Steven Coats*

# Lexicon *geupdated*: New German anglicisms in a social media corpus

https://doi.org/10.1515/eujal-2019-0001

**Abstract:** The German verbal lexicon has been enriched by numerous English borrowings, particularly within the past 100 years, but while many verbal anglicisms are frequently used and sanctioned by language authorities, the status of new, non-standard, and rare verbal anglicisms in German has not been subject to extensive research attention. In this study, a new method is used to analyze non-standard German verbal anglicisms in a large and novel corpus compiled from the social media platform Twitter. After a review of previous work, the methods used to create a corpus of German-language tweets and to automatically extract new verbal anglicisms are described, and the semantics of some of their most frequent types are analyzed, including forms with separable and inseparable prefixes. Then, present and past participles are considered according to assimilation to standard German orthography, use as participle or attributive adjective, and stem vowel quality. In the final set of results, the focus is on the productivity of the verbalizing morpheme *-ier-*, a historically important element for the integration of foreign word material into German. The study demonstrates that non-standard verbal anglicisms are widely used, and that their morphological behavior is mediated by frequency effects as well as phonological, pragmatic, and semantic considerations.

**Keywords:** Corpus linguistics, German, Anglicisms, Morphology, Social Media, Twitter

**Zusammenfassung:** Die Anzahl der Verbalanglizismen im Deutschen ist in den letzten 100 Jahren kontinuierlich gestiegen. Viele Verbalanglizismen sind inzwischen fest verankert im deutschen Sprachgebrauch, aber der Verwendungsumfang und das morphologische, syntaktische, und semantische Verhalten von neuen und seltenen Verbalanglizismen ist noch nicht intensiv erforscht worden. In dieser Studie werden neue Verbalanglizismen anhand eines von der Sozialmedienplattform Twitter kompilierten Korpus analysiert. Nach einer Skizze der Forschungslage werden die Methoden der Korpuskompilation und der automatischen Generierung von neuen Anglizismen vorgestellt. Eine Häufigkeitsanalyse der häufigsten Verbalanglizismen und deren Bedeutungen erfolgt, sowie eine

***Corresponding author: Steven Coats,** University of Oulu, English Philology, 90014 Oulu, Finland, e-mail: steven.coats@oulu.fi

Analyse von Verbalbildugen mit trennbaren und nicht-trennbaren Präfixen. Bei Partizipien wird der Zusammenhang zwischen orthographischer Normierung, grammatischer Funktion, und phonemischer-graphemischer Übereinstimmung aufgezeigt. Im darauffolgenden Abschnitt wird der Status des alten Verbalisiermorphems *-ier-* berücksichtigt. Zusammenfassend kann eine reiche Vielfalt an neuen Verbalanglizismen festgelegt werden: ihre Orthographie und Beugung hängen von Worthäufigkeit ab, wobei lautliche, pragmatische, und semantische Faktoren eine Rolle spielen.

**Schlüsselwörter:** Korpuslinguistik, Deutsch, Anglizismen, Morphologie/Wortbeugung, Sozialmedien

**Resumen:** El léxico verbal alemán se ha enriquecido con numerosos préstamos en inglés, particularmente en los últimos 100 años, pero mientras que muchos anglicismos verbales son frecuentemente utilizados y sancionados por las autoridades lingüísticas, otros son menos utilizados. En este estudio, se analizan nuevos y no estándar anglicismos verbales alemanes en un amplio y novedoso corpus compilado desde la plataforma de medios sociales Twitter. Después de una revisión de trabajos anteriores, se describen los métodos utilizados para crear un corpus de tweets en alemán y para extraer automáticamente nuevos anglicismos verbales. Entonces, la prevalencia de los anglicismos verbales es analizada: los participios presentes y pasados son considerados de acuerdo con la asimilación a la ortografía alemana estándar, el uso como participio o adjetivo atributivo, la clase semántica, la longitud de la palabra y la calidad de las vocales del tallo. En la tercera sección, se analizan las frecuencias y el comportamiento morfológico de las formas verbales prefijadas, y en la cuarta sección, el enfoque se centra en la productividad del morfema verbalizador *-ier-*, un elemento históricamente importante para la integración de material de palabras extranjeras en alemán. El estudio demuestra que los anglicismos verbales no estándar son ampliamente utilizados, y que su comportamiento morfológico está mediado por efectos de frecuencia, así como por consideraciones fonológicas, pragmáticas y semánticas.

**Palabras clave:** Lingüística de corpus, Alemán, Anglicismos, Morfología, Medios sociales

# 1 Introduction

Borrowing, a universal linguistic phenomenon, is one of the processes by which the lexicon undergoes change. In recent centuries, and particularly since 1945, the German lexicon has been a receptor of borrowings from English – primarily of noun forms, but also of verbs, which due to their inflectional richness and variable constituency status within the phrase, are well-suited to shed light not only upon processes of lexical and semantic change, but also of changes in the way the grammar and syntax of a language typically pattern. The semantics of borrowings and of loan words (borrowings that have become an accepted part of the lexicon) are typically transparent: non-finite verbal anglicisms in German have meanings that are (usually) straightforward and predictable according to their source language semantics, and their morphological behavior (mostly) conforms to German grammatical paradigms. The process by which loan words such as anglicisms are morphologically and orthographically modified and become integrated into the standard language of a lexicon, however, has not been well documented in corpus-based studies. Because anglicisms are prevalent in German, non-finite verbal anglicisms are suitable targets for a corpus-based and quantitative analysis of lexical and morphological change.[1]

Borrowings and loan words exist on a continuum: the former can be defined as lexical items from a source language that exist as lone elements within a sentence or utterance in a receptor language, while the latter are 'nativized' borrowings that show some degree of integration into the morphosyntax and/or orthography of a receptor language (see Kachru 1990: 59–60; Myers-Scotton 1997). In this study, the focus is on a particular class of non-finite verbal anglicisms: New words that exhibit some degree of orthographic integration in the German lexicon, but which are not yet widely attested or codified in sources such as dictionaries. Most new anglicisms (and other loan words) lead a fleeting existence: coined as neologisms in particular situations, they fail to become established in the lexicon of a receptor language and to gain acceptance within the broader speech community. Nevertheless, because rare and new anglicisms exhibit variability in morphological assimilation to standard paradigms and because they are often coined in order to fill a semasiological gap in a receptor language, they are well-suited for the study of certain processes of semantic and morphological change, especially if their prevalence and variability can be documented and quantified using a corpus-based approach.

**1** This paper consists of a revised and expanded version of Coats (2018b), with some modifications to the methodology.

This study introduces two main methodological improvements to the study of anglicisms in German. First, the corpus created for the study, sampled from the social media platform Twitter, is larger than the corpora used in most recent studies of anglicisms in Germany. Because larger corpora are more likely to record rare usages than are smaller corpora (Biber 1993), this study can identify some patterns in the usage of verbal anglicisms that have not previously been noted in the research literature. Second, rather than relying on existing lists of anglicisms, such as anglicisms whose status as German words has been well-established or anglicisms from a particular lexical field, this study introduces a method for the automatic generation and identification of new anglicisms in German by creating a large number of possible German word forms from English-language verbal stems. The principal advantage of this approach is that it dispenses with preordained analytical categories, allowing the relevant lexical types to 'emerge from the corpus' in terms of their frequencies. Instead of analyzing a list of types of interest based on researcher intuition, which may or may not correspond to the inherent variability in anglicisms in the data, the approach used in this study identifies the most frequent types from hundreds of thousands of potential word forms; the frequency information then guides the ensuing analysis of semantic, morphological, and orthographic parameters of variation on the basis of the attested word forms (cf. the multidimensional approach to grammatical features pioneered by Biber 1988).

