

Changes in the productivity of word-formations patterns.  
Some methodological remarks

Kristian Berg, Carl von Ossietzky  
Institut für Germanistik  
Universität Oldenburg  
26111 Oldenburg, Germany  
kristian.berg@uni-oldenburg.de

Accepted for publication in *Linguistics*

Acknowledgements: I would like to thank Frank Anshen, Mark Aronoff, Nanna Fuhrhop, Wolfgang Klein, Oliver Schallert, an anonymous reviewer, and particularly Harald Baayen and Stefan Hartmann for valuable comments on a first version of this paper.

1 Abstract  
2 Changes in the productivity of word-formation patterns are often investigated using hapax  
3 legomena. In this paper, I argue that at least in diachronic investigations of productivity, a  
4 measure based on first attestations is a viable alternative to hapax-based measures. I show  
5 that such a measure is a more direct proxy to new words than hapax-based measures – it  
6 measures what we want to measure, which is not always true for the latter. I present a  
7 method that deals with the common problem of varying sub-corpus sizes (I suggest we  
8 randomly resample the sub-corpora up to a predefined size), and to the problem of old  
9 words appearing as new at the start of the corpus (I suggest we take an earlier corpus and  
10 determine a point in time when almost all old words have registered). Armed with these  
11 instruments, we can determine the ratio of new types to existing types for a time span,  
12 which can be regarded as the renewal rate of the respective category.

13  
14 Keywords: Productivity, word-formation, diachronic morphology, new words

15 1. Introduction

16  
17 What is morphological productivity and how can it be operationalized? This question has  
18 occupied morphological theory for decades. The common denominator is that productivity is  
19 based on new words. This holds for the locus classicus of productivity, Bolinger's definition  
20 of "the statistically determinable readiness with which an element enters into *new*  
21 *combinations*" (Bolinger 1948:18, my emphasis), but it holds for many others as well, cf. e.g.  
22 Booij's (2012:70) assertion that productive patterns "can be used to form new words", or  
23 Bauer's (2001: 98) definition of productivity as the potential of a morphological process "for  
24 repetitive non-creative morphological coining". Baayen's (1993:183) recapitulation still  
25 holds: "There is a broad consensus that productivity concerns the property of morphological  
26 processes to give rise to new words."

27  
28 From this conceptual foundation on new words arises the major dilemma for synchronic  
29 productivity: In a synchronic corpus, we cannot know which words are (or will be) new. This  
30 is where Baayen's (1989, 1992, 1993) elegant idea comes into play: We may not know how  
31 many words are new, but we can use rare words as an indicator for those new words – a  
32 canary in the coal mine of sorts, as Mark Aronoff put it (p.c.). This works because productive  
33 patterns tend to have many low-frequent formations, whereas unproductive processes are  
34 characterized by more high-frequency types, cf. e.g. Aronoff (1983), Bauer (2001: 147).

35  
36 The usefulness and the ease of application of Baayen's measures, and the fact that they  
37 often correlate with linguists' intuitions (cf. e.g. Baayen 1993), has made them the gold  
38 standard for synchronic productivity. They are also attractive for diachronic investigations of  
39 productivity, cf. (among others) Scherer (2005), Štichauer (2009), and Trips (2009). It is  
40 crucial, however, to remind ourselves that "P is an indirect measure of productivity" (Bauer  
41 2001: 153). There is a much more direct measure when dealing with diachronic data: New  
42 words (cf. e.g. Cowie 1999, Anderson 2000). It is easy to determine for each time span in a  
43 diachronic corpus which words are new. There are some methodological challenges;  
44 addressing them is one of the purposes of this paper. The quintessential point is this:  
45 Productivity is conceptually based on new words, and we can determine new words in a  
46 diachronic corpus. A measure based on new words has a number of other advantages, as I  
47 will show below.

1 Like most research on morphological productivity, this investigation will be limited to  
2 derivational word formation. I use five German derivational suffixes to demonstrate the  
3 usefulness of the proposed measures.  
4  
5 Of course, there are other aspects of productivity that a focus on new words does not  
6 capture. For example, both the number of types and the number of tokens in use at any  
7 given time highlight important aspects of a pattern's morphological productivity (see e.g.  
8 Baayen 2009 for an intuitive commercial analogy). Those aspects are largely neglected in the  
9 present paper as measures in their own right. This is simply due to space limitations, not  
10 because I believe them to be unimportant.

11  
12 The structure of this paper is as follows: In the next section (2.), I will discuss related work on  
13 diachronic productivity, before I give an overview of the corpus I have used for the  
14 investigation and the derivational suffixes I have looked at (3.). After that, I will give a brief  
15 overview of approaches to diachronic word formation, and building on this, I will develop a  
16 measure of diachronic productivity based on new words (4.). I will then use this measure in  
17 an investigation of five German suffixes (5.). The measure is compared to the most widely  
18 used measure in diachronic productivity, Baayen's  $\mathcal{P}$ , in the subsequent section (6.). The  
19 main findings are summarized and discussed in the last section (7.).  
20

## 21 2. Related work: Diachronic productivity

22

23 The field of research on productivity is vast. In the following, I will limit the discussion to  
24 diachronic productivity, and more precisely, to quantitative approaches<sup>1</sup>. These approaches  
25 basically come in two different flavors: They are either based on new formations, or they use  
26 Baayen's hapax-centered measures<sup>2</sup>. I will comment on both in turn.  
27

28 Approaches that measure productivity using new words rely mostly on dictionaries. One of  
29 the first works of this kind is Neuhaus (1971, 1973), who measures the number of first  
30 attestations of *-ish* words in the OED per decade. In a similar manner, Anshen & Aronoff  
31 (1989) use a reverse dictionary (Walker's rhyming dictionary) to collect words with a given  
32 suffix (in this case, e.g. *-ness* and *-ity*). These words are then checked for the date of their  
33 first attestation in the Oxford English Dictionary (OED). This way, they can determine which  
34 pattern produces more new words. Arguably, the greatest weakness of this approach is that  
35 only those words with an entry in the reverse dictionary can be checked for their first  
36 attestation. Ten years later, Anshen & Aronoff (1999) manage to overcome this limitation by  
37 using a more sophisticated method on an electronic version of the OED. They semi-  
38 automatically extract every word in the OED with *-ment* and *-ity* to determine the factors for  
39 the decline in productivity of the former and the stability of the latter. Anshen & Aronoff  
40 essentially use the raw numbers of new words per time span (in their case, half-century) and  
41 compare the numbers for the two suffixes: Whichever has more new words per time span is

---

<sup>1</sup> For a comprehensive general overview of productivity, see e.g. Bauer (2001) or Baayen (2009); for a comprehensive overview of productivity measures see Hilpert (2013: 127ff.); for a general overview of diachronic productivity see Scherer (2015).

<sup>2</sup> There are also approaches like Plag (1999) and Bolozky (1999) that use measures based on new words as well as those based on hapaxes.

1 more productive. Note that this method actually only allows Anshen & Aronoff to compare  
2 the productivity of different patterns in the same time span. Statements like “the number of  
3 borrowed -ity forms slowly increases and then, starting at the second half of the 17th  
4 century, decreases” (Anshen & Aronoff 1999: 21) are not wrong, but they cannot be  
5 generalized to mean that the productivity declined from 1750 onwards: It may simply be  
6 that the OED contains less data in that period. To avoid this problem, the data need to be  
7 normalized. This is a notorious issue in quantitative diachronic linguistics, and I will return to  
8 it several times in this paper. The question of what to normalize for is far from trivial. Lindsay  
9 & Aronoff (2013), for example, building on the method used in Anshen & Aronoff (1999),  
10 decide to use the number of types in the OED for each time span.

11

12 According to Baayen & Renouf (1996: 69f.), there is a fundamental problem with the use of  
13 dictionaries. Truly productively formed words, which are semantically transparent and very  
14 rare, will often not be included in the lexicon, primarily for economic reasons. As a  
15 consequence, dictionary-based measures underestimate the actual productivity, at least of  
16 productive processes (but cf. Plag 1999: 97:99 who argues that arguments like this against  
17 the use of dictionaries do not apply to the OED).

18

19 Dalton-Puffer (1996), Cowie (1999) and Cowie & Dalton-Puffer (2002) avoid the pitfalls of  
20 dictionaries by counting new words in diachronic corpora. The problem of normalization  
21 remains, though: If the corpus is not evenly sized, that is, if certain time-spans consist of  
22 more running words than others, what should the count data be normalized for? Cowie &  
23 Dalton-Puffer (2002: 427) argue against the use of tokens – as in e.g., ‘In 1800, there are 31  
24 new -isch types per 1,000,000 tokens’ – because that would amount to “not counting like  
25 out of like”. Yet then the authors go on to do just that, and normalize new words for 100,000  
26 tokens (p. 430f.). I will return to this problem below (section 4).

27

28 Hapax-based measures, on the other hand, have been developed by Baayen (1989, 1992,  
29 1993) for synchronic investigations. As briefly noted above, they provide an elegant way to  
30 gauge the synchronic productivity of a word formation pattern without access to  
31 information about which words are new. The central assumption is that productive patterns  
32 produce many low-frequency words. The reasoning for this is as follows: Low frequency  
33 words, especially those with just one attestation, so-called hapax legomena, are probably  
34 semantically transparent; if they were not, their meaning could hardly be inferred.  
35 Productive patterns, on the other hand, produce semantically transparent words; it thus  
36 seems probable for productive processes to produce many rare formations. The most widely  
37 used measure  $\mathcal{P}$  is called ‘productivity in the strict sense’ (Baayen & Lieber 1991: 817),  
38 ‘potential productivity’ (Baayen 2009), or ‘category conditioned degree of productivity’  
39 (Baayen 2001:157). It relates the number of hapaxes of a category to the total token count  
40 of the category ( $\mathcal{P} = \frac{n_1}{N}$  with  $n_1$ =hapaxes and  $N$ =sum of all tokens). As such,  $\mathcal{P}$  is the  
41 probability of encountering a hapax after  $N$  tokens.

