörterbuch (Duden, dictionary of foreign words) in 1994, a specialized dictionary for loan words. This first entry gave ‘Qilgong’ as the orthographic standard and [tʃiˈkɔŋ] as the words correct phonetic adaptation to German pronunciation. Across all dictionaries and dictionary types, revisions mainly concerned the phonetic representations of <g>, <i>, and <o> (here presented as the graphemes used for the adaptation into the German writing system).

Overall results present only two orthographic word forms (‘Qigong’ and ‘Qi Gong’), but 10 varying phonetic transcriptions. Whereas the orthographic standard of *Qigong* remains unaltered in all of the nine dictionaries listed in Table 1, the phonetic transcriptions of the first and the second syllable vary across both years and dictionaries. Three dictionaries have revised the phonetic transcription after the first inclusion of the lemma during the 2000s. In contrast, only three of the nine dictionaries revised information on standard pronunciation in the entries for *Qigong* between 2009 and 2019. Only one dictionary (Duden—*Das Aussprachewörterbuch*) presents a second alternative pronunciation of the second syllable. The general dictionary *Deutsches Universalwörterbuch* (Duden) was the last to include *Qigong* in its headword list in 2007 and did not alter the phonetic transcription in following editions from the 2010s until 2015.

Interestingly, only the Bertelsmann *Die deutsche Rechtschreibung* (dictionary of German orthography) from 1999 and its revised edition published through Wahrig (2002s edition of Wahrig—*Die deutsche Rechtschreibung* in Table 1), included the spelling ‘Qi Gong’ and the pronunciation standard [dʒi ɡʊŋ] in their respective first entries for *Qigong*. This orthographic adaptation, which resembled the original language’s two-character spelling, was not included in any following edition of other dictionaries presented above. The last major lexicographical revision of the dictionaries compiled for this case study was the inclusion of [tʃiˈɡɔŋ] as an alternate pronunciation standard in the 2015 edition of *Das Aussprachewörterbuch* (Duden, dictionary of pronunciation), that was not included in the 2017 edition of *Die deutsche Rechtschreibung* (Duden).

Further lexicographical explanation on the difference between spelling and pronunciation of the first syllable ‘Qi’ of *Qigong* was not included in the dictionary entries, even though official standards of orthography⁶ do not offer applicable rules for the assignment of phonetic representations⁷ of the grapheme <q> occurring as a single consonant as of today. Instead, <q> is assigned the phoneme [k], if it occurs as the grapheme combination <q> and <u> in a foreign word, e.g. in the case of the French Mannequin (['manəkɛ̃]). In case of native German words, however, <q> and <u> are assigned to the phonemes [k] and [v], e.g. Quitte (['kvitə]) or quellen (['kvɛlən]). There is no standard rule for the assignment of the phoneme order [tʃi] to the graphemes <q> and <u> (yet).

⁶ As of today, the official rules and word register of the council for German orthography (*Rat deutscher Rechtschreibung*) does not include rules concerning the phonetic transcription of <q>. (http://www.recht-schreibrat.com/DOX/rfdr_Regeln_2016_redigiert_2018.pdf, last revised in 2018).

⁷ In the case of a loan word, a foreign phoneme can be substituted for a German phoneme with phonetic similarity or for a German phoneme following the orthography of the assimilated lexeme (cf. Schippan 1992: 265).

Die deutsche Rechtschreibung (Duden, dictionary of German orthography) was the first to include the alternate standard-conforming spelling ‘Chi’ in the dictionary article for *Qi* (“vital energy”) in its 2009 edition, but did not alter the orthographic information in the entry of the headword *Qigong*. As of December 2019, entries in the dictionaries compiled in Table 1 do not include ‘Chigong’ as a standard-conforming orthographic variation of the lemma.