The text of the study is organized as follows: After a review of some previous literature, the methods used to create the corpus and to generate the lists of potential new German anglicisms are described and the morphological behavior of non-finite verbs in German briefly reviewed. The 'Results and Analysis' section is organized in four parts: First, the extent to which non-standard verbal anglicisms are present in the corpus is quantified and some of the most frequent types analyzed in terms of their semantics. The second part explores the productivity of verbal prefixation of new verbal anglicisms, both by separable and inseparable prefixes, and considers the semantics of prefixed forms on the basis of type frequencies. The third part considers the degree to which participial forms assimilate to standard German orthography and postulates a relationship between type frequency and orthographic assimilation, modulated by factors such as status as verbal or noun phrase element or stem vowel quality. In the fourth part of the results, the status of the *-ier-* morpheme as a verbalizer of English lexical material is analyzed on the basis of corpus frequencies.

In the ensuing discussion, the possible semantic, phonological, and pragmatic factors that affect the behavior of new verbal anglicisms are addressed, and an outline for future work on anglicisms is provided.

## 2 Previous work

English has been an important source language for borrowings for at least 100 years, especially for European receptor languages (Görlach 2001). In German, the number of anglicisms began to increase in the 19th century (Eisenberg 2011; 2013), but especially since 1945, a result of social, economic, and political factors (von Polenz 1999). Research into anglicisms in German has focused on the semantics, grammar, and contexts of use of anglicisms. Several studies have made use of corpora to quantify the prevalence of anglicisms and document their increasing use over time.

An early corpus-based study of English borrowings in German is Carstensen (1965), in which the semantics, grammar, and syntax of anglicisms were analyzed on the basis of texts from West German newspapers and magazines printed from 1961–1964, primarily the weekly news magazine *Der Spiegel*. Carstensen distinguished between anglicisms that denote new concepts (*Bedürfnislehnwörter*, 'necessary borrowings'), for which no lexeme exists in the receptor language, and anglicisms that denote concepts that are also denoted by existing lexemes (*Luxuslehnwörter*, 'luxury borrowings').[2] Galinsky (1967) categorized anglicisms according to their stylistic or pragmatic functions. Yang (1990), also analyzing a corpus of texts from *Der Spiegel*, attested an increase in the relative frequency of anglicisms over time, as did Onysko in a more recent corpus-based analysis of anglicisms in *Der Spiegel* (2007). Burmasowa (2010) came to similar conclusions regarding the increase in anglicisms over time in an analysis of texts from the daily newspaper *Die Zeit*. Anglicisms in German were cataloged in a three-volume dictionary (Carstensen and Busse 1993, 1994, 1996).

The semasiological distinction proposed by Carstensen and earlier researchers was revisited by Onysko and Winter-Froemel (2011), who introduced terms from classical rhetoric, *catachrestic* and *non-catachrestic*, to classify anglicism types in German. Drawing on Levinson's theory of conversational implicature (Levinson 2000), the authors proposed that the terms distinguish between pragmatic implicatures of 'informativeness', which are typically unmarked and communicate stereotypical information, versus implicatures of 'manner', which are marked and "license inferences towards a non-stereotypical interpretation" (2011:

**2** The distinction was noted by earlier researchers. For example, according to Hermann Paul, "*zur aufnahme fremder wörter in der muttersprache veranlasst natürlich zuerst das bedürfnis. Es werden demgemäss wörter für begriffe aufgenommen, für welche es dieser noch an einer bezeichnung fehlt.*" [Naturally, the first motivation for the adoption of foreign words into the mother tongue is necessity. Accordingly, words will be adopted for concepts for which the mother tongue lacks a designation.] (1886: 339).

1555). As with earlier classifications of borrowings, however, the distinction catachrestic/non-catachrestic is also based on the 'newness' of the concept a borrowing denotes: "the fundamental linguistic criterion to distinguish between these two types of innovations is tied to the question of whether the concept designated by the new expression (the innovation) is already expressed by another lexical unit in the language or not" (2011: 1554), suggesting that conceptually, the new terms are not substantially different from the earlier terms *Bedürfnis-/Luxuslehnwort*. The authors then listed the 101 most frequent anglicisms from Onysko's *Spiegel* corpus and rated them as catachrestic or non-catachrestic; non-catachrestic borrowings (e. g. *Job, Manager, Trend*) were found to constitute approximately two-thirds of this group.

A corpus-based chronological account of anglicisms in the 20th and early 21st centuries is Eisenberg (2013), who analyzed their frequency and use in a corpus compiled from popular, scientific, journalistic, and literary texts published in the periods 1905–1914, 1948–1957, and 1995–2004. Eisenberg noted that the proportion of anglicisms that are verbal forms has remained relatively constant over time at approximately 5 % of all anglicisms. The relative paucity of verbal anglicisms is likely due in part to the additional morphosyntactic constraints on verbal forms as compared to noun phrase constituents (2013: 79). The most common verbal anglicism types have changed over the course of a century: in the period 1905–1914 they were *beordern* 'to order (s. o.)', *boykottieren* 'to boycott', *pokern* 'to play poker, to wager', *streiken* 'to go on strike', *tippen* 'to type', *interviewen* 'to interview', *hinbeordern* 'to order (s. o.) somwhere', *starten* 'to start', *trainieren* 'to exercise/practice (sports)", *boxen* 'to box', *lynchen* 'to lynch', *flirten* 'to flirt', *toasten* 'to toast', *chartern* 'to charter', *bluffen* 'to bluff', and *paddeln* 'to paddle'; in 1995–2004 they were *starten*, *stoppen* 'to stop', *testen* 'to test', *trainieren*, *parken* 'to park', *schockieren* 'to shock', *surfen* 'to surf', *flirten*, *interviewen*, *klicken* 'to click', *boomen* 'to boom/increase rapidly', *kicken* 'to play football (soccer)", *ordern* 'to order', *managen* 'to manage', *boykottieren*, *schocken* 'to shock', *checken* 'to check', *pokern*, *driften* 'to drift', and *outen* 'to out' (84; 93). The evolution of the most frequent types reflects some societal changes (disappearance of *lynchen*, appearance of *outen*) or technological developments that motivate 'necessary' loans (appearance of *surfen* and *clicken*). Two verbs with the same meaning derived from 'to shock' are attested among the most recent frequent forms in 1995–2004: *schocken* and *schockieren*. The latter of these two represents a somewhat older derivational process – it incorporates the *-ier-* morpheme, long used to introduce foreign lexical material into the German verbal system (see Section 4.4). Anglicisms as past or present participles are rather rare in the older subsection corpus used by Eisenberg (2013), but more frequent in contemporary German. Notably absent from all sections of the corpus are prefixed verbal anglicisms (2013: 114).

Winter-Froemel et al. (2014) analyzed the frequency of 50 anglicisms (47 nouns and 3 adjectives) and their German lexical equivalents or near-equivalents in a corpus compiled from different sources. They regressed anglicism frequency with several variables, and found that for words that replicate the semantic content of existing lexemes (*non-catachrestic* borrowings), older anglicisms, shorter anglicisms, and anglicisms from the lexical field of information technology are more likely to be used.

Baeskow (2017) discussed verbal anglicisms with the inseparable prefixes *er-*, *be-*, *ent-*, *ver-*, and *zer-*, focusing on verbs derived from corporate names in the field of information technology (e. g. *ergoogeln*, *vertwittern*). She noted that such verbs can be found by searching the web using Google's search engine, as well as in some online corpora of German, and that the semantics of inseparable prefixes, particularly as they pertain to the grammatical category of lexical aspect, tend to be preserved in the attested prefixed verbal anglicisms.