42

43 The ease of computation and the fact that the measure yields intuitive results has led to the  
44 acceptance of  $\mathcal{P}$  as the de-facto standard for measuring productivity. It is not surprising that  
45 it is regularly used as the central measure in diachronic investigations as well, cf. e.g. Scherer  
46 (2005), Schneider-Wiejowski (2011), Trips (2009), or Schröder (2011). The way the measure  
47 is used is that the corpus is divided into consecutive sub-corpora, with  $\mathcal{P}$  values computed  
48 for each sub-corpus. There are two main problems with this. First, the corpus has to have a

1 certain minimal size: In smaller corpora, there will be more hapaxes. To put it differently,  
2 many words that are not really rare will nevertheless appear only once in a small corpus  
3 simply because it is not large enough to include the respective second attestations (cf. Dal  
4 2003: 17). Second, the pitfalls discussed above apply: To be able to compare two  $\mathcal{P}$  values,  
5 the sub-corpora should be of equal size, or they should be normalized. Otherwise,  
6 differences between the values are much harder to interpret; they may simply be due to one  
7 time span accidentally containing more tokens than the other. However, as the relation  
8 between tokens and hapaxes is not linear (cf. e.g. Baayen 1992: 113), any linear  
9 normalization is necessarily too simplistic.

10  
11 A solution to this problem is to extrapolate the number of hapaxes for sample sizes larger  
12 than the actual size (cf. e.g. the model of Baroni & Evert 2005). And as Gaeta & Ricca (2006)  
13 have argued, it is not the sample size of the (sub-)corpus that needs to be standardized, but  
14 the token count for individual patterns. The larger the number of tokens of a pattern, the  
15 lower is  $\mathcal{P}$ . If a pattern is instantiated by many tokens, its  $\mathcal{P}$  value is automatically lower  
16 than that of an infrequent pattern. Comparing the unadjusted  $\mathcal{P}$  values will thus “always  
17 imply an overestimation of the values of P for the less frequent suffixes”, as Gaeta & Ricca  
18 (2006: 63) state. Accordingly, contemporary investigations of diachronic productivity (e.g.  
19 Schneider-Wiejowski 2011, Hartmann 2018) still center around Baayen’s  $\mathcal{P}$ , but they modify  
20 their corpora with elaborate mathematical extrapolation procedures like finite Zipf-  
21 Mandelbrot models.  
22

### 23 3. Data bases and suffixes

24  
25 Before we turn to the methods I would like to put forward, let us take a brief look at the  
26 data that will be used to test these methods. I will use the large diachronic corpus Deutsches  
27 Textarchiv (DTA) as the main data base in this paper (cf. [www.deutschestextarchiv.de](http://www.deutschestextarchiv.de)). As of  
28 December, 2017, it contains 2,419 printed texts from the late 15th century to the early 20th  
29 century, totaling 138 million tokens. It is not balanced for genre; it includes newspaper texts,  
30 scientific publications, novels, pamphlets, etc. The main advantage of the corpus, however,  
31 and the primary reason for its particular useability for a diachronic investigation like this one,  
32 is that it is graphemically normalized and lemmatized. The original spelling of every word is  
33 accessible, but it is enriched with information about graphemic and lexical invariants: Each  
34 running word is mapped onto a graphemically normalized token form (e.g., the spelling  
35 variants *Kleid*, *Kleidt*, *Kleyd*, and *Kleydt* ‘dress’ are all mapped onto the same invariant form  
36 *Kleid*). Each graphemically normalized token is then mapped onto one lemma (e.g. the  
37 normalized tokens *Kleid* ‘dress’ and *Kleider* ‘dresses’ are mapped onto the lemma *Kleid*).  
38 With this lemmatization and graphemic normalization, it is possible to easily extract all  
39 lemmas that end with a certain suffix.

40  
41 Reasons of validation (see next section) make it necessary to also use corpora that cover the  
42 time before and after the main corpus. To cover the time before the DTA corpus, I chose the  
43 Early High German (EHG) corpus *Bonner Frühneuhochdeutschkorpus*  
44 (<http://www.korpora.org/fnhd/>, Lenders & Wegera 1982). This corpus contains 40 Early New  
45 High German texts from the 14th to the 17th century with a total of around 600,000 tokens.  
46 Like the DTA corpus, it is lemmatized, which facilitates the extraction of types. The time  
47 following the main corpus is covered by the 20<sup>th</sup> century core corpus of the *Digitales*

1    *Wörterbuch der Deutschen Sprache* ([www.dwds.de](http://www.dwds.de)). It contains between 10 and 13 million  
2    running words per decade and is balanced for genre (cf. Geyken 2007). Even though the  
3    corpus is not fully accessible due to copyright laws (it cannot be downloaded, for example,  
4    and only a limited number of results can be displayed), it has an application programming  
5    interface (API) which is used to automatically collect frequency data from the web server of  
6    the corpus.

7  
8    The suffixes I investigate are all derivational suffixes from different domains, adjective  
9    suffixes (-*isch*, -*bar*), nominal suffixes (-*nis*, -*tum*), and a verbal suffix (-*ieren*). The suffixes  
10   exhibit different degrees of productivity: -*isch* and -*bar* are supposedly highly productive in  
11   modern-day German (see below for references); -*nis* and -*tum*, on the other hand, are  
12   regarded as hardly productive, and the status of -*ieren* is unclear. For one suffix, -*isch*, there  
13   is a recent investigation concerning its diachronic productivity (Kempf 2016); this is not the  
14   case for the other suffixes. Accordingly, the following short summary will mostly refer to the  
15   kind of base the suffix operates on, the semantic relation between base and derivative, and  
16   the synchronic productivity.

17  
18   -*isch*: This pattern yields adjectives from nouns, which are mostly native person nouns (e.g.  
19   *mörderisch* ‘murderous’). It also serves to integrate bound non-native bases as  
20   adjectives (e.g. *elektrisch* ‘electric’). In her diachronic investigation, Kempf (2016:  
21   248ff.) observes a strong increase in productivity between 1450 and 1700 (using a  
22   variety of hapax- and neologism-based measures). Kempf hints at a slight decline in  
23   the 19th century. Today, the pattern is regarded as highly productive (cf. e.g. Lohde  
24   2006: 184).

25   -*bar*: This is an adjective suffix that operates mostly on verbal bases. These verbal bases  
26   are in turn mostly transitive verbs that form a regular passive (cf. Fleischer & Barz  
27   2012: 332ff.). The resulting adjectives have a passive-like quality and denote the  
28   possibility of the action the verb describes (e.g. *essbar* ‘able to be eaten’). Today, the  
29   pattern is the most productive pattern for de-verbal adjectives (Fleischer & Barz  
30   2012: 332). Historically, the suffix also operated on nominal bases, e.g. *dankbar*  
31   ‘thankful’, cf. Flury (1964).

32   -*nis*: This suffix forms abstract nouns from verbs (e.g. *Besäufnis* ‘the act of getting drunk’).  
33   Nominal and adjectival bases are very rare. The pattern is unproductive today (cf.  
34   Motsch 2004: 359).

35   -*tum*: This noun suffix operates mostly on nominal bases, more precisely, on person nouns.  
36   It yields abstract nouns that may denote both the key qualities of the person (or  
37   group of persons) that the base refers to (as in e.g. *Abenteurertum* ‘the quality of  
38   being an adventurer’) or the collectivity of those persons (as in e.g. *Bürgertum* ‘the  
39   collection of all citizens [i.e., the bourgeoisie]’). This pattern is only marginally  
40   productive today (cf. Lohde 2006: 107).

41   -*ieren*: This verb suffix operates mostly on foreign bases; many of them are bound bases  
42   (e.g. *informieren* ‘to inform’), but there are a few native free bases as well (e.g.  
43   *amtieren* ‘to hold office’). If the bases are free, they are mostly nouns. Semantically,  
44   the formations are very heterogenous (cf. Fleischer & Barz 2012: 432).

45  
46

1    4. Methodology

2  
3    The method proposed here is reasonably straightforward: Divide a diachronic corpus into  
4    time-spans and count the new words with a given word formation pattern. As with any  
5    apparently straightforward measure, the devil is in the detail: What is an appropriate time  
6    span, and which words belong to the word formation pattern in question? The size of the  
7    time span hinges on the corpus size, but is apart from that largely a matter of convenience.  
8    From 1700 on, the DTA contains for the most part at least 4 million tokens per decade (data  
9    before 1700 will be discarded for reasons independent of periodization, see below). A  
10   minimum of 4 million words per decade sub-corpus seems to be a reasonable size compared  
11   to much smaller synchronic corpora like the 600,000 token Eindhoven corpus, for which  
12   Baayen (1993:187) claims that “the major patterns of productivity already emerge very  
13   clearly”.

14

15   Which words belong to the word formation pattern in question? Is e.g. *Nutzbarkeit*  
16   ‘useability’ an instance of the word formation pattern *-bar* (*nutzbar* ‘useable’), which is in  
17   turn derived with the nominal suffix *-keit*? I argue with (among others) Baayen (2009: 903)  
18   that only those words should be included where “a given rule was the last to apply” (in a  
19   synchronic-structural sense, not historically). This has the advantage that the extraction of  
20   the pattern from the corpus is greatly facilitated: We simply search for lemmas that end with  
21   the suffix in question and meet the part of speech criteria<sup>3</sup>. To avoid results like *frisch* ‘fresh’  
22   for *-isch*, the search pattern demands that the suffix be preceded by at least three letters;  
23   there are only a few German stems with less than three letters (cf. Berg 2017: 115ff.). The  
24   results are then corrected and cleaned manually. In some instances, a faulty lemmatization  
25   leads to two distinct entries for one lemma (e.g. *Kaisertum*, *Keyserium* ‘empire’); this is  
26   adjusted. In other cases, the pattern in question is not the last step in the derivation: for  
27   example, *undruckbar* ‘unprintable’ has to be analyzed as a derivation of the adjective  
28   *druckbar* ‘printable’, and the entry thus has to be excluded. The same holds for the nominal  
29   suffixes, where compounds with *-tum* and *-nis* as the second element has to be filtered (e.g.  
30   *Familienheiligtum* ‘family sanctuary’). And finally, many Latin words (or words of Latin origin)  
31   ending in the letter sequence *nis* or *tum* are excluded (e.g. *tyrannis*, *Quantum* ‘quantum’).  
32   Entries without free bases, however (like e.g. *elektrisch* ‘electric’), and semantically opaque  
33   entries are not excluded (contra Gaeta & Ricca 2006:74f.): These entries still conform to the  
34   output constraints of the word formation pattern (cf. e.g. Aronoff 1976).