5.3 Conclusion

The results of the investigative case study on *Qigong* confirm the hypothesis that the degree of grapheme–phoneme correspondence correlates with the stability of a lemma’s integration into the German language system. Information regarding the standard-conforming orthography and pronunciation of the lemma *Qigong* varied over the examined timespan across all dictionary types and was last revised in the 2015’s edition of the dictionary for pronunciation issued by the Duden publishing company. The inclusion of *Qigong* without further orthographical assimilation of the syllable *Qi* corresponds to only one of the systems of Romanisation for Mandarin Chinese, Pinyin, with a lower degree of grapheme–phoneme correspondence than the Wade-Giles system that propose to transliterate the character 氣 (气) as ‘ch’i’. Albeit the fact that the transcription of the ideographic word into Latin alphabet was based on its phonetic features and not the “shape” of the character in the first place, the lexicographical description of the orthographic standard might have reinforced a low level of grapheme–phoneme-correspondence. In conclusion, future research needs to take into account the relation between grapheme–phoneme correspondence of transliterations from alternate writing systems and the degree of assimilation of loan words or word formation elements, as well as interferences of grapheme–phoneme-correspondence in the recipient language and the donor language caused by indirect borrowing processes.

6 Corpus-based development of lexicographic information in the online Neologismenwörterbuch

Today, tools for the corpus-based analysis of lemma candidates combined with benefits of electronic lexicography allow lexicographers to describe the integration of a borrowed lexeme as a process, i.e. how a word, word formation element or multi-word expression entered a language, adapted to the language system and became more frequent in use. The use of customized computer tools for the automated detection of neologisms like the *NeoCrawler* (cf. Kerremans and Prokic 2018), the *Logoscope* (cf. Falk et al. 2014) or *Neoloog* (cf. Falk 2014, Waszink 2019), and scientific corpus management systems such as Cosmas II and KorAP⁸ or the commercially

⁸ Cosmas II and its successor KorAP are two corpus linguistic analysis tools run on the Deutsche Referenzkorpus (DeReKo) and several historical corpora (<https://cosmas2.ids-mannheim.de/cosmas2-web/>; <https://korap.ids-mannheim.de/>).

Fig. 4 Cleared word form list for the search query q^*i++g^*ng ODER ($q^*i\ gong$) run on the corpus W1 of the DeReKo corpora, sorted by overall occurrence, number of texts and relative frequency per word form/type

23	21	0.175	QI GONG
7	7	0.053	QI Gong
5.673	4.039	43.206	Qi Gong
20	19	0.152	Qi-gong
6	5	0.046	qi gong
4	4	0.030	QI-GONG
2	2	0.015	QI-Gong
922	702	7.022	Qi-gong
21	17	0.160	Qi-gong
1	1	0.008	qi-gong
1	1	0.008	Qi-Gung
65	49	0.495	Qigang
2	1	0.015	Qigiong
24	23	0.183	QIGONG
346	264	2.635	QiGong
5.603	3.710	42.673	Qigong
25	12	0.190	qigong
1	1	0.008	qingong
4	1	0.030	Qoigung
198	176	1.508	Qui Gong
1	1	0.008	Qui gong
42	36	0.320	Qui-Gong
3	3	0.023	Quigang
5	4	0.038	QuiGong
115	95	0.876	Quigong

available Sketch Engine (cf. Kilgariff et al. 2014) has come to benefit lexicographers, dictionary makers and researchers alike, by providing methods for automated quantitative analyses of a new word, its grammatical features and collocational relations, which serve for more unbiased, objective lexicographical descriptions. The following sections use *Qigong* as an exemplary case to illustrate the corpus-based lexicographical assessment of word forms and their frequency that serves as a basis for the orthographic and phonetic information given for loan words in the Neologismenwörterbuch.