Although the concepts of *Bedürfnislehnwörter* and *Luxuslehnwörter* (and the related terms *catachrestic* and *non-catachrestic*) provide a convenient starting point for a semasiological analysis of new anglicisms, in some cases, the assignment of an anglicism to one of the two categories is not straightforward. Onysko and Winter-Froemel (2011) and Winter-Froemel et al. (2014) point out that some anglicisms derived from polysemous English words are themselves polysemous, with one meaning corresponding to a new concept, but another equivalent to an existing German lexeme.[3] Furthermore, for some anglicisms, a German-language equivalent may be attested but extremely rare, or morphologically possible but not attested, additionally complicating the decision as to which anglicisms are to be categorized as 'necessary' and 'luxury' loans. Even for words for which an equivalent or near-equivalent exists in German, the semantic and collocational profiles (the 'semantic prosody'; Sinclair 1991) of an anglicism and a German lexical type with the same meaning will not be completely equivalent. The decision as to the necessity of a borrowing may ultimately be based on judgments as to the relative frequency of rare words, the grammaticality of new word formations, or the meanings of polysemous items. Manual classification of the meanings of anglicisms and disambiguation of polysemous lexemes can be useful for the analysis of existing anglicisms, but can be time-consuming for large numbers of types, and may not be reliable for new wordforms. While beyond the scope of this study, using word embeddings to explore the meanings of anglicism types in large corpora may also be a useful approach (cf. Coats 2018a).

**3** The example *surfen* is given: The meaning 'to surf online' is new, but the meaning 'to surf waves in the ocean' is covered by *wellenreiten*.

Before the methods used to collect the data and generate the new non-finite verbal anglicisms are discussed, a brief overview of the morphology of non-finite forms of the German verb and verb derivational processes is provided.

## 2.1 Non-finite verbal forms

Infinitives of German verbs end in *-(e)n*. The present participle (*Partizip I*) is formed via affixation of *-d* to the infinitive. Verbs borrowed into German or derived from borrowed lexical material typically assimilate to the weak inflectional paradigm, forming the past participle (*Partizip II*) via circumfixation of *ge-* and *-(e)t*.

In Table 1, *sagen* and *lernen* belong to the core German lexicon, whereas *mailen* and *batteln* are anglicisms. In *batteln*, metathesis of <*le*> has occurred in order to adhere to the German norm for phoneme-grapheme correspondence, and the schwa of the infinitive suffix *-en* has been elided after a liquid. Verbs formed from English words with the same phonological shape (e. g. *rattle*, *bubble*, etc.) are usually subject to this process and their orthography is adapted (Duden 2016: § 38, § 92–94; Eisenberg 2011: 242–244), although for recent borrowings, variation exists (e. g. *googeln* and *googlen* 'to google'). The participles of the forms in Table 1 are formed according to the standard pattern for German weak verbs.

**Table 1:** Infinitive, *Partizip I*, and *Partizip II* of weak verbs in German.

| Infinitive | Present participle | Past participle |
|---|---|---|
| *sagen* 'to say' | *sagend* 'saying' | *gesagt* 'said' |
| *lernen* 'to learn' | *lernend* 'learning' | *gelernt* 'learned' |
| *mailen* 'to email' | *mailend* 'mailing' | *gemailt* 'emailed' |
| *batteln* 'to battle' (esp. rap battles) | *battelnd* 'battling' | *gebattelt* 'battled' |

Verbal borrowings in German typically assimilate to the German orthographic paradigm for the formation of the past participle, but for many verbs, variants exist that show different degrees of assimilation. In examples (1) and (2) from the corpus, the past participle of the anglicism *liken* ('to like', esp. social media) exhibits full assimilation to the paradigmatic norm for weak verb past participles (*gelikt*, in the first example) or partial assimilation (*geliked*), in which the English *-ed* ending is retained. In the first example, the text notes that the Duden publish-

ing house, an important German language authority, officially adopted the fully assimilated form in its dictionary in 2017.[4]

(1) *@user Jetzt ist es offiziell: du hast gelikt, er/sie/es likt. #Duden* [Now it's official: you have liked, he/she/it liked. #Duden]

(2) *@user Grade erst gesehen :3 Das meist geliked Video auf mein Kanal mittlerweile, Dankeschön!!!* [Just saw it :3 The most liked video in my channel in the meantime, Thankyou!!!][5]

Some previous research has considered variation in the degree of assimilation of the past participle to German weak-verb norms. Onysko suggested that the word class of an anglicism may influence its orthography: words borrowed as inflectable verbs may be more likely to assimilate to standard German orthographic norms when building derived forms (e. g. *canceln → gecancelt, chatten → gechattet*), whereas forms borrowed as adjectival participles from English may retain their English spelling (e. g. *relaxed*), especially if their phonological realization in English and German more or less coincide (2007: 235–237).

## 2.2 Verb derivation via affixation

Prefixation of a verbal stem with a separable or an inseparable particle is a productive process in German verb formation. The separable prefixes *ab-*, *an-*, *auf-*, *aus-*, *durch-*, *ein-*, *her-*, *herauf-*, *herein-*, *herum-*, *herunter-*, *hin-*, *hinzu-*, *los-*, *mit-*, *nach-*,*vor-*, *voran-*, *voraus-*, *weg-*, *zu-*, *zurück-*, and *zusammen-*, most of which are prepositions or adverbs, can delimit or specify the scope of the verbal stem, often spatially or temporally, as in *kommen* 'to come' → *hereinkommen, mitkommen* 'to come in', 'to come with, accompany', or *gehen* 'to go' → *vorausgehen* 'to go before, precede'. Inseparable prefixes, historically derived from prepositions and adverbs, have been grammaticalized as prefixes that can express a wide range of possible meanings (see Duden, § 1054–1076). Often, inseparable verbal prefixes express aspect or *Aktionsart* (Wischer and Habermann 2004), as in *denken* 'to think' (imperfective) → *erdenken* 'to devise, conceive of' (perfective). In Section 4.2, the frequencies of prefixed new verbal anglicisms are analyzed in terms of the lexical fields they are drawn from.

**4** Usernames and URLs in the examples in this study have been anonymized.

**5** The past participle *geliked* is used as an attributive adjective in this example; according to standard German grammar it should take the inflectional ending *-e*. Partial retention of the English orthography may inhibit such a form's integration into the standard German inflectional paradigm. See Section 4.3.2.

The morphemes *-ier-* and *-isier-*, in verbs such as *faszinieren* ('to fascinate') or *legalisieren* ('to legalize'), have historically been the most important morphemes for the integration of foreign lexical material (usually Romance in origin) into the German verbal system, productive since at least the 12th century (Öhmann 1970). Historically, relatively few verbs have been formed in German via fusion of a borrowed English verbal stem with the *-ier-* morpheme. Non-standard *-ier-* formations, however, are prevalent in the corpus used in the present study. In some cases, *-ier-* forms are attested in addition to forms showing simple suffixation of *-en* (e. g. *attackieren* vs. *attacken*, both 'to attack'). The use of the *-ier-* derivation may have pragmatic or stylistic significance; the question is addressed in Section 4.4.

# 3 Data and methods

The corpus used for the study consists of 36,240,530 tweets in German (534,211,366 tokens) collected from users located in Germany, Austria, or Switzerland. The corpus was created in several steps. First, 653,457,659 tweets with 'place' metadata were collected globally from the Twitter Streaming API from November 2016 until June 2017 using *Tweepy* (Roesslein 2015). From this 'seed' data, users with a 'place' attribute in Germany, Austria or Switzerland who had broadcast at least one German-language tweet were identified (70,986 users). All tweets of these users, or the most recent 3,250 (the upper limit for tweet downloads for a single user from Twitter's REST API), were downloaded in April 2018. Of the 70,986 users of interest, the timelines of 60,683 were available – the others having been presumably set to private, deleted by the user, or banned by Twitter. 61,118,733 tweets were downloaded from these 60,683 users. Tweet metadata was used to select only those written in German (59.3 %) for the final corpus. Tokenization was undertaken using the nltk tokenizer (Bird et al. 2009).[6]

New verbal anglicisms were created automatically from lists of English verbs. The 1,000 most frequent base verbal forms (English infinitives without *to*) were compiled from the *British National Corpus*, the *Corpus of Contemporary American English*, and the *Wikipedia Corpus of English* (Davies 2004–, 2008–, 2015)[7] and combined with 1,413 forms from the *Pattern Dictionary of English Verbs* (Hanks

**6** A list of the tweet IDs for the tweets in the corpus is available at https://github.com/stcoats/Germananglicisms; the corpus can be generated by downloading the tweets from Twitter's API.
**7** http://corpus.byu.edu.