35

36   Apart from these questions, there is a general problem that is an almost chronic issue for  
37   diachronic investigations: Many diachronic corpora vary in size for their time spans. As a  
38   case in point, take the corpus used in this paper. Figure 1 shows that the decades from 1470  
39   to 1900 vary greatly with respect to the tokens they contain. This holds both for the sum of  
40   tokens in each decade as well as for the tokens in each genre.

41

---

<sup>3</sup> While this catches most of the word forms we are looking for, some still escape because of faulty lemmatization. An additional token-based search for inflected forms and spelling variants is necessary (e.g. for *-tum* I searched for tokens that end with either *t(h)um*, *t(h)um(e)s*, *t(h)ümer*, *t(h)ume* or *t(h)ümern* (and that have not already been collected)).

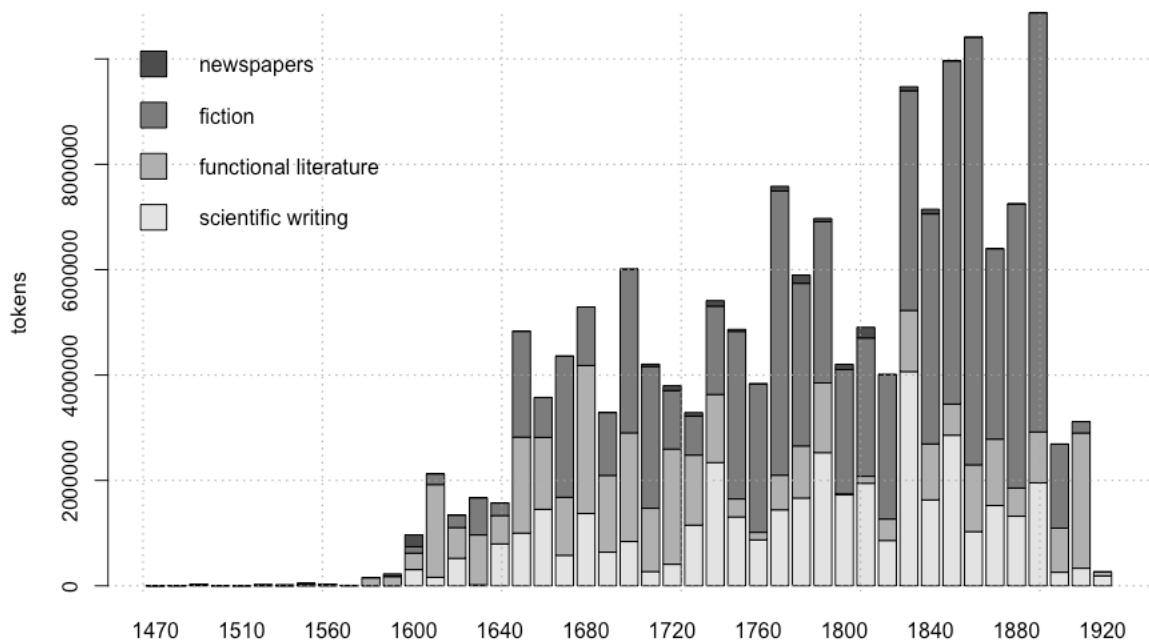
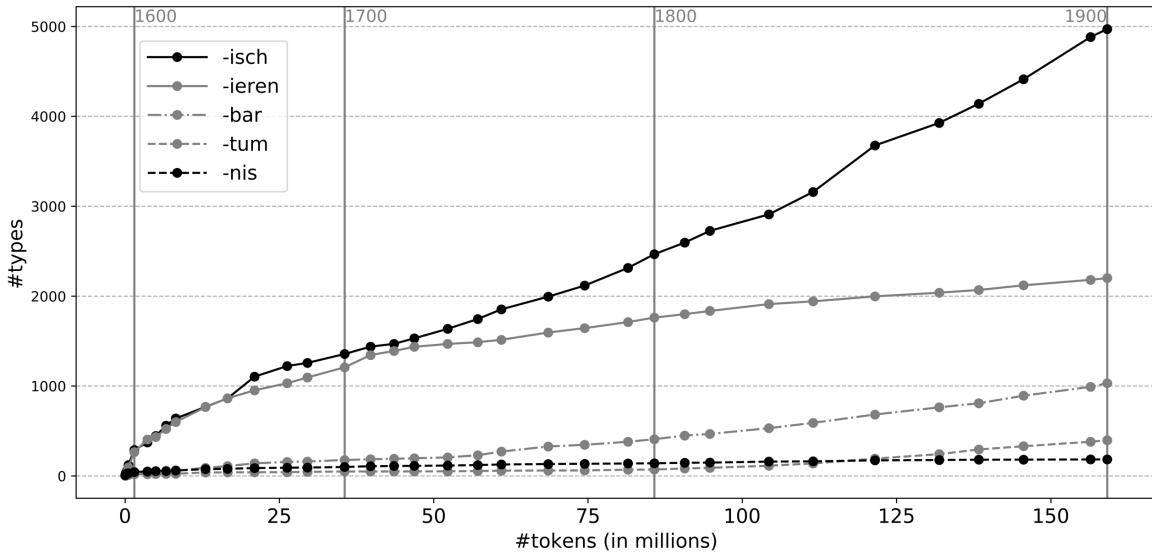


Figure 1: Barplot of the number of tokens per decade in the corpus Deutsches Textarchiv.

Not surprisingly, more recent decades contain more data (although there are exceptions). For a productivity measure based on new words, this is problematic because more tokens mean more types – but, as mentioned above, the relation is not linear. Thus we cannot simply list the set of new words for each decade if the decades differ in size as considerably as they do in the DTA, and we can also not normalize for tokens and state the number of new words relative to some chosen number of tokens (e.g., ‘in 1710, there are 14 new words with *-isch* per 1,000,000 tokens’).

The easiest way to deal with this issue is to look at the number of new types as a function of the number of tokens in that particular time slice. Figure 2 plots this relationship for the five suffixes. The y-axis could equally well be labelled ‘vocabulary’; the graph shows a diachronic vocabulary growth curve in the sense of Baayen (1992: 113). In some sense, this is the ‘purest’ visualization of the data; it does not exclude any data and rests on a minimum of presuppositions. Each dot represents the number of types of a word formation pattern in a given decade.

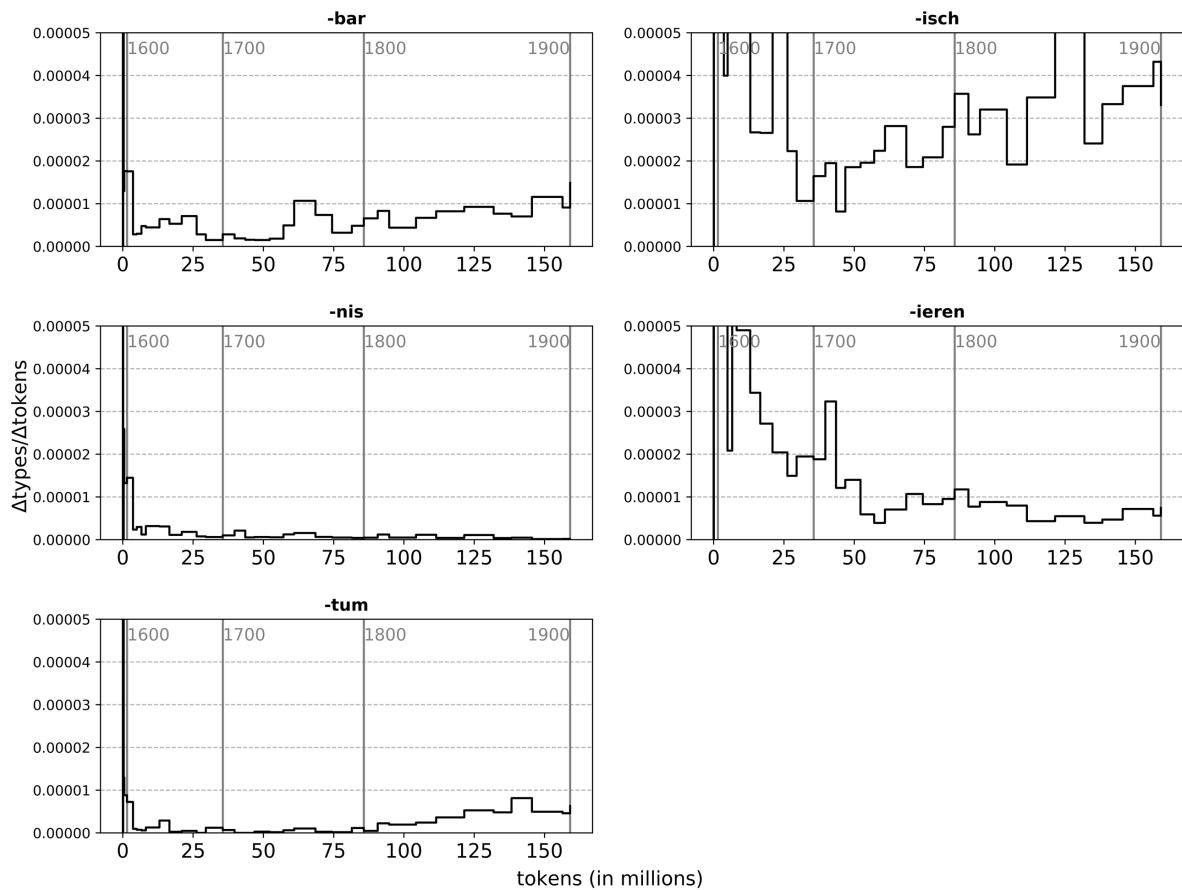


1  
2 *Figure 2: The number of types as a function of the number of tokens for the five patterns (-bar, -isch, -nis, -ieren,  
3 and -tum) Data base: corpus Deutsches Textarchiv (DTA).*

4  
5 Figure 2 reveals several facets of the data. First, the size of the decades and centuries (in  
6 tokens) is visible as the horizontal distance between the dots. It varies with respect to the  
7 tokens they contain: the 19th century has more tokens than the 18th, which in turn has  
8 more than the 17th.