6.1 Corpus-based lexicographical assessment

The identification of candidates accounting for potential orthographic variations or alternative word forms is achieved with the help of *Cosmas II*, a corpus search, management and analysis database, which allows lexicographers to compile and filter computed word formation lists manually to elicit non-related word forms and formations in the resulting word form list. Occurrences of word forms on the result list are then analyzed regarding their lexical agreement with the lexeme in question, conformity with standardization rules of German and overall frequency in the corpus.⁹

For our example *Qigong*, a search query was modeled manually and run in corpus W1 (one out of 4 regularly updated text corpora in DeReKo). Non-related word

⁹ cf. Klosa and Lungen (2018) for further information on corpus-linguistics methods used for the detection and evaluation of neologism candidates.

Fig. 5 Orthography and pronunciation of *Qigong* in the NWB dictionary article (<https://www.owid.de/artikel/315716?module=neo&pos=1>)

Schreibung und Aussprache	
Weitere normgerechte	
Schreibung:	Qi-Gong
Nichtnormgerechte	
Schreibung:	Qi-gong
	QiGong
	Qi Gong
	Qi gong
	qigong
	Quigong
	Qui-Gong
	Qui-gong
	QuiGong
	Qui Gong
	Qui gong
Worttrennung:	Qi gong
Aussprache:	[tʃiˈɡoŋ, -ˈɡʊŋ]

forms were cleared off the initial word form list and the remaining orthographically close word forms were compiled to generate word form list *Qigong_W1* (Fig. 4). Frequently occurring word forms within corresponding context and with close graphematic overlap were taken into consideration for inclusion in the entry of the dictionary article of *Qigong*, as presented in Fig. 5.

6.2 Representation of corpus findings in the dictionary entry

The section *Schreibung und Aussprache* (orthography and pronunciation) presented in Fig. 5 contains the standard-conforming word form ‘Qigong’ (following official standards of German orthography), one or multiple alternate standard-conforming word forms (in black) and multiple not standard-conforming but frequent word forms (in gray). Further lexicographic information and references concerning the differentiation between standard and non-standard word forms are offered through the user manual of the dictionary (*Benutzerhinweise*), which is hyperlinked within each dictionary article.

Interestingly enough, the orthographic variation ‘Qi Gong’ that is included in both the NWB (as a not standard-conforming variation) and the 1999 edition of the Bertelsmann dictionary (*Die deutsche Rechtschreibung*) appears to be the most frequent word form (as measured by number of overall occurrences) in the W1-corpus, with its relative frequency being the second highest after the word form ‘Qigong’. Despite the absence of ‘Qi Gong’ as an orthographic variation of the lemma *Qigong* in entries of common and specialized dictionaries of the 2000s and 2010s, actual occurrences of the word form in the corpus W1 confirm its frequent use in written texts:

Zum *Qi Gong* gehören sanft bewegte Übungen ebenso wie solche, die ruhig stehend oder sitzend ausgeführt werden. (St. Galler Tagblatt, 08.01.2009)
 [‘Softly performed exercises are part of Qi Gong, as well as exercises that are performed in seated or standing positions.’]

Das Gong, gesprochen Gung, im Wort *Qi Gong* bedeutet passenderweise Arbeit. Qi Gong ist also die Arbeit, um die Lebensenergie fließen zu lassen. (Berliner Zeitung, 25.05.2019)
 [‘It is fitting, that the Gong, pronounced Gung, in the word Qi Gong means work. Qi Gong is the work that is performed to let the energy of life flow.’]

Evidence from the corpus that was compiled for this paper also confirms the lexicographical decision to include the not standard-conforming variation ‘Quigong’ (‘Qui-Gong’) as a prominent variation that remained frequent in use in German newspaper articles during the past two and the current decade:

die Kurse, die über die übliche Akupunkturausbildung weit hinausgingen und auch Spezialgebiete wie die Zungentherapie oder die Atemtherapie “*Quigong*” umfassen. (Nürnberger Nachrichten, 20.01.1995)
 [‘classes, that went far beyond the usual vocational training for acupuncture also included special subjects like tongue training or the breath therapy „Quigong“’]

viele moderne, stressgeplagte Menschen finden durch tägliches Üben von *Quigong* oder Tai Ji zum inneren Gleichgewicht zurück. (Rhein-Zeitung, 07.10.2004)
 [‘a lot of modern, stressed-out people find their way back to inner balance through daily practicing of Quigong or Tai Ji.’]