2013).[8] This aggregated list was used to create 2,739 unique German infinitives, including some variant spellings (e. g. *googeln* and *googlen*). Then, present and past participles, as well as prefixed verbal forms, were created using string substitutions based on regular expressions, taking into account German phonotactics and orthographic conventions. The inseparable prefixes (*be-*, *er-*, *ent-*, *emp-*, *miss-*, *ver-*, *zer-*, *über-*) and the separable prefixes (*ab-*, *an-*, *auf-*, *aus-*, *durch-*, *ein-*, *her-*, *herauf-*, *herum-*, *herunter-*, *hin-*, *hinzu-*, *mit-*, *voran-*, *los-*, *mit-*, *vor-*, *weg-*, *zurück-*, *zusammen-*) were used; in addition, deverbal forms with an infixed *-zu-* (e. g. *anzutwittern*) were generated. Finally, the same collection of verbal forms were automatically generated from verbal stems via suffixation of *-ier-* and *-isier-*. The searches analyzed in Section 4.3.2 also took into account present and past participles with adjectival inflection (e. g. *ein geliktes Foto, das gelikte Foto* 'a liked photo', 'the liked photo').

Because the focus of this study is primarily on new verbal anglicisms that are not yet part of the standard German lexicon or in widespread use, it was necessary to remove actual German word forms from the list of anglicisms. This was done using a list of 238,650 unique German word types (Kleuker 2016) aggregated from corpora maintained by German CLARIN centers: the Berlin-Brandenburg Academy of Sciences, the Leipzig Corpora Collection of the University of Leipzig, and the Institute for the German Language in Mannheim. To catch word forms not attested in Kleuker (2016) but which can be considered standard German words, the morphological analyzer SMOR was used (Schmid et al. 2004, Fitschen 2004).

Additionally, a step was taken to remove automatically-generated anglicisms that are actual English words: Because some tweets in the corpus consist of code-switched passages of German and English text, automatically-generated anglicisms whose orthography corresponds to an existing English word may be false positives. For example, a form such as *driven* was generated automatically as a German infinitive from English 'to drive'. In code-switched tweets, however, the form appears as the English past participle in an English-language passage. Types such as this were removed by matching the generated anglicisms with an English word list of 236,736 types from nltk (Bird et al. 2009). Finally, inflected participial forms were generated by suffixing *-e*, *-em*, *-en*, *-er*, *-es*, *-este*, *-estem*, *-esten*, *-ester*, and *-estes* to the present and past participles. Only words not attested in standard German and not orthographically equivalent to standard English words are considered in the analysis below.[9]

**8** http://pdev.org.uk.

**9** An aggregated list of the automatically-generated infinitives and past participles without adjectival inflection is available at https://github.com/stcoats/Germananglicisms.

The procedure used for the creation of potential anglicisms generated a very large number of potential German word forms. For example, from the English verb *to fail*, ten 'base' forms were generated (*failen, failend, gefailt, gefailed, failieren, failierend, failiert, failisieren, failisiernd, failisiert*), seven of which are participles, and for each base form, 28 prefixed forms (*abfailen, anfailen, auffailen, ausfailen*, etc.). Inflecting the 196 participial forms for each English word stem with ten adjectival suffixes results in 1,960 variants, meaning that for a single new verbal anglicism, well over 2,000 word forms were automatically generated. While most of the automatically-generated forms were not attested in the corpus, the procedure used to generate non-finite forms allowed for rare usages to be discovered. Overall, new anglicisms exhibit significant orthographic and morphological variation in the corpus.

# 4 Results and analysis

## 4.1 Overall frequencies

New non-finite verbal anglicisms in the corpus are attested from diverse semantic fields and exhibit variation in orthography. A total of 3,228 types in the corpus produced matches with the automatically-generated list, comprising 124,615 tokens, or 233.27 per million words. To compare, Onysko (2007) reported that 1.11 % of tokens (11,000 per million words) in his corpus from *Der Spiegel* consisted of anglicisms (2007: 114). The difference is not surprising, as Onysko considered established anglicism types that are widely used in German, whereas this study considers novel and rare anglicism types that are not widely used. Nevertheless, to cursorily assess the degree to which the Twitter corpus used in this study also contains established anglicisms, a list of German anglicisms[10] was processed to only take into account individual lexemes (i. e. not multi-word anglicisms), including hyphenated anglicisms; the frequencies of the 4,778 types were then tabulated in the corpus. Using this method, 8,007,698 'old' anglicisms were detected (not including hashtags), a rate of 1.50 % (14,990 per million words). The comparability of this result with that of Onysko (2007) suggests that, in line with other research, the overall rate of anglicism use in German continues to increase. It should be noted, however, that genre differences between a corpus of news articles and a corpus of tweets likely also play a role; because established anglicisms are not the focus of this study, this question has not been further pursued. As for the new

**10** https://de.wiktionary.org/wiki/Verzeichnis:Deutsch/Anglizismen.

**Table 2:** Most frequent new anglicisms.

| | Type | Freq | | Type | Freq |
|---|---|---|---|---|---|
| 1 | *twittern* | 28921 | 11 | *adden* | 1214 |
| 2 | *streamen* | 9248 | 12 | *geupdated* | 1188 |
| 3 | *getwittert* | 6567 | 13 | *haten* | 1146 |
| 4 | *liken* | 5928 | 14 | *rendern* | 1054 |
| 5 | *googlen* | 2829* | 15 | *coden* | 1000 |
| 6 | *gestreamt* | 2232 | 16 | *followen* | 831 |
| 7 | *geliked* | 1415 | 17 | *gevotet* | 810 |
| 8 | *supporten* | 1370 | 18 | *cachen* | 782 |
| 9 | *gefixt* | 1300 | 19 | *tracken* | 781 |
| 10 | *geflasht* | 1271 | 20 | *sharen* | 758 |

*6388 if *googeln*, whose stem is in the SMOR lexicon, is included.

anglicisms, some of the types in the corpus are relatively frequent: Table 2 shows the 20 most frequent types.

Among the most frequently attested anglicism types in the corpus, many are 'necessary loans' for new concepts or activities in the sphere of information technology with specific semantic profiles that do not correspond exactly to existing German lexemes: *twittern*, *liken*, *adden*, *updaten*, *followen*, and *sharen* (and their participles) are words used in the context of communication on social media platforms. *Coden*, *googlen*, *rendern*, and *streamen* are also used in the context of internet/IT. The meanings of these words correspond closely to the English-language meanings of the items from which they have been borrowed. *Gefixt* is used in the corpus primarily in the sense of 'to repair/fix' (an online service or website). Only three of the 20 most frequent types in the corpus denote activities in non-IT contexts: *supporten* 'to support' denotes support for a cause, an idea, or a team, as in (3), (4), and (5) (equivalent to *unterstützen*). *Geflasht* is used as a predicate adjective meaning 'excited' or 'enthusiastic', as in (6), but also to denote rewriting the memory of an IT device.[11] *Haten*, in (7) and (8), is a stylistically marked equivalent to standard German *hassen* 'to hate'.

(3) *@user wenn es denn dem Schutz des Heidegrundwassers dient, sollen doch die Reicheren unter sich mächtig dafür supporten.* [@user if it is to protect the groundwater of the moor, the richer of them should support it strongly amongst themselves.]