9  
10 Second, the size of the combined types at the end of the corpus varies: *-isch* has the largest  
11 vocabulary over time (4,971 types), and *-nis* the smallest (183 types). Take a moment to  
12 note, though, that the number of types at the end of the corpus is not the number of types  
13 in the last decade; it is the accumulated set of types that was gathered over time, and it  
14 contains words that have fallen out of use, and many words that were used just once a long  
15 time ago. The last substantial decade in the corpus, 1890–1899, contains 2,088 different  
16 types with *-isch* (as compared to the 4,971 types in the whole corpus), and 65 different types  
17 with *-nis* (as compared to the 183 types in the whole corpus). Baayen (2009: 901f.) is thus  
18 right in his assessment that the size of the vocabulary of a word formation pattern (what he  
19 calls “realized productivity”) is – at least partly – a “past achievement”; it contains the  
20 successful word formations from the past, as well as new ad-hoc formation; but it does not  
21 contain past ad-hoc formations. Even though realized productivity is part of a synchronic  
22 investigation of productivity, it is a rather coarse measure for a diachronic investigation. We  
23 can disentangle past and present achievements for each time span by using first  
24 attestations, and thus arrive at a more accurate measure, as I will show below.

25  
26 Third and most importantly, figure 2 shows the rates with which new types appear, and how  
27 this rate changes over time. For example, new *-isch* words are produced at a higher rate  
28 than new *-ieren* words from around 1750; the *-isch* curve rises more steeply than the curve  
29 for *-ieren*. The slope of the line thus indicates changes in productivity. This slope of the line is  
30 the difference in types divided by the difference in tokens,  $\Delta$ new types/ $\Delta$  tokens (if we used  
31 approximation functions to the lines in figure 2, we would look at their first derivative, cf.  
32 e.g. Baayen 2001: 49). Note that the difference in types in each time span equals the newly  
33 attested words. The result for the five suffixes is plotted in figure 3:



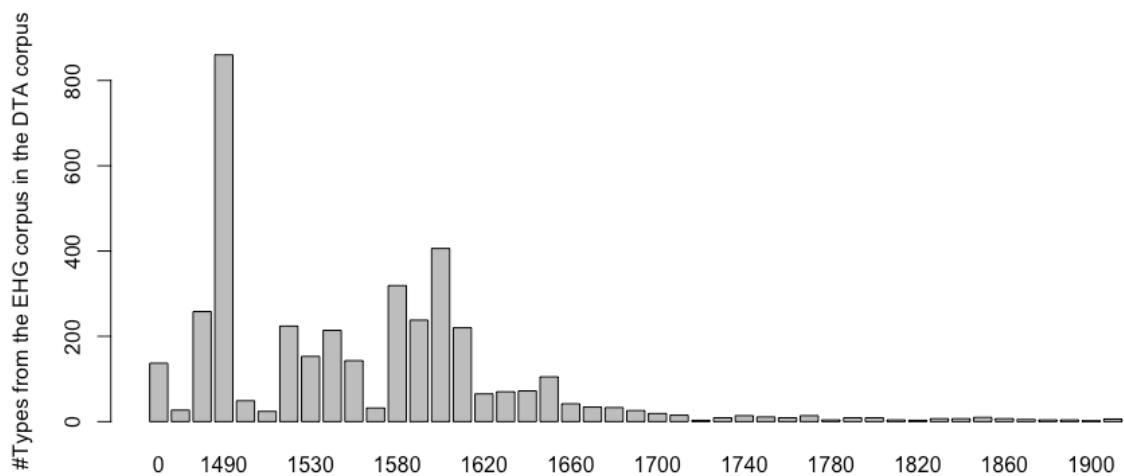
1  
2 *Figure 3: Change in the number of new words with a given pattern divided by the change in tokens with that*  
3 *pattern for the five patterns (-bar, -isch, -nis, -ieren, and -tum). Data base: corpus Deutsches Textarchiv (DTA).*

4  
5 In this representation, we can clearly see the level of productivity and its changes over time.  
6 The suffix *-isch* has the highest rate with around 0.00003 new types per new token, or 30  
7 new types per 1,000,000 tokens. The level is rising slightly over time. The level for *-ieren* is  
8 stable (at least from around 1750 onwards) at a considerably lower rate of below 0.00001  
9 new types per new token. The level for *-bar* slowly rises to 0.00001 new types per new  
10 token. For *-tum*, the rise starts later (around 1800), and it does not reach the level of *-bar*.  
11 Finally, there do not seem to be any changes in productivity for *-nis* – the word formation  
12 pattern is unproductive.

13  
14 There are three major problems with this kind of methodology. One is that the curves in  
15 figure 3 all show a bias towards higher productivity at earlier stages. A second one is that we  
16 are de facto normalizing over tokens, even though, as was stated above, the relation  
17 between types and tokens is not linear. And finally, if we are calculating new types as a  
18 function of tokens, we are “not counting like out of like” (Cowie & Dalton-Puffer 2002:427). I  
19 will address these issues in turn.

20  
21 The bias towards higher productivity levels at the beginning of the corpus has often been  
22 observed (cf. e.g. Cowie & Dalton-Puffer 2002: 429; Kempf 2016: 116). It follows from the  
23 fact that words which predate the corpus “must initially register” (Cowie & Dalton-Puffer  
24 2002: 429). The higher levels of new words that we observe are (for the most part) not new  
25 at all, but older words that make their first appearance in the corpus. The levels can thus not  
26 be taken at face value. There are several ways to deal with this issue. Cowie (1999) suggests

1 using a different, earlier corpus as a ‘starting lexicon’: If any given word already appears in  
 2 the older corpus, it does not count as a new word. While this is generally a good idea, a  
 3 prerequisite is that the older corpus is large enough to actually contain all (or most) words  
 4 that predate the later corpus.  
 5  
 6 I would like to present a different way to handle this problem, one that also works if the  
 7 older corpus is considerably smaller than the actual corpus (for a similar but more qualitative  
 8 approach, see Kempf’s (2016) and Kempf & Hartmann’s (2018) “comparative dating  
 9 method”). The idea is simple: We determine the set of types in the earlier corpus and track  
 10 their first attestation in the actual corpus. This should lead to a curve which is high at initial  
 11 stages of the newer corpus (older words are registering), but which should flatten at some  
 12 point – and it is this point (or period of time) that we are interested in. From here on, the  
 13 majority of older words has been encountered.<sup>4</sup>  
 14  
 15 For the present investigation, I used the Early New High German corpus introduced above  
 16 (section 3). This corpus overlaps with the DTA corpus, which starts in 1473; as a  
 17 consequence, I only use texts up until 1500 from the Early New High German corpus. From  
 18 these 20 texts, a set of 4,919 distinct types is extracted, and the first attestation of each of  
 19 these types in the DTA is determined.  
 20  
 21 Of these 4,919 types in the Early New High German corpus, 3,789 also appear in the DTA  
 22 corpus (77%). As expected, most of them are first attested in the early stages of the DTA.  
 23 Figure 4 shows a barplot of the number of types according to the decade of their first  
 24 attestation in the DTA:



25  
 26 Figure 4: Number of first attestation of types from the Early High German corpus in the DTA corpus per decade.  
 27 Example: In 1490, more than 800 types from the earlier corpus are first attested in the DTA corpus.  
 28

---

<sup>4</sup> Another method that could even do without an older corpus is to use morphologically simple words as a baseline and track their vocabulary growth curve. Morphologically simple words predate *cum grano salis* today’s corpora. Once most of these morphologically simple words have appeared (i.e., once the growth curve flattens), we can infer that most older words in general have registered. If the corpus is not annotated morphologically, we could use short words as an approximation to simple words, e.g. the set of four letter words (I would like to thank Mark Lindsay for this idea).

1 Words from the earlier corpus are first attested even in the latest decades of the DTA. But  
2 this is mostly due to the lemmatization practice of the Early New High German corpus: For  
3 example, the token *paraliticus*, which appears twice in the corpus, is clearly a Latin loan  
4 word (evidenced by the Latin suffix *-us*); yet it is lemmatized as *Paralytiker*, which first  
5 appears in the DTA corpus around 1900. Thus the actual distribution is probably even  
6 clearer, with fewer types at later stages. In any case, 1700 is a sensible cut-off point: As can  
7 be seen in figure 4, the number of newly attested (old) types after this decade is marginal. In  
8 total, only 140 of the 3026 types from the Early High German corpus are first attested 1700  
9 or later; that is a ratio of 4.9%. That means that the DTA data before 1700 is potentially  
10 (though not necessarily) skewed when it comes to first attestations, and we should focus on  
11 the decades from 1700 on.

12  
13 The second problem with figure 3 mentioned above is that we are normalizing the token  
14 count – we arrive at statements like ‘for *-isch*, there are about 30 new types for every  
15 1,000,000 tokens’. But again, the relation between types and tokens is not linear, and we  
16 have every reason to believe that this also holds for the relation between new types and  
17 tokens. The remedy I suggest is twofold: Make the time spans similar with regard to their  
18 token count, and normalize for types instead of tokens. Both steps lead to a measure that is  
19 at the same time more stable and a more direct proxy to new words, as I will show below.  
20 This measure also takes care of the third issue raised above, the fact that we were counting  
21 types out of tokens.

22  
23 For corpora that are not balanced over time (as the DTA, see figure 1 above), adjusting the  
24 size of the time spans is the easiest way to ensure comparability of different time spans. The  
25 first step is to determine the target size of the time span. As mentioned above, the minimum  
26 size almost all decades after 1700 have is 4,000,000 tokens (remember that we are excluding  
27 data before 1700), so we aim for this size in each decade. Instead of carefully curating texts  
28 for the decades (which is of course possible), I suggest we use a Monte Carlo simulation: We  
29 randomly re-sample each decade’s texts out of all texts for this decade until we reach the  
30 target size<sup>5</sup>, then we move on to the next decade until all decades have been re-sampled.  
31 We do this 100,000 times, and we thus end up with 100,000 alternatively re-sampled  
32 versions of the main corpus. We note for each permutation of the corpus the number of  
33 new types and the number of all types with the word formation pattern in question. After all  
34 permutations have been computed, we calculate the mean and the range into which 95%  
35 and 99% of the data fall.