Donnerstags gibt es *Quigong*, eine chinesische[sic!] Heilgymnastik (Mannheimer Morgen, 27.07.2018)
 [‘Quigong, a Chinese therapeutic gymnastic, is offered on Thursdays’]

Despite its higher degree of grapheme–phoneme correspondence, the spelling variant *Chigong* was not included in the dictionary entry for *Qigong* presented above (Fig. 4). The exclusion corresponds to the type’s corpus-based frequency, which serves as one of several criteria for inclusion of a word form. An additional search query modeled for *Chigong* yielded 55 occurrences in 49 subcorpora for Germany, Austria and German newspapers in Swiss and Tirol in corpus W1.

Pronunciation (*Aussprache*) presented in Fig. 5 is based on the manual assessment of online audio and video data and follows the International Phonetic Alphabet (IPA). For *Qigong*, the NWB included both [tʃiˈgʊŋ] and [tʃiˈgʊŋ], corresponding to the standard pronunciation given in *Das Aussprachewörterbuch* in 2015, the dictionary for German pronunciation published by the Duden.

6.3 Benefits and issues

As shown above, tools for the corpus-based detection and analysis of neologisms combined with benefits of electronic lexicography allow lexicographers to include a detailed representation of the current level of integration of a loan word by describing examples of actual occurrences of the word. However, recent studies on quantitative methods for computed analyses of large corpora (cf. Müller-Spitzer et al. 2018; Kopleinig 2016) have brought up justified concerns related to the equation of a word's significance in a language and its frequency in corpora compiled from easily and quickly available texts, i.e. for a large part newspaper articles. Assuming that journalists and publishers of newspapers aim to adhere to the official standards of orthography, corpus-based lexicographic information on patterns of use for a word can only represent a sample of its actual usage and needs to be described accordingly. In line with general concerns on the representative nature of compiled corpora, the designation of an official standard, i.e. by the council for German orthography, needs to be considered as a major influence on the integration process of new words.

7 Outlook

Internet lexicography offers more and more adequate tools and features to describe language in online dictionaries through efficiently interconnected infrastructures within the dictionary or by making use of further information available on the internet (cf. Müller-Spitzer 2018: 321–324). But do dictionary users need lexicographic information on non-standard conforming variations of a word or do we overwhelm them by adding orthographic information that differs from their (print) dictionary-using habits? To give dictionary users the most conclusive lexicographical description of a loan word and its integration into the language system, lexicographers need to take a closer look at natural language usage of a word in the recipient language (e.g. by exploring data compiled from the web). Since English accounts as a transfer language for a majority of newly borrowed lexemes in German, previous adaptations to the English language system (serving as a tool for pre-assimilation to a Latin-based writing system) might account for a lower grapheme–phoneme-correspondence of the word's adaptation to German. Further research ought to investigate (a) potential interferences between the orthographic and phonetic adaptation of a word caused by borrowing through a first or second transfer language and (b) differences between Romanization systems of non-Latin writing systems, to assess potential standard-conforming adaptations of the word in question.

Acknowledgement Open Access funding provided by Projekt DEAL.