**11** With the meaning 'excited' or 'enthusiastic', the participle *geflasht* qualifies as an derivational pseudo-anglicism, or a borrowing from English whose meaning is not attested for the source-language lexeme from which it has been borrowed.

(4) *@user Wir suchen immernoch einen Sponsor der sich traut unser Projekt zu supporten, da hilft also jeder Kontakt der euch einfällt!* :-) [@user We are still looking for a sponsor who dares to support our project, any contact that occurs to you helps! :-)]

(5) *so kinder, jetzte jehts los. kurz vorm olympiastadion. supporten fuer hertha und die relegation. alle die daumen druecken!!!* [so children, now it begins. just in front of Olympic Stadium. supporting hertha and relegation. everyone cross your fingers!!!]

(6) *@user das war auch einfach geil!!! ich bin immer noch total geflasht!!!* [@user it was simply great!!! I'm still totally excited!!!]

(7) *Ich bin ja ganz vorne mit dabei wenns darum geht den #EmojiFilm zu haten... aber den Trailer find ich gar nicht mal so scheiße.* 😶 [I'm among the first to agree when it comes to hating the #EmojiFilm... but the trailer is not even so shitty. 😶]

(8) *Wie kann man eigentlich #RTL für solch ein geiles Programm haten? Das ist pure Unterhaltung egal wie schlecht gemacht, immer gewesen!* [How can you actually hate #RTL for such a great program? It has always been pure entertainment, no matter how poorly made!]

While it is not surprising that new anglicisms are introduced to denote items or activities originating in the (mostly English-language) world of information technology, a large number of anglicisms in the corpus bear no relation to IT. The frequency distribution for the attested forms exhibits a classic Zipfian profile and has a 'long tail': a large number of types that occur only once in the corpus (*hapax legomena*). The 1,345 *hapax* types are drawn from diverse lexical fields, mostly unrelated to information technology. A sample – the meanings of which are transparent from the verbal stem – is shown in (9).

(9) *annoyen*, *breathen*, *ercapturen*, *zurückcheaten*, *gehealed*, *mitgementioned*, *gelookt*, *gepeeltes*, *killiert*, *encouragierend*, *failiert*

None of the forms in (9) could be considered 'necessary' borrowings, as the underlying concepts could potentially be expressed with existing lexemes such as *nerven*, *atmen*, *fangen*, *zurückschummeln/zurückbetrügen*, *geheilt*, *miterwähnt*, *geschaut*, *geschältes*, *getötet*, *ermutigend*, or *versagt*. The form *ercapturen* is used in a tweet that makes reference to a geocaching game in which players are instructed (in German) to *capturen* hidden items; the use with the inseparable *er-* prefix may reflect the author's intention to express perfective aspect. This interpretation is in line with the finding of Baeskow (2017) that inseparable pre-

fixes are used with new verbal anglicisms to delimit aspectual features of the verb.

Present participles, a relatively infrequent word class, were few in the corpus: the most common type was the necessary borrowing *twitternd*, but 'luxury' types such as *giggelnd*, *entertainend*, *snackend*, or *goend* were also attested.

## 4.2 Prefixed forms

A large number of types with separable and inseparable prefixes are present in the corpus. Of the 807 types with separable prefixes, the most frequent are *angefixed* 'be fixated on, addicted', *abfucken* 'to fuck up',[12] *antwittern* (and participle forms) 'to twitter to (s. o.)', *abchillen* 'to chill out', *durchliken* 'to like through' (i. e. to 'like' all the posts on a page or all the images in a social media album), and *angepingt* 'to be pinged on a telecommunications or social media network'. Most of these forms pertain to internet or IT-related activities and could thus be considered necessary borrowings. In some cases, such as *durchliken*, the semantics of German verbal prefixation make it possible to more succinctly express a meaning than would be possible with the original English lexeme – 'to like all of the posts' is a lengthy circumlocution. The forms *angefixed*, *abfucken*, and *abchillen* all denote meanings for which there are German equivalents or near-equivalents (e. g. *süchtig nach etwas sein*, *verwahrlosen/herunterkommen/vermasseln*, *entspannen*). The choice to use one of the anglicism forms, however, can convey additional stylistic or pragmatic information.

For inseparable prefixes, 166 types were attested: *vertwittern* 'to twitter away (one's time, e. g.)', *entfollowen* 'to stop following on social media', and *entliken* 'to stop liking on social media', were the most common, followed by *ertwittern* 'to twitter (sth., perfective meaning)', *vertrumpen*, and *vertellen*. Again, most forms denote internet and social-media-related activities. The form *vertrumpen* combines the negative-evaluative connotation of the *ver-* prefix with the name 'Trump' to cast aspersion on the American president. *Vertellen* is an interesting form in that some uses are clearly anglicisms, as in (10), while others are low German, as in (11).

(10) *Ruft mich einer an, um mir seine Lebensgeschichte zu vertellen. Diggah, ich bin Pannenhilfe, nicht Domian. #RiasArbeit* [Someone calls me and

**12** The past participle form, *abgefuckt*, is not counted because as an established and frequent anglicism, it is attested in the word lists used to filter out existing German lexemes.

tells me his life story. Dude I'm the breakdown service, not Domian. #RiasArbeit][13]

(11) *Hinrich un Florian vertellen eis wedder wat up #Plattdüütsch! Dissemol is Tina Landgraf vun dat Mecklenburgisches Staatstheater (Fritz-Reuter-Bühn) doarbi! Sonnawend 03.03.2018: 13 Uhr – LOHRO! #watupplatt #plattdeutsch #fritzreuterbühne #mecklenburgischesstaatstheater* [Hinrich and Florian tell us something in #Plattdüütsch again! This time Tina Landgraf from the Mecklenburg State Theater (Fritz-Reuter-Stage) is present! Saturday 03.03.2018: 1 PM – LOHRO! #watupplatt #plattdeutsch #fritzreuterbühne #mecklenburgischesstaatstheater]

The prefixed infinitive form with infixed *-zu-* was attested by 70 types: *abzufucken*, *mitzutwittern*, and *anzutwittern* were the most frequent. For both the separable and the inseparable prefixed verbal forms, a large number of types are attested only once. These include forms such as *heruntertwittern* 'to twitter down to' and *weggeliked* 'liked away' that relate to internet activities, but also a great many 'luxury' borrowings such as *mitgejammt*, *abwriten*, *beinfluencen*, *ershooten*, *erfreshen*, or *bepainten*, all of whose meanings seem to be the same as those of the English stems from which they are derived.

The method used to automatically generate new anglicisms resulted in some false positives. Several of the *hapax legomena* and other low-frequency types are not anglicisms, but rather non-standard spellings of standard German words. This can be confirmed by examining the tweets in which the forms occur. For example, the form *erfaren*, automatically generated in the anglicism list by combining 'to fare' with the inseparable prefix *er-*, is a non-standard spelling of standard German *erfahren* ('to experience' or 'experienced'). Similar forms include *überagend* (from *to age*), a misspelling of *überragend* 'outstanding', *forden* and *erforden* (from *to ford*), misspellings of *fordern* 'demand' and *erfordern* 'require', *gestatet* (from *to state*), an orthographical variant of *gestattet* 'allowed', *ausgerut* (*to rut*), meaning *ausgeruht*, 'rested', and *verwanten* (*to want*), denoting *verwandten* 'related/relations'. In total, these false positives are so infrequent in the corpus (90 tokens) that their presence doesn't greatly affect the overall frequency of new non-verbal anglicisms; hence, the corresponding base verbal forms were not removed from the list used to generate potential anglicism types.

**13** 'Domian' was a radio and television show in the public media of North Rhine-Westphalia in which moderator Jürgen Domian answered telephone calls from listeners/viewers.