36  
37 If the time spans are of equal size, we can simply count and compare the numbers of new  
38 words in each decade without normalizing. I will call this measure  $V_{neo}$ . Suppose pattern A  
39 has 200 new words in a given decade, and pattern B has 100 new words in the same decade,  
40 then pattern A has a higher  $V_{neo}$  value; it is more productive according to this measure.

41  
42 However, comparing the changes in new words for two different word formation patterns is  
43 potentially difficult if their respective vocabulary sizes differ. Using the example above,  
44 suppose the 200 new words for pattern A occur alongside 200 old words, while the 100 new

---

<sup>5</sup> More precisely, we add the decade’s texts in random order until we exceed the target size; for the last text, we cut the excess tokens so that the total number of tokens matches the target size. By keeping the texts intact, we accommodate for the fact that words are not distributed randomly throughout texts and collections of texts (cf. Säily 2016: 132).

1 pattern B words occur with just 50 old words. This changes the picture: Pattern B has a  
2 higher renewal rate, and we should take that into account. This renewal rate is the same as  
3 Bolozky's (1999: 5) R1, the "pattern's potential to 'regenerate itself'", and it is similar in spirit  
4 to Baayen's (1993) category conditioned degree of productivity<sup>6</sup>. The reasonably simple  
5 formula for the neologism-based productivity  $P_{\text{neo}}$  for a time span  $t$ , a corpus size  $N$ , and a  
6 word formation pattern  $wfp$  is given in (1). We count all types that first occur in  $t$  and divide  
7 them by the count of all types attested in  $t$ :

8

$$9 \quad (1) \quad P_{\text{neo}}(wfp, t, N) = \frac{\text{new types } (wfp, t, N)}{\text{all types } (wfp, t, N)}$$

10

11  $P_{\text{neo}}$  can be interpreted as a probability, the probability of encountering new types with this  
12 pattern; the  $V_{\text{neo}}$  measure introduced above is simply a count measure. The main advantage  
13 of  $P_{\text{neo}}$  is that it is directly interpretable as the fraction of all types in a time span that are  
14 new. To give an example from the actual investigation (see below), in 1840–1849, there are  
15 1,372 types with the pattern *-isch* in the first of 100,000 4-million token versions of the DTA  
16 corpus, and 250 of them had never occurred before (i.e., in the previous 4-million word  
17 samples leading up to 1840). That means 18.2% of the types in that decade were new.  
18 Compare that to a pattern which is usually believed to be unproductive, *-nis*: In the same  
19 decade, only three of the 68 total types were new, which is a ratio of 4.4%.

20

21 Baayen's  $\mathcal{P}$ , on the other hand, denotes the fraction of all tokens with a word formation  
22 pattern that occur only once. It rests on the assumption that hapaxes are indicative of new  
23 words. For diachronic investigations, the productivity measure introduced in (1) is  
24 potentially more useful: First, it is more direct in that it rests on first attestations, which are  
25 a close approximation of 'actual' new words in the language community; we do not need to  
26 take the (synchronously necessary) detour via hapax legomena. Second, the new words can  
27 in turn be investigated qualitatively: How did the new formations change, and what does  
28 that say about the respective constraints of the word formation process? And third, if we tie  
29 productivity to the 'birth' of words, we can also take the 'death' of words into account and  
30 look at the life-span of word formations. Are they nonce formations, or do they become  
31 institutionalized or lexicalized? And the measure in (1) is not only more useful in diachronic  
32 investigations, it is also more stable than hapax-based measures. I will demonstrate this in  
33 the following sections.<sup>7</sup>

34

35 A few methodological issues need to be addressed. When is a word a neologism? This seems  
36 to be a trivial question, but there are several pitfalls. First of all, we can only ascribe this  
37 status to words relative to the corpus we are using. In a larger corpus, we will probably find

---

<sup>6</sup> Cowie (1999: 133f.) uses the same measure, although without keeping the time spans at a fixed size (and without giving it a name). I would like to thank Stefan Hartmann for pointing this out.

<sup>7</sup> There is one caveat: We are only making statements for the productivity of the patterns in a corpus of size X, and a time span of size Y. Yet with this in mind, we can very well compare patterns that differ in their total token count. Gaeta & Ricca (2006) argue that Baayen's  $\mathcal{P}$  measure is systematically too high for low-frequency patterns, and they suggest to calculate the expected number of hapaxes for larger token counts. This is generally not necessary if we remind ourselves that the productivity values we obtain are always only for certain corpus sizes. After all, that is psycholinguistically plausible: Language users gauge the productivity of a word formation pattern on the basis of the input they get – and that input (in the form of running words of text) is the same for all word-formation patterns, which in turn means that some (like compounding in German) are much more frequent than others (for example derivation with *-nis*).

1 earlier attestations for many words. If we accept that, then a word in the corpus is new if it  
2 was never attested before in the corpus. As a corollary of this definition, words that appear  
3 again after a longer period of absence are not new. Take the adjective *katalanisch* ‘Catalan’,  
4 for example. It is first attested in the DTA corpus 1653, where it appears once; the next  
5 attestation is in 1821, after 168 years of radio silence. In a certain way, the 1821 attestation  
6 is also new: The writer had probably never seen the word *katalanisch* in print before, and  
7 one can argue that the word was thus new in the 1821 document. But this distribution may  
8 as well be an effect of the corpus size and the low frequency of the word itself; after all, we  
9 have to remind ourselves: We are only monitoring a small sample of language use over time.  
10 It may well be that the large gap we witness is accidental. For this reason, we adhere to the  
11 rather strict definition given above: A word is new when it is first attested; from then on, it is  
12 old, no matter for how long it is not recorded after that first attestation.<sup>8</sup>

13  
14 Second, and more general, the approach to newness is very naïve. All words that occur for  
15 the first time at some point in the corpus are new, but of course, at least some of them were  
16 coined deliberately and intentionally, for example by writers (I would like to thank Harald  
17 Baayen for pointing this out).

18  
19 Third, quantitative approaches cannot readily capture changes in the function of the suffix.  
20 Take the German suffix *-mäßig* as in *wettermäßig* ‘weather-wise’. The Old High German  
21 version of this suffix, *-māzi/-māzig* referred more explicitly to the base *māza* ‘measure’:  
22 *Xmāzi* meant ‘having the dimension of X’ (Fleischer & Barz 2012: 346).

23  
24 Fourth, it is well documented that productivity in word formation also varies with register or  
25 genre (cf. e.g. Plag, Dalton-Puffer & Baayen 1999). Even though the DTA corpus is annotated  
26 for genre, it is not balanced for genre, and the contributions of the genres to the overall  
27 token count vary considerably, making it difficult to use the sampling method introduced  
28 above. Accordingly, in what follows I will ignore the influence of genre – knowing full well  
29 that I am probably discarding information that could account for some variation in the data.

30  
31 And finally, using a fixed target size of tokens for all decades, we run into the conceptual  
32 problem that the historic population of the German speaking countries grew over time  
33 (which in turn lead to a steadily growing output of texts over time), although not  
34 monotonously (cf. the Thirty Years’ War). Cutting each decade at a fixed target size ignores  
35 the link to the development in population (and text production), as Harald Baayen observes.  
36 However, this link is only implicit in the DTA corpus anyway: it strives neither for complete  
37 coverage of available texts, nor is the number of texts in the corpus proportional to  
38 population growth or to the number of published texts. For this reason, I will use the  
39 methodology sketched above and limit each decade to a fixed target size. The conceptual  
40 problem remains, of course – I simply cannot address it with the data at hand.

41  
42  
43

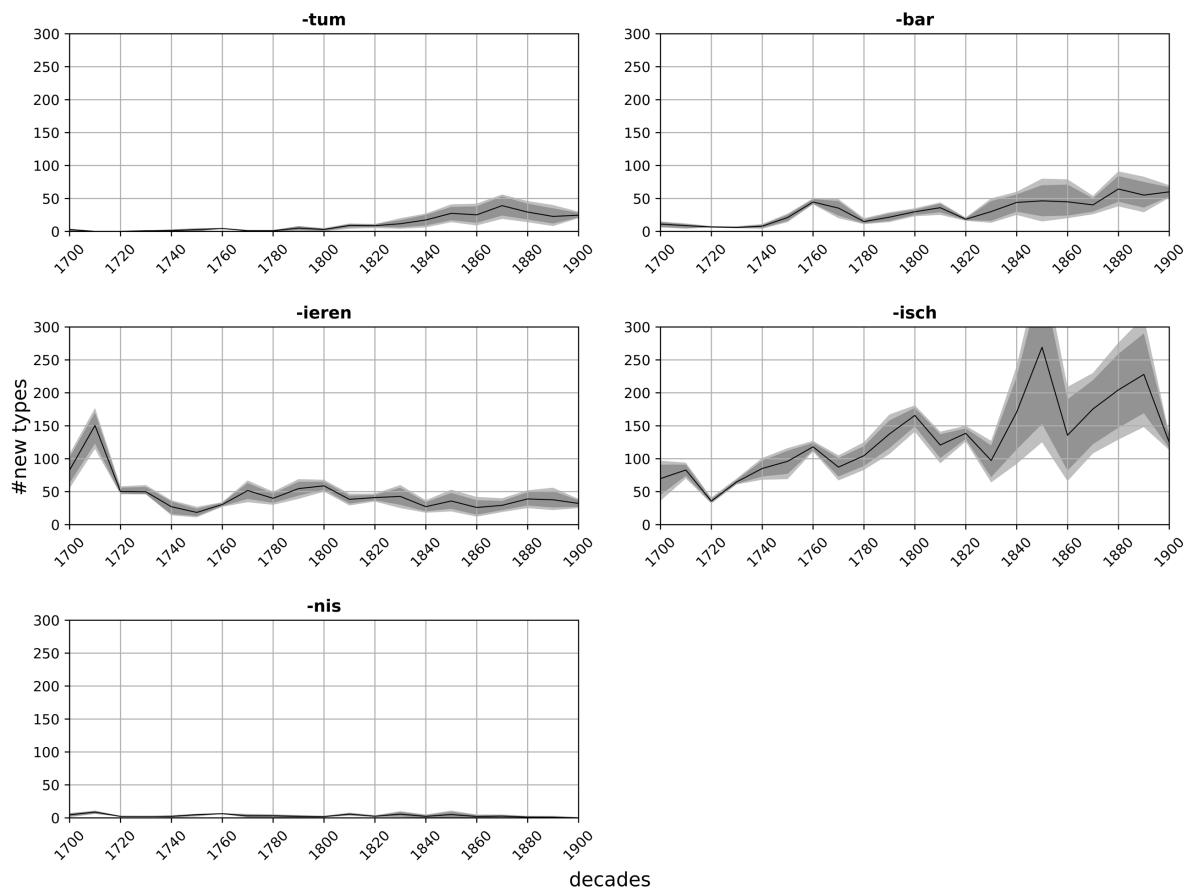
---

<sup>8</sup> This is, as Harald Baayen rightly observes, a cumulative societal perspective on the lexicon. For other perspectives, other measures might be more appropriate. For example, to describe the knowledge that individual speakers have of a word formation pattern, the type count may be more suitable: neologisms tend to be rare and – depending on the amount a person reads – may not ‘reach’ that person at all.