Compliance with ethical standards

Conflict of interest The author declares that there is no conflict of interest regarding the publication of

this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Busse, Ulrich. 2004. *Toll Collect* ['toi..., engl,'toul... oder'to:l...]—ein Fall fürs Tollhaus oder den Duden?. Standard und Variation bei der Aussprache von Anglizismen im Deutschen. *Jahrbuch des Instituts für Deutsche Sprache*. <https://doi.org/10.1515/9783110193985.207>.
- Dargiewicz, Anna. 2013. Fremde Elemente in Wortbildungen des Deutschen: zu Hybridbildungen in der deutschen Gegenwartssprache am Beispiel einer raumbundenen Untersuchung in der Universitäts- und Hansestadt Greifswald. *Schriften zur diachronen und synchronen Linguistik* 15. Frankfurt am Main/Bern/Bruxelles/New York/NY/Oxford/Wien: Lang.
- DeReKo—Leibniz-Institut für Deutsche Sprache. 2019. *Deutsches Referenzkorpus/Archiv der Korpora geschriebener Gegenwartssprache 2019-11*. Mannheim: Leibniz-Institut für Deutsche Sprache. Online: <http://www.ids-mannheim.de/DeReKo>.
- Duden—*Das Aussprachewörterbuch*. 2000. 4th revised edition, ed. Max Mangold, Dudenredaktion. Mannheim/Leipzig/Wien/Zurich: Dudenverlag.
- Duden—*Das Aussprachewörterbuch*. 2015. 7th revised edition, ed. Stefan Kleiner, Ralf Knöbl and Dudenredaktion. Berlin/Mannheim: Dudenverlag/Institut für Deutsche Sprache.
- Duden—*Das Fremdwörterbuch*. 1997. 6th revised edition, ed. by the scientific council of the Dudenredaktion. Mannheim/Leipzig/Wien/Zurich: Dudenverlag.
- Duden—*Das Fremdwörterbuch: auf der Grundlage der aktuellen amtlichen Rechtschreibregeln*. 2010. 10th revised edition, ed. Dudenredaktion. Mannheim/Zürich: Dudenverlag.
- Duden—*Das Fremdwörterbuch: auf der Grundlage der neuen amtlichen Rechtschreibregeln*. 2007. 9th revised edition, ed. by the scientific council of the Dudenredaktion. Mannheim/Leipzig/Wien/Zurich: Dudenverlag.
- Duden—*Das große Fremdwörterbuch: Herkunft und Bedeutung der Fremdwörter*. 1994. Ed. by the scientific council of the Dudenredaktion. Mannheim/Leipzig/Wien/Zurich: Dudenverlag.
- Duden—*Das große Fremdwörterbuch: Herkunft und Bedeutung der Fremdwörter*. 2000. 2nd revised edition, ed. by the scientific council of the Dudenredaktion. Mannheim/Leipzig/Wien/Zurich: Dudenverlag.
- Duden—*Deutsches Universalwörterbuch*. 2007. 6th revised edition, ed. Ralf Osterwinter, Dudenredaktion. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- Duden—*Deutsches Universalwörterbuch*. 2015. 8th revised edition, ed. Werner Scholze-Stubenrecht, Dudenredaktion. Berlin: Dudenverlag.
- Duden—*Die deutsche Rechtschreibung: auf der Grundlage der aktuellen amtlichen Rechtschreibregeln*. 2017. 27th revised edition, ed. Kathrin Kunkel-Razum, Dudenredaktion. Berlin: Dudenverlag.
- Duden—*Die deutsche Rechtschreibung: auf der Grundlage der neuen amtlichen Rechtschreibregeln*. 2000. 22nd revised edition, ed. Werner Scholze-Stubenrecht, Dudenredaktion. Mannheim/Leipzig/Wien/Zurich: Dudenverlag.
- Falk, Ingrid, Bernhard, Delphine, Gérard, Christophe. 2014. From non word to new word: automatically identifying neologisms in French newspapers. *LREC—The 9th edition of the Language Resources and Evaluation Conference, May 2014, Reykjavik, Iceland*.