## 4.3 Variation in the past participle

Variation between the assimilated form of the past participle, with the ending *-(e)t*, and the partially-assimilated form, with the ending *-ed*, was found for 219 past participle types. In Table 3, the counts of the variant forms are shown alongside the logarithm of the odds ratio, a convenient effect size metric,[14] for forms for which both variants are attested at least once and at least one of the variants is a new anglicism. Figure 1 shows the log odds ratio versus the log of number of occurrences of both variants for all 219 attested base forms (to enhance legibility, not all values are annotated with their word forms). Points with negative log odds values have the *-(e)t* ending; those with positive values end in *-ed*. The distribution is limited by the function $y = -\log x$ and its reflection accross the y-axis. The moderately strong negative correlation between the log of the total number of occurrences versus the log of the odds ratio (Spearman's $\rho = -0.39$, $p = 3.2 \times 10^{-12}$) suggests that as a new verbal anglicism becomes more widely used, it tends to assimilate to standard orthography, taking on the *-(e)t* ending for the past participle.

**Table 3:** Variation in Past Participles.

| | Type | Freq | Type | Freq | logOR |
|---|---|---|---|---|---|
| 1 | *geblogged* | 100 | *gebloggt* | 8840 | −4.48 |
| 2 | *getwittered* | 4 | *getwittert* | 6567 | −7.40 |
| 3 | *geblocked* | 29 | *geblockt* | 5862 | −5.31 |
| 4 | *gechecked* | 55 | *gecheckt* | 3111 | −4.04 |
| 5 | *gerocked* | 5 | *gerockt* | 2433 | −6.19 |
| 6 | *gegoogled* | 127 | *gegoogelt* | 2276 | −2.89 |
| 7 | *gestreamed* | 121 | *gestreamt* | 2232 | −2.91 |
| 8 | *geliked* | 1415 | *gelikt* | 197 | 1.97 |
| 9 | *geupdated* | 1188 | *geupdatet* | 404 | 1.08 |
| 10 | *geflashed* | 309 | *geflasht* | 1271 | −1.41 |

Among the most frequent 20 types, the partially-assimilated forms ending in *-ed* are more common for the types *geliked* and *geupdated*. Among the less frequent types, *gefeatured*, *geshared*, *geclimbed*, and *gemaked* are preferred to variants with standard orthography. Overall, however, variants with assimilated inflection are more frequent in the corpus.

**14** The measure, $\log \frac{n_x}{n_y}$, is symmetrical about zero and results in positive values when $x$ is more frequent and negative values when $y$ is more frequent.

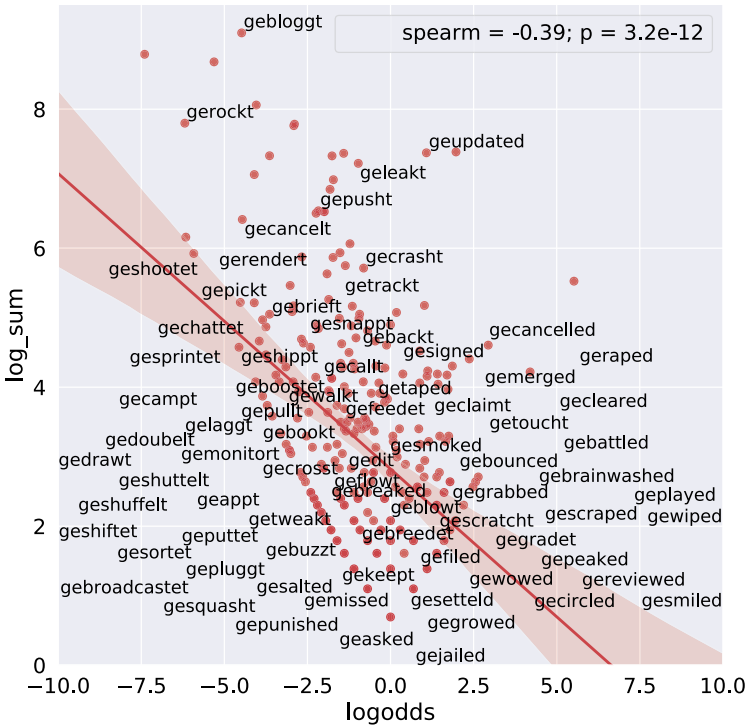


**Figure 1:** Assimilated and partially-assimilated past participle forms.

One factor that may influence the degree of orthographic assimilation is anglicism age: Older borrowings, such as *gerockt*, are more likely to exhibit standard inflection than are newer borrowings such as *geliked* or *geupdated* (see also Winter-Froemel et al. 2014).

The degree to which English and German orthography overlap in the representation of vowel sounds may also influence assimilation to German inflection, with assimilation less likely for forms that would have a different stem vowel realization according to standard German orthographic-phonemic correspondence conventions, such as *gelikt* or *geshart*. In such cases, retention of partially-English orthography may help recognition of the diphthongs [aɪ] and [eɪ]. Many of the forms more readily assimilated to German participial inflection (those with negative log-odds ratio values) have stem vowels similar in quality to those of the original English verbs or participles.

### 4.3.1 Stem vowel

While older anglicisms in German typically underwent changes in orthography in order to conform to standard German grapheme-phoneme correspondence, such as the 19th-century anglicisms *Streik*, 'strike' (planned work stoppage by industrial laborers) or *Dschungel* 'jungle' (cf. Görlach 2003: 72 ff.), most anglicisms introduced since 1945 retain their English orthography. However, a lack of correspondence between the orthographic and phonological shape of a word may cause the word to be less frequently used or affect its morphological productivity. Poplack and Sankoff (1984), investigating code-switching among Spanish-English bilinguals in New York City, found an association between frequency and degree of phonological assimilation for borrowings (in this case, English lexical items used in Spanish spoken discourse). Winter-Froemel et al. (2014) included phoneme-grapheme 'markedness' as a factor in a multiple regression of loanword frequency, apparently by manually annotating each of the 50 words in their analysis as to whether it contains graphemes or grapheme sequences whose phonetic realization in English and in German are not equivalent.

In order to test whether stem vowel quality is associated with morphological assimilation in words where the graphemes *<i, a, o>* correspond to the diphthongs [aɪ], [eɪ], and [oʊ] in the original English word forms (and hence also in anglicisms such as *liken*, *updaten, shapen*, or *followen*), past participle anglicisms with intraconsonantal *<i, a, o>* in the verbal stem and with a frequency of 10 occurrences or more were identified in the corpus using regular expressions and classified as diphthong or non-diphthong. Counts for diphthong types and all other new anglicism past participle types were then compared in a contingency table (Table 4) according to assimilation (*-(e)t*) or partial assimilation (*-ed*) using a log-likelihood test (Dunning 1993).

**Table 4:** Stem vowel quality and past participle assimilation.

| | Diphthong types | Other types |
|---|---|---|
| *-(e)t* | 2297 | 63097 |
| *-ed* | 3601 | 5521 |

The test statistic and associated p-value ($G = 9097.35$, $p < 1.00 \times 10^{-320}$) suggest that at least for this subsample of the corpus, stem vowel quality may interact with morphological assimilation, hindering it when there is a diphthong in the verbal stem. Thus, forms with the dipthongs [aɪ] (*liken*), [eɪ] (*updaten, shapen*),

and [oʊ] (*followen*) may be more likely to retain the English participial ending due to phonemic-graphemic incongruence between English and German. In the long term, however, it seems that these forms (or at least those of them that enter into widespread use) will also gradually be replaced by the variant exhibiting the standard German inflection, as in (1).