1    5. Case study: German derivational suffixes

2  
3    Let us apply the measures developed in the last section to the five word formation patterns  
4    in the 4 million token versions of the DTA corpus<sup>9</sup>. As mentioned above, we are not dealing  
5    with one version of the corpus, but with 100,000 randomly sampled corpora. As a  
6    consequence, we can compute a mean value of new words, but we can also compute the  
7    probability distribution of new words: We can state the probability with which any new  
8    resampling of the corpus will fall above or below certain limits. These limits are given for the  
9     $p < 0.01$  level (light grey area) and the  $p < 0.05$  level (dark grey area) in the following plots.  
10   Figure 5 shows the  $V_{neo}$  values for the five suffixes:

11



12

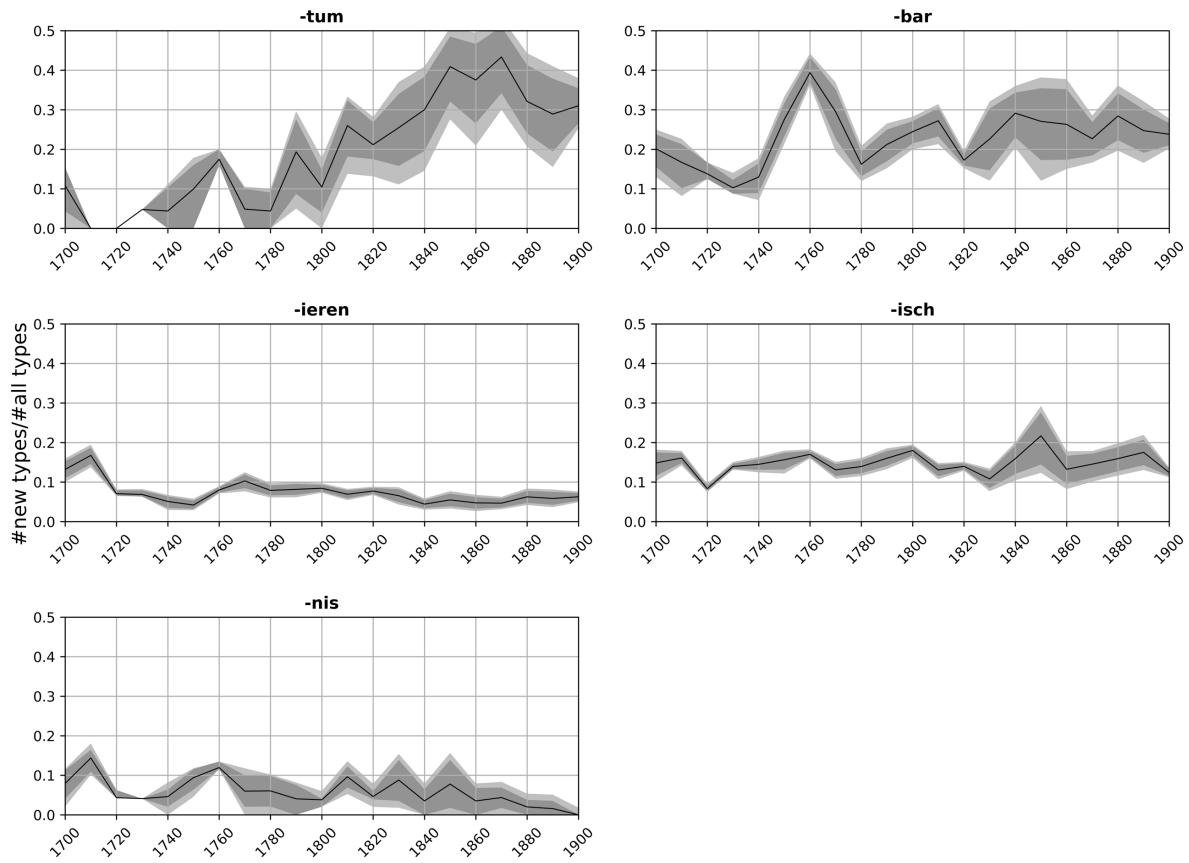
13   Figure 5:  $V_{neo}$  values for five German word formation patterns (-tum, -bar, -ieren, isch, and -nis); light grey area:  
14   99% confidence intervals, dark grey area: 95% confidence intervals. Data base: 100,000 times randomly  
15   resampled 4 million token version of the corpus Deutsches Textarchiv (DTA).

16

17   The pattern -isch produces the highest number of new words per decade, and the amount is  
18   slightly rising. The patterns -ieren and -bar yield less new words, but they are in the same  
19   range, with -bar producing slightly more words, and the pattern also rising. The number of  
20   new -tum words increases as well, but the numbers are well below those for the other  
21   suffixes, except for -nis: This seems to be a truly unproductive pattern in the period  
22   investigated. It would be premature, though, to reduce productivity to the  $V_{neo}$  values: After  
23   all, there are many more -isch types (not just the new ones) than e.g. -tum types. We should  
24   thus take the size of the morphological category into account, as explained above. Doing this  
25   yields the following  $P_{neo}$  values:

<sup>9</sup> The dataset and the Python scripts are available at <https://github.com/kristian-berg/Linguistics>.

1



2

3 *Figure 6:  $P_{neo}$  values for five German word formation patterns (-tum, -bar, -ieren, isch, and -nis); light grey area:*  
4 *99% confidence intervals, dark grey area: 95% confidence intervals. Data base: 100,000 times randomly*  
5 *resampled 4 million token version of the corpus Deutsches Textarchiv (DTA).*

6

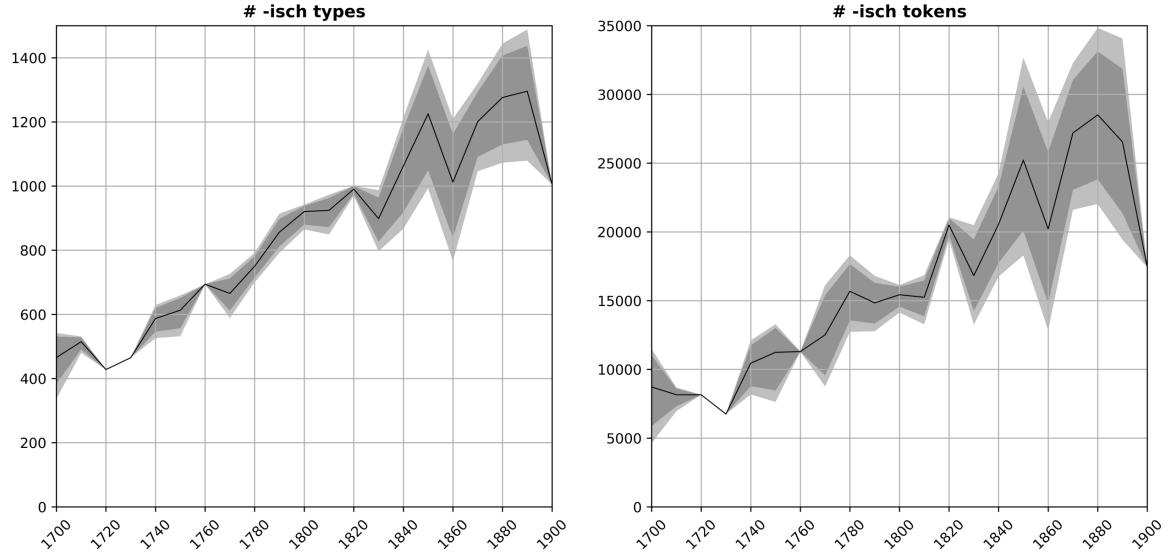
7 One of the most visible differences between the plots is in the confidence intervals: For **-isch**  
8 and **-ieren**, they are comparatively small, while they are relatively large for e.g. **-tum**. This  
9 mirrors the difference in the number of types (and new types) between the patterns in the  
10 raw (i.e., unsampled) data: While there are, for example, 947 **-isch** types in 1800 (153 of  
11 which are new), there are only 28 **-tum** types (two of which are new). The probability for  
12 variation due to different samples is thus much higher for patterns with a lower type count.  
13 These results have to be handled with care; however, we know in which area the next  
14 sample will fall with a 95% or a 99% chance.

15

16 Turning back to the plots, we can observe the following: The adjective suffix **-isch** has the  
17 steadiest distribution. The  $P_{neo}$  value is constant around 0.1 for the 19th century, with  
18 slightly higher values in the 18th century (this corroborates Kempf's findings about the  
19 productivity of **-isch** on the basis of type counts, token counts, hapaxes and neologisms, cf.  
20 Kempf 2016:248ff.). That means this word formation pattern had a constant renewal rate of  
21 around 10% for a long time: In each decade, every tenth word with this suffix was new.