- Grundlagen der deutschen Rechtschreibung*. 2004–2018. Zwischenstaatliche Kommission für deutsche Rechtschreibung/Rat für deutsche Rechtschreibung. Online: <https://www.rechtschreibrat.com/regeln-und-woerterverzeichnis/>.
- Haynie, Hannah, Claire Bown, Patience Epps, Jane Hill, and Patrick McConvell. 2014. Wanderwörter in languages of the Americas and Australia. *Ampersand* 1: 1–18.
- Haspelmath, Martin. 2009. Lexical borrowing: Concepts and issues. In *Loanwords in the World's languages: A Comparative Handbook*, ed. Martin Haspelmath and Uri Tadmor, 35–54. Berlin: de Gruyter Mouton.
- Herberg, Dieter. 2004. Das Projekt “Neologismen der 90er Jahre des 20. Jahrhunderts“. In *Sprachkultur und Lexikographie. Von der Forschung zur Nutzung von Wörterbüchern*, ed. Jürgen Scharnhorst, 331–353. Frankfurt am Main/Berlin/Bern/Bruxelles/New York/Oxford/Wien: Lang.
- Herberg, Dieter, Kinne, Michael, Steffens, Doris. 2004. Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. *Schriften des Instituts für Deutsche Sprache* 11. Berlin: de Gruyter.
- Hermann, Ursula. 1999. *Die deutsche Rechtschreibung*. Nachschlagewerke zur deutschen Sprache im Bertelsmann-Lexikon-Verlag. Revised edition, ed. Lutz Götze. Gütersloh/München: Bertelsmann-Lexikon-Verlag.
- Hermann, Ursula. 2002. *Wahrig, Die deutsche Rechtschreibung*. 3rd revised edition, ed. Lutz Götze. Gütersloh/München: Wissen-Media-Verlag.
- Horx, Matthias. 1995. *Trendwörter von Acid bis Zippies: Lexikon*. Düsseldorf: ECON.
- Kerremans, Daphné, and Jelena Prokić. 2018. Mining the web for new words: semi-automatic neologism identification with the NeoCrawler. *Anglia* 136 (2): 239–268.
- Kilgariff, Adam, et al. 2014. The Sketch Engine: ten years on. *Lexicography* 1: 7–36.
- Kirkness, Alan. 1976. Zur Lexikologie und Lexikographie des Fremdworts. Probleme der Lexikologie und Lexicographie. In *Probleme der Lexikologie und Lexikographie. Jahrbuch des Instituts für Deutsche Sprache* 1975, ed. Hugo Moser, 226–241. Düsseldorf: Schwann.
- Klosa, Annette and Lünen, Harald. 2018. New German words: detection and description. IDS Publikationen. https://ids-pub.bsz-bw.de/frontdoor/deliver/index/docId/7718/file/Klosa_Luengen_New_german_words.pdf. Accessed 20.11.2019.
- Klosa-Kückelhaus, Annette. 2019. From *chatten* through *podcasten* to *youtuben*. Social media neologisms from the 1990s to the 2010s in German. *Neologica* 13: 107–123.
- Koplenig, Alexander. 2016. *Analyzing lexical change in diachronic corpora*. Dissertation. Mannheim: Universität Mannheim.
- Krome, Sabine. 2011. *Brockhaus, Wahrig, Die deutsche Rechtschreibung*. 8th edition, ed. Sabine Krome, Wahrig-Redaktion. Gütersloh: Wissenmedia in der Inmedia-ONE-GmbH.
- Lemnitzer, Lothar. 2010. Neologismenlexikographie und das Internet. *Lexicographica* 26: 65–78.
- Library of Congress. 2011. *ALA-LC Romanization Table for Chinese*. Online: <https://www.loc.gov/catdir/cpsd/romanization/chinese.pdf>, last viewed 15.01.2019.
- McConvell, Patrick. 2009. Loanwords in Gurindji, a Pama-Nyungan language of Australia. In *Loanwords in the World's languages: A Comparative Handbook*, ed. Martin Haspelmath and Uri Tadmor, 790–822. Berlin: de Gruyter Mouton.
- Müller-Spitzer, Carolin. 2018. Wörterbuchbenutzungsforschung. In *Internetlexikografie. Ein Kompendium*, ed. Annette Klosa-Kückelhaus, Carolin Müller Spitzer, 291–342. Berlin/Boston: de Gruyter.
- Müller-Spitzer, Carolin, Wolfer, Sascha and Koplenig, Alexander. 2018. Quantitative Analyse lexikalischer Daten. Methodenreflexion am Beispiel von Wandel und Sequenzialität. In *Wortschätze. Dynamik, Muster, Komplexität. Jahrbuch des Instituts für Deutsche Sprache* 2017, ed. Stefan Engelberg et al., 245–266. Berlin/Boston: de Gruyter.
- Neologismenwörterbuch*. 2006—today. In OWID—Online Wortschatz-Informationssystem Deutsch. Mannheim: Institut für Deutsche Sprache. Online: <http://www.owid.de/wb/neo/start.html>, last viewed 24.01.2020.
- Onysko, Alexander. 2007. *Anglicisms in German. Borrowing, lexical productivity, and written codeswitching. Linguistik—Impulse und Tendenzen* 23. Berlin, New York: de Gruyter.
- OWID—Online Wortschatz-Informationssystem Deutsch. Mannheim: Institut für Deutsche Sprache. Online: <http://www.owid.de/wb/neo/start.html>, last viewed 24.01.2020.
- Quasthoff, Uwe. 2007. *Deutsches Neologismenwörterbuch. Neue Wörter und Wortbedeutungen in der Gegenwartssprache*. Berlin: de Gruyter.
- Schippan, Thea. 1992. *Lexikologie der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