### 4.3.2 Grammatical function of participle

German past participles (mostly those of transitive verbs) can be used as attributive and superlative adjectives, but partial assimilation may affect the degree to which forms such as *geupdated* are used as attributive adjectives (e. g. *die geupdatete Webseite* 'the updated website' vs. *?die geupdatede Webseite*). Because according to German grammar, the uninflected participle only occurs as a verbal constituent, adjectival use of the past participle can be investigated by simply counting occurences of particples with the inflectional suffixes *-e*, *-em*, *-en*, *-er*, *-es*, *-este*, *-estem*, *-esten*, *-ester*, and *-estes* and comparing their frequencies with those of the uninflected forms. Table 5 shows the ten most frequent fully assimilated past participles, their frequencies as participial or adjectival attributes, and the verbal to adjectival log odds ratio. While the tendency to be used as a verbal component or an adjectival attribute depends on the semantics of the verb, verbal use is more common — the verbal to adjectival log odds ratio for all fully-assimilated participles is 2.93, meaning the forms are almost 19 times more likely to be used as verbal elements.[15]

**Table 5:** Use as VP element or Attributive Adjective: Assimilated Anglicisms.

| | Type | Freq_part | Freq_adj | logOR |
|---|---|---|---|---|
| 1 | *gebloggt* | 8840 | 67 | 4.88 |
| 2 | *getwittert* | 6567 | 209 | 3.45 |
| 3 | *geblockt* | 5862 | 172 | 3.53 |
| 4 | *gecheckt* | 3111 | 7 | 6.10 |
| 5 | *gerockt* | 2433 | 2 | 7.10 |
| 6 | *gegoogelt* | 2276 | 28 | 4.40 |
| 7 | *gestreamt* | 2232 | 49 | 3.82 |
| 8 | *gechillt* | 1487 | 377 | 1.37 |
| 9 | *geleakt* | 993 | 411 | 0.88 |
| 10 | *gefixt* | 1300 | 20 | 4.17 |

**15** In this data, the ratio is similar for other, non-anglicism past participles.

**Table 6:** Use as VP element or Attributive Adjective: Partially-assimilated Anglicisms.

| | Type | Freq_part | Freq_adj | logOR |
|---|---|---|---|---|
| 1 | *geliked* | 1415 | 3 | 6.16 |
| 2 | *geupdated* | 1188 | 0 | inf |
| 3 | *geleaked* | 375 | 4 | 4.54 |
| 4 | *geflashed* | 309 | 0 | inf |
| 5 | *gefeatured* | 250 | 0 | inf |
| 6 | *gefixed* | 223 | 0 | inf |
| 7 | *gehacked* | 197 | 0 | inf |
| 8 | *getagged* | 164 | 0 | inf |
| 9 | *gevoted* | 131 | 0 | inf |
| 10 | *gefollowed* | 130 | 1 | 4.87 |

The partially-assimilated participles, on the other hand, are much less likely to be used as attributive or superlative adjectives (Table 6), with a significantly greater log odds ratio for participial-adjectival use of 5.40.

This result suggests that assimilation to standard German morphology and orthography may be required before a borrowed lexical item can be used in the full range of grammatical and syntactic slots possible for the word class.

## 4.4 *-ier-* derivations

234 types created via derivation with *-ier-* or *-isier-* were attested in the corpus (including infinitives, participles, and prefixed forms). The most frequent attested type was *makiert* (140 occurrences) and its infinitive *makieren* (114 occurrences), which are not new, non-standard anglicisms, but rather orthographical variants of *markieren* 'to mark/tag' (e. g. an image on a social media platform). The relatively high frequency of these forms suggests that they are not the result of keyboarding errors committed by persons typing quickly – the non-standard spellings discussed at the end of Section 4.2, in contrast, occur only once or a handful of times in the corpus. The type may represent a phonetic spelling of a relatively uncommon standard German word, corresponding to an approximate pronunciation of the word in parts of Western Germany.[16] Other frequent types with *-ier-* or *-isier-* morphemes are not typically used in standard German in Germany, but are established dialect words (e. g. the Swiss German words *grillieren*

**16** See König (1989: 74 ff.) on *R-Schwund* in spoken German, cf. Elspaß and Möller (2009): http://www.atlas-alltagssprache.de/runde-1/f16a-b/

'to grill/barbecue' or *parkieren* 'to park a car') or, like *makieren*, are non-standard spellings of established lexical items, such as *abbonieren* instead of *abonnieren* 'to subscribe to (e. g. a newspaper)', *boycottieren* instead of *boykottieren* 'to boycott', or *debatieren* instead of *debattieren* 'to debate', and thus do not represent new anglicisms. Genuine new anglicisms with *-ier-* or *-isier-* include *trumpieren* (7 occurrences) and *trumpisieren*, referencing the American president but used with various meanings. *-ier*-derived forms of the most common new verbal anglicisms, those pertaining to social media and IT, are infrequent: *twitterieren* occurs once in the corpus, as does *updatieren*. A form with the denominal verbalizer *-isier-* occurs with *twitterisieren*, attested three times.

A few *-ier-* lexemes appear to be new borrowings from English: *relatieren* 'to relate (to something)' occurs 15 times, all with the modal collocate *können* 'to be able', as in (12) and (13):

(12) *@user Kann relatieren, hatte den gleichen Mist vor nem halben Jahr.* [@user I can relate, had the same crap half a year ago.]

(13) *@user baller das hier mal und sag mir ob du damit auch relatieren kannst URL* [@user take a shot at this and tell me if you can also relate to it URL]

*Relatieren* is one of the types for which an *-ier*-derived form is in competition with a form created via simple affixation of the infinitive suffix *-(e)n*: the form *relaten* occurs 58 times in the corpus, as in (14) and (15):

(14) *@user Hat ja nichts damit zu tun dass es nicht lustig ist, ich kann nur nicht relaten* [@user doesn't have anything to do with the fact that it's not funny, I simply can't relate]

(15) *Meine liebste Variable ist die Dummy-Variable, da kann ich relaten. URL* [My favorite variable is the dummy-variable, I can relate to that. URL]

In its participial form, however, *relatiert* is often used as an attributive adjective meaning 'relevant/similar/related', as in (16) and (17).

(16) *Wenn man sich da mal die relatierten Videos ansieht – türkische Popmusik ist schon irgendwie ne Parallelwelt. :o* [If one takes a look at the related videos – Turkish pop music is somehow a parallel universe. :o]

(17) *@user /soc/-relatierte sachen sind so kompliziert URL* [@user /soc/-related things are so complicated URL][17]

**17** /soc/ refers to a message board at 4chan.org often used to facilitate types of social interaction such as cam sessions or physical meetings. 4chan may also have played a role in the borrowing

Verb formation from borrowed lexical items via the *-ier-* morpheme, although still somewhat productive in German, appears to be less common than suffixation of a borrowed stem with the *-en* infinitive suffix. Word length considerations and communicative economy may also play a role, especially considering the character limitation inherent to Twitter.

## 5 Conclusions and future outlook

This study attests a large number of new verbal anglicisms in a German-language social media corpus from Twitter. The significance of the study is fourfold: First, the study introduces a methodological innovation for anglicism research – generating potential anglicism types automatically from large lists of genuine English words via regular expressions – that better accounts for variation in lexical innovations, compared to methods that consider only a small number of pre-defined types. The method will prove to be useful not only for the corpus-based study of anglicisms in German, but for the study of anglicisms in other languages as well. Second, and in line with findings from recent research on anglicisms, the study shows that many of the most frequent new anglicisms are 'necessary borrowings' – they denote activities or technologies from the lexical field of communication or information technology, especially social media. However, there are a very large number of infrequent types in the corpus, most of which are not drawn from the lexical field of information technology, suggesting that the creation of new anglicisms is a constant low-level phenomenon in some genres and registers of contemporary German. Third, variation in the degree to which the past participle (*Partizip I*) of new anglicisms assimilates to standard German inflectional morphology, by taking on a *-(e)t* ending, is mediated by frequency effects, and influenced by phoneme-grapheme correspondence in the stem vowel: types with stem vowel diphthongs whose orthography does not correspond to the standard German representation for these sounds are slower to assimilate. In addition, types showing partial assimilation are less likely to be used as attributive adjectives. Finally, the study shows that the verbalizing morphemes *-ier-* and *-isier* are still productive in German, but relatively poorly attested, and forms derived from these morphemes

of this meaning of *relatiert*, which may be derived by analogy from a common discourse framing device used on the site's message boards: A user will upload an image and accompanying text; a particular aspect of the post (or the entire post) will be then highlighted with the formulation 'pic related'. This interpretation, however, is speculative, and an analysis of data from 4chan and other image boards would be necessary before it could be stated with confidence.

are less common than forms created via suffixation of a borrowed English verbal root with *-(e)n*.