22

23 As an effect of this constant influx of new words, both the type and token counts of this  
24 pattern rise over time; the pattern is expanding because the productivity is stable at a  
25 certain non-negligible level. Figure 7 shows the number of **-isch** types and tokens over time:

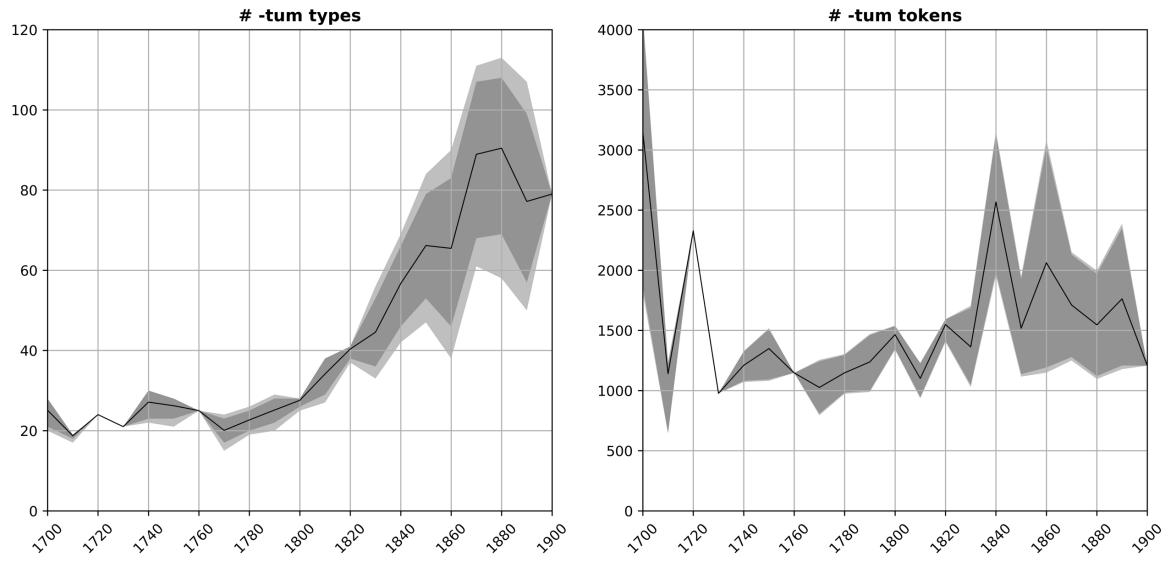


1  
2 *Figure 7: Number of types (left panel) and tokens (right panel) with -isch over time; light grey area: 99%*  
3 *confidence intervals, dark grey area: 95% confidence intervals.*

4  
5 As is visible, an apparently moderate renewal rate of 10% is mirrored as a considerable and  
6 almost linear increase in both type and token frequency for -isch over time. A constant  $P_{neo}$   
7 value leads to a constant rise in types and tokens. It would be interesting to see whether this  
8 generalization holds beyond the data presented here.

9  
10 The second pattern in figure 6 above, -ieren, is also constant, but at a lower level of around  
11 0.05–0.1. In contrast to that, the patterns -bar and -tum both show rising  $P_{neo}$  values over  
12 the course of the 19th century.  $P_{neo}$  for -bar even reaches the 0.3–0.4 mark, which means  
13 that roughly a third of all types with this suffix in the respective decades were new. Of the  
14 patterns investigated here, -tum and -bar are the most productive ones in the 19th century.

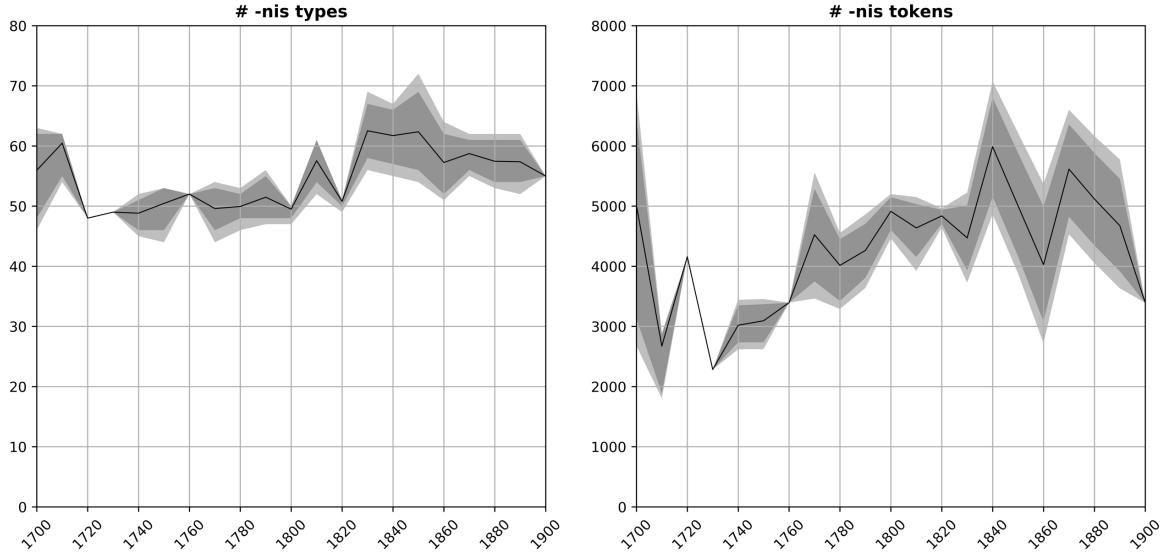
15  
16 For -tum, the rise of productivity in the 19th century is mirrored in a rapid increase in -tum  
17 types in the 19th century: The lexical richness triples between 1800 and 1890 (figure 8, left  
18 panel).



1  
2 *Figure 8: Number of types (left panel) and tokens (right panel) with -tum over time; light grey area: 99%*  
3 *confidence intervals, dark grey area: 95% confidence intervals.*

4  
5 This effect is not mirrored in the token count, however (figure 8, right panel), which –  
6 despite occasional peaks – rises only slightly. One explanation for this distribution is that a  
7 lot of the new words are rare and carry no weight compared to the established, high-  
8 frequency types. For example, in 1800–1809, the three most frequent types *Reichtum*  
9 ‘riches’, *Eigentum* ‘property’, and *Irrtum* ‘error’ account for more than half of the total -tum  
10 tokens of that decade (860 of 1,372 tokens, 63%). The situation has changed only slightly by  
11 1890, when the three most frequent types, *Eigentum*, *Christentum* ‘christendom’, and  
12 *Reichtum* account for less than half the tokens (2,094 of 4,935, 42%). As a contrast, the three  
13 most frequent -isch types in 1800 (*politisch* ‘political’, *französisch* ‘French’, and *römisch*  
14 ‘Roman’) account for only around 9% of all tokens in that decade (1,411 of 16,589). The -tum  
15 pattern is thus much more dominated by high frequency types than the -isch pattern.

16  
17 The last word formation pattern in figure 6, -nis, is only marginally productive. In five  
18 decades between 1700 and 1890, the productivity rises to around 0.1, but in the remaining  
19 decades, the productivity ranges between 0 and 0.05, which means that at least 95% of the  
20 types in the corresponding decades are old. What is more, there are only comparatively few  
21 types overall: In any given decade, there are hardly more than 60 types (figure 9, left plot):



1  
2 *Figure 9: Number of types (left panel) and tokens (right panel) with -nis over time; light grey area: 99%*  
3 *confidence intervals, dark grey area: 95% confidence intervals.*

4  
5 However, as witnessed in the P<sub>neo</sub> plot above, the pattern is not totally unproductive; the  
6 vocabulary increases, albeit slowly compared to the other patterns. The token count  
7 increases even more slowly (figure 9, right plot). It is rather high considering the limited  
8 number of types: This is due to a number of high frequency types; for example, the three  
9 most frequent -nis words in 1800, *Verhältnis* ‘relationship’, *Kenntnis* ‘knowledge’, and  
10 *Bedürfnis*, ‘need’ account for 59% of the total token count in that decade (3,106 out of 5,287  
11 tokens). The fact that there are new words with -nis at all is surprising given both measures,  
12 the small vocabulary size and the relatively high token count. In cognitive models of word  
13 formation, a large vocabulary size is often seen as a facilitating factor for productivity, and  
14 the same holds for a large number of low-frequency words (cf. e.g. Taylor 2002); both are  
15 not the case for -nis. The growth of the pattern is even more surprising when we look at the  
16 frequency of the new formations: Many of them become lexicalized or institutionalized, i.e.,  
17 they are in a sense very successful word formations. This may be the key to understanding  
18 the marginal productivity of this pattern: There was an onomastic demand. I will come back  
19 to this point shortly.

20  
21 One of the advantages of the measure used here is that we can take the set of neologisms in  
22 each time period as a starting point for a qualitative analysis, much like e.g. Kempf (2016) or  
23 Hartmann (2016). Changes in productivity can be caused by changes in onomastic demand  
24 (cf. e.g. Scherer 2007: 268 for German -er); sometimes, the productivity of a pattern changes  
25 with the productivity of a competing pattern (cf. e.g. Lindsay & Aronoff 2013 for English -ity  
26 vs. -ment); sometimes a pattern seems to fall out of fashion (cf. e.g. Berg submitted for  
27 German -nis). In all of these cases, an analysis based on the set of neologisms per time  
28 period can shed more light on the language-internal (competition) or external (onomastic  
29 demand, changing fashions) factors. Changes in productivity also sometimes go hand in hand  
30 with changes in the input restrictions of a word formation pattern. As a pattern becomes  
31 more productive, the set of possible bases widens. Turned into a deterministic rule  
32 (productivity as inversely proportional to the amount of input restrictions, e.g. Booij 1977:  
33 5), this is clearly wrong: There is a variety of reasons why patterns become productive or

1 stop being productive. Still, both may be related, and a qualitative analysis can uncover this  
2 relationship.  
3 In the following, I will sketch such a qualitative analysis that focuses on input restrictions for  
4 the suffix *-tum*. I will use all neologisms from the corpus, not just the neologisms from the  
5 4,000,000 token subcorpora. The reason is that we would have to find a way to generalize  
6 over the potentially different sets of first attestations in the 100,000 iterations, which is not  
7 trivial. Using all neologisms with their first attestations in the larger corpus is unproblematic  
8 as long as we abstain from making quantitative statements.