- Schröder, Martin. 1997. Brauchen wir ein neues Wörterbuchkartell? Zu den Perspektiven einer computerunterstützten Dialektlexikographie und eines Projekts „Deutsches Dialektwörterbuch“. *Zeitschrift für Dialektologie und Linguistik* 64: 57–66.
- Steffens, Doris, al-Wadi, Doris. 2015. Neuer Wortschatz. Neologismen im Deutschen 2001–2010. 2 Volumes. Mannheim: Institut für Deutsche Sprache.
- Wahrig, Gerhard. 1999. *Fremdwörterlexikon*. Edited and revised by Renate Wahrig-Burfeind. München: Dt. Taschenbuchverlag.
- Wahrig, Gerhard. 2000. *Deutsches Wörterbuch*. 7th revised edition, ed. Renate Wahrig-Burfeind. Gütersloh: Bertelsmann-Lexikon-Verlag.
- Wahrig-Burfeind, Renate. 2004. *Wahrig-Fremdwörterlexikon*. 5th revised edition, edited and revised by Renate Wahrig-Burfeind. Gütersloh/München: Wissen-Media-Verlag.
- Wahrig-Burfeind, Renate. 2011. *Brockhaus—Wahrig Fremdwörterlexikon*. 8th revised edition, ed. Renate Wahrig-Burfeind. Gütersloh: Wissenmedia.
- Wahrig-Burfeind, Renate. 2011. *Brockhaus, Wahrig, Deutsches Wörterbuch: mit einem Lexikon der Sprachlehre*. 9th revised edition, ed. Renate Wahrig-Burfeind, Sabine Krome. Gütersloh/München: Wissenmedia.
- Waszink, Vivien. 2019. *Neologisms in an Online Portal*. Online: https://globalex.link/wp-content/uploads/2019/05/gwln2019_waszink.pdf, last viewed 13 Apr 2020.
- Wortwarte. 2000–today. *Die Wortwarte. Wörter von heute und morgen. Eine Sammlung von Neologismen*, ed. Lothar Lemnitzer. Online: <http://www.wortwarte.de>.
- Yang, Wenliang. 1990. *Anglizismen im Deutschen: am Beispiel des Nachrichtenmagazins Der Spiegel*. Tübingen: Niemeyer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.