A caveat must be provided – a corpus from Twitter is not necessarily representative of German as it is used in other text types or in spoken language. For this reason, corpora from different written genres and corpora of transcribed spoken language should be analyzed in terms of new anglicism frequencies as well. One possibility for future work would be to analyze selected 'luxury' borrowings in terms of their semantics, which may be undergoing a process of differentiation from equivalent German lexical items. Here, both qualitative analysis and quantitative approaches such as creating semantic embeddings in a large corpus may lead to interesting results. Another possible analysis could focus specifically on verbal prefixes by comparing their productivity in new anglicisms to their overall prevalence in the German verbal lexicon.

Most new anglicisms are used infrequently, and many may never become established. Nevertheless, by studying the prevalence and grammatical behavior of new verbal forms derived from English in German, we can gain insight not only into how new meanings become necessary in our changing world as a result of technological developments, but also into the processes by which the morphology and semantics of new word forms are reconciled within the larger framework of the language's lexicon and grammatical paradigms. For German, as for other languages, the influence of anglicisms is unlikely to decrease in the immediate future, and thus the study of new verbal anglicisms provides us with an opportunity to investigate and compare lexical and grammatical phenomena that are not only restricted to the German language. In summary, new verbal anglicisms in German are often related to new technologies, but not always, and they gradually assimilate to the German inflectional paradigm, moderated by frequency effects and phonological considerations. In a broader perspective, the study of new anglicisms may provide insight into issues that are more general than German morohology alone, and thus shed light on language evolution and change.

# 6 References

Baeskow, Heike. 2017. #Virtual Lexicality: The semantics of innovative prefixed verbal anglicisms in German. *Word Structure* 10(2). 173–203.

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge, UK: Cambridge University Press.

Biber, Douglas. 1993. Representativeness in corpus design. *Literary and Linguistic Computing* 8. 243–257.

Bird, Steven, Edward Loper & Ewan Klein. 2009. *Natural Language Processing with Python*. Newton, MA: O'Reilly.
Burmasowa, Svetlana. 2010. *Empirische Untersuchung der Anglizismen im Deutschen am Material der Zeitung 'Die Welt'*. Bamberg: University of Bamberg Press.
Carstensen, Broder. 1965. *Englische Einflüsse auf die Deutsche Sprache nach 1945*. Heidelberg: Carl Winter Verlag.
Carstensen, Broder & Ulrich Busse. 1993, 1994, 1996. *Anglizismen-Wörterbuch (3 Bde.)*. Berlin: de Gruyter.
Coats, Steven. 2018. Exploring code-switching and borrowing using word vectors. Paper presented at the 14th Conference of the European Society for the Study of English, Brno, Czechia, 1 September 2018. http://cc.oulu.fi/~scoats/ESSE18.html
Coats, Steven. 2018. Variation of new German verbal Anglicisms in a social media corpus, In Reinhild Vandekerckhove, Darja Fišer & Lisa Hilte (eds.), *Proceedings of the 6th Conference on CMC and Social Media Corpora for the Humanities*, 27–32.
Davies, Mark. 2004–. *BYU-BNC. (Based on the British National Corpus from Oxford University Press)*. https://corpus.byu.edu/bnc/
Davies, Mark. 2008–. *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. https://corpus.byu.edu/coca/
Davies, Mark. 2015. *The Wikipedia Corpus: 4.6 million articles, 1.9 billion words*. https://corpus.byu.edu/wiki/
Duden. 2016. *Die Grammatik* (9th edition). Berlin: Dudenverlag.
Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19(1). 61–74.
Eisenberg, Peter. 2011. *Das Fremdwort im Deutschen*. Berlin and New York: de Gruyter Mouton.
Eisenberg, Peter. 2013. *Anglizismen im Deutschen*. Reichtum und Armut der deutschen Sprache: Erster Bericht zur Lage der deutschen Sprache, 57–119. Edited by Deutsche Akademie für Sprache und Dichtung, Union der deutschen Akademien der Wissenschaften. Berlin: de Gruyter.
Elspaß, Stephan & Robert Möller. 2009. *Atlas zur deutschen Alltagssprache*. http://www.atlas-alltagssprache.de/
Fitschen, Arne. 2004. *Ein Computerlinguistisches Lexikon als komplexes System*. Ph.D. Thesis, Universität Stuttgart.
Galinsky, Hans. 1967. Stylistic aspects of linguistic borrowing: A stylistic view of American elements in modern German. In Broder Carstensen & Hans Galinsky, (eds.) *Amerikanismen der deutschen Gegenwartssprache: Entlehnungsvorgänge und ihre stilistischen Aspekte*, 35–72. Heidelberg: Carl Winter.
Görlach, Manfred (ed.). 2001. *A Dictionary of European Anglicisms*. Oxford, UK: Oxford University Press.
Görlach, Manfred. 2003. *English Words Abroad*. Amsterdam: John Benjamins.
Hanks, Patrick. 2013. *Lexical analysis: Norms and exploitations*. Cambridge, MA: MIT Press.
Kachru, Braj. 1990. *The alchemy of English: The spread, functions, and models of non-native Englishes*. Urbana/Chicago: University of Illinois Press.
Kleuker, David. 2016. *Wortliste*. https://github.com/davidak/wortliste
König, Werner. 1989. *Atlas zur Aussprache des Schriftdeutschen in der Bundesrepublik Deutschland, Band 1: Text*. Ismaning: Max Hueber.
Levinson, Stephen. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Myers-Scotton, Carol. 1997. Code-switching. In Florian Coulmas (ed.), *The handbook of sociolinguistics*, 212–237. Oxford, UK: Blackwell.

Öhmann, Emil. 1970. Suffixstudien VI: Das deutsche Verbalsuffix -ieren. *Neuphilologische Mitteilungen* 71(3). 337–356.

Onysko, Alexander. 2007. *Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching*. Berlin: de Gruyter.

Onysko, Alexander & Emse Winter-Froemel. 2011. Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43(6). 1550–1567.

Paul, Hermann. 1886. *Principien der Sprachgeschichte* (2. Auflage). Halle: Max Niemeyer.

von Polenz, Peter. 1999. *Deutsche Sprachgeschichte vom Spätmittelalter bis zur Gegenwart. Band III: 19. und 20. Jahrhundert*. Berlin: de Gruyter.

Poplack, Shana & David Sankoff. 1984. Borrowing: the synchrony of integration. *Linguistics* 22. 99–135.

Roesslein, Joshua. 2015. *Tweepy*. Python programming language module. https://github.com/tweepy/tweepy

Schmid, Helmut, Arne Fitschen & Ulrich Heid. 2004. SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 1263–1266.

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.

Winter-Froemel, Esme, Alexander Onysko & Andreea Calude. 2014. Why some non-catachrestic borrowings are more successful than others: a case study of English loans in German. In Amei Koll-Stobbe & Sebastian Knospe (eds.), *Language Contact Around the Globe*, 119–142. Frankfurt am Main: Peter Lang.

Wischer, Ilse & Mechthild Habermann. 2004. Der Gebrauch von Präfixverben zum Ausdruck von Aspekt/Aktionsart im Altenglischen und Althochdeutschen. *Zeitschrift für germanistische Linguistik* 32. 262–285.

Yang, Wenliang. 1990. *Anglizismen im Deutschen: Am Beispiel des Nachrichtenmagazins Der Spiegel*. Tübingen: Niemeyer Verlag.