9  
10 34 types are first attested before 1700. If we sort them according to their respective bases,  
11 we arrive at the following set of rather diverse categories:

- 12  
13 (2) a. Heiligtum, Eigentum, Weistum  
14 'sanctuary', 'wealth', 'property', 'a (wise) legal decision'  
15 b. Altertum, Besitztum, Beweistum  
16 'antiquity', 'possession', 'evidence'  
17 c. Herzogtum, Erzherzogtum, Burggraftum  
18 'duchy', 'archduchy', 'burggravedom'  
19 d. Priestertum, Meistertum, Bürgermeistertum  
20 'priesthood', 'mastery', 'mayordom'  
21 e. Christentum, Judentum, Heidentum  
22 'christendom', 'Jewry', 'heathendom'

23  
24 The words in (2a) are derivations on an adjectival basis (*heilig*, 'holy', etc.); all other words  
25 have nominal bases. The bases of the words in (2b) are inanimate nouns (e.g. *Besitz*  
26 'possession')<sup>10</sup>, the ones in (2c–2e) are animate, and more precisely, they denote persons or  
27 groups of persons. The words in (2c) are formations on titles of nobility (e.g. *Herzog* 'duke');  
28 closely related to those are the words in (2d), which have occupational titles as bases (e.g.  
29 *Bürgermeister* 'mayor'). The bases in (2e), on the other hand, denote followers of belief  
30 systems (e.g. *Christ* 'Christian'). Only two words do not fit well into those categories:  
31 *Menschentum* 'humankind' has as its base the most generic term for humans, while  
32 *Luthertum* 'Lutherdom' has a very specific base, namely one person, Martin Luther. All in all,  
33 the set of words used before 1700 is rather limited, but its input restrictions are very  
34 heterogeneous (adjectival and inanimate as well as animate nominal bases).

35  
36 In the first half of the 18th century, only ten new types are attested. All but one (*Martertum*  
37 'affliction') have bases that denote person nouns. In most words, this base is a title of  
38 nobility (e.g. *Markgrafentum*, 'markgravedom'), or an occupational title (e.g.  
39 *Großmeistertum*, 'grandmasterdom'). The base of one word denotes a follower of a belief  
40 system (*Mohammedanertum* 'Islam'). The bases of the two remaining words are ethnonyms  
41 (e.g. *Türkentum* 'Turkdom'). This is a new type of formation.

42  
43 The second half of the 18th century also sees an expansion of input restrictions. While most  
44 of the 18 newly attested formations have bases that are either titles of nobility (e.g.

---

<sup>10</sup> There are two words that would synchronically be analyzed as derivations from a verbal base, *Wachstum* 'growth' on the verbal stem *wachs-* 'to grow' and *Irrtum* 'error' on the verbal stem *irr-* 'to err'. Diachronically, these words seem to be derived from nominal bases (cf. e.g. Tschentscher 1956).

1    *Königtum* ‘kingdom’<sup>11</sup>) or occupational titles (e.g. *Studententum* ‘studenthood’), there are  
2    three exceptions to this, and they all involve an evaluation of the person in question, either  
3    positive (*Heldentum* ‘heroism’), or negative (e.g. *Pfaffentum* ‘priesthood (pejorative)’).

4  
5    The input restrictions are again expanded in the next half-century 1800–1850. During this  
6    time, there are 70 new words. Bases that denote titles of nobility or occupational titles are  
7    attested, but they are in the minority (e.g. *Junkertum* ‘junkerdom’ or *Beamtentum*  
8    ‘officialdom’). The evaluative pattern is also attested with seven new words (e.g.  
9    *Virtuosentum* ‘virtuosity’), as is the pattern with followers of (political) belief systems  
10   *(Jakobinertum* ‘Jacobinism’). The largest subpattern with 15 types are ethnonyms (e.g.  
11   *Römertum* ‘Romanhood (i.e. Roman times)’, *Persertum* ‘the Persian people’). There are also  
12   a number of formations with inanimate bases (e.g. *Schrifttum* ‘the set of all writing (i.e.  
13   literature)’, *Kirchentum* ‘churchdom’) or adjectival bases (e.g. *Deutschtum* ‘Germanhood’). In  
14   this period, three new types of base are attested. The most prominent among them are  
15   bases that denote social status or class (e.g. *Bürgertum* ‘bourgeoisie’ or *Vasallentum*  
16   ‘vasalldom’). A second group consists of (nominal) bases that refer to characteristic features  
17   of the person other than their origin, social status or occupation, e.g. *Jungfrauentum*  
18   ‘virgindom’. And finally, there is also one type with a base that refers to an animal  
19   (*Bienentum* ‘beedom’).

20  
21   In the latter half of the 19th century, there are 241 new words; this is clearly the most  
22   productive phase of the pattern in the DTA corpus (cf. figure 6). Among the formations, we  
23   find the categories introduced above: Some bases are occupational titles (e.g.  
24   *Unternehmertum*, ‘entrepreneurship’); some of the bases denote a social class (*Arbeitertum*,  
25   ‘the collectivity of all workers’); others are ethnonyms (*Japanertum*, *Europäertum*); some are  
26   evaluative (*Idiotentum*, ‘idiocy’); others refer to characteristic traits of the person, e.g.  
27   *Waisentum* ‘orphandom’); some are titles of nobility (*Zarentum* ‘tsardom’), others evaluate  
28   their referent (*Maulheldentum* ‘the quality of being a loud-mouth’); some denote followers  
29   of belief systems (e.g. *Republikanertum* ‘republicanism’), yet others are inanimate  
30   (*Phrasentum*, *Episodentum*), and some denote animals (*Löwentum* ‘lionhood’). But again,  
31   the set of possible bases expands in this time period. There are four bases that are personal  
32   names (e.g. *Wagnertum* ‘Wagnerism’), and there is one base that is a cardinal number (von  
33   Humboldt’s *Fünftum* ‘fiveness’).

34  
35   Summing up, the number of input restrictions decreases over time, and the pattern expands  
36   to other types of bases. Table 1 summarizes the findings sketched above (shaded cells  
37   indicate that a pattern is attested):<sup>12</sup>

38  
39  
40  
41  
42  
43  
44  
45

---

<sup>11</sup> This is not a corpus artefact – the word, even though it appears rather basic today, was indeed coined as late as the second half of the eighteenth century, cf. Tschentscher (1956).

<sup>12</sup> The category “cardinal number” is omitted because it consists of just one type.

	pre 1700	1700–50	1750–1800	1800–1850	1850–1900
adjective					
inanimate noun					
title of nobility					
occupational title					
follower					
proper name					
ethnonym					
evaluation					
social class					
feature					
animal					

7 *Table 1: Summary of the input restrictions for -tum formations over time.*

8  
9 In the case of *-tum*, both  $P_{neo}$  (figure 6) and the number of types (figure 8) increase in the  
10 19<sup>th</sup> century. This increase is mirrored by an expansion of possible bases in the same time  
11 period. Why is this interesting? Because it does not have to be this way. As long as the initial  
12 categories are large enough, an increase in productivity can theoretically happen within  
13 these initial categories. Yet it is the case here, and it would be interesting to know how often  
14 it is the case for other patterns that expand (or shrink). Once enough quantitative and  
15 qualitative analyses of the kind presented here have been conducted, we can start drawing  
16 generalizations which in turn can inform our theoretical models of productivity.

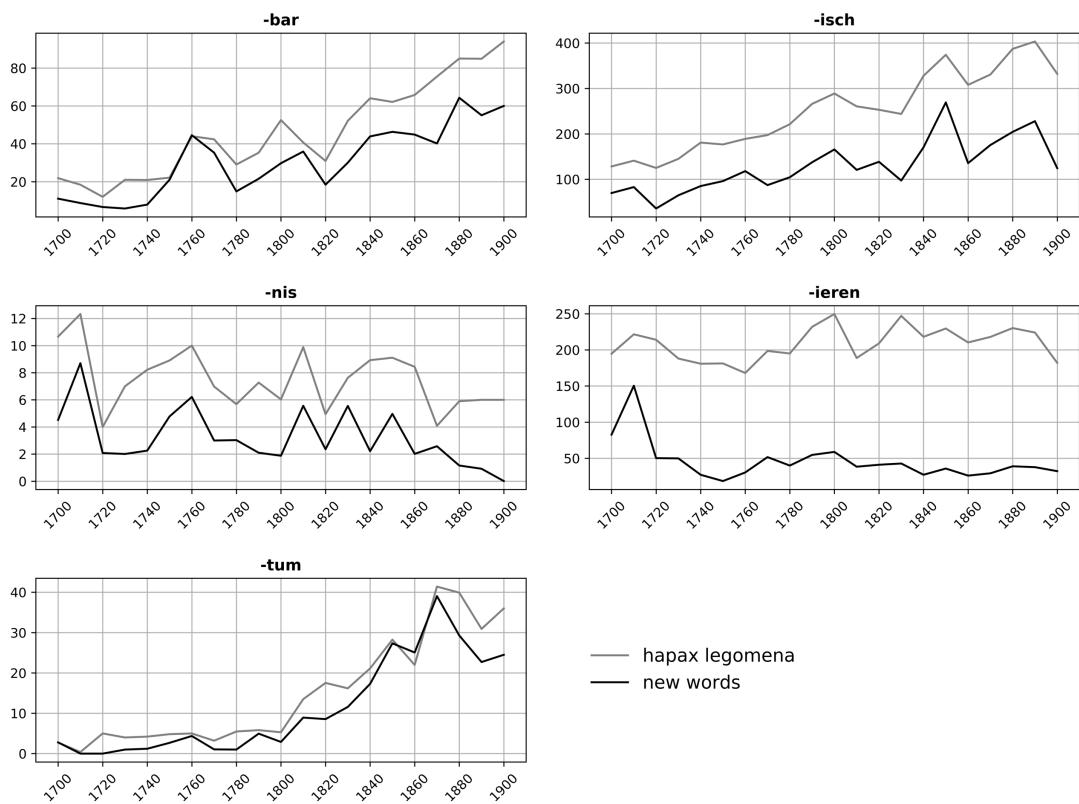
17  
18 Qualitative analyses need not stop here; we could also investigate the linking element or the  
19 semantic function of *-tum*. There has obviously been a high degree of onomastic demand in  
20 19th century Germany for those very abstract terms; it would be interesting to link this  
21 morphological finding to cultural studies of the time.

22  
23 6. Evaluation:  $P_{neo}$  vs.  $\mathcal{P}$

24 Let us take a moment to evaluate the productivity measure introduced here. The advantage  
25 of  $P_{neo}$  is that it can be interpreted as the renewal rate of the morphological category. It  
26 identifies how big a fraction of the vocabulary is new in any given time span. As such, it is a  
27 more direct measure compared with Baayen's  $\mathcal{P}$ .  $\mathcal{P}$  is defined as the quotient of the number  
28 of hapaxes with a given word formation pattern and the total number of tokens with the  
29 pattern ( $\mathcal{P} = \frac{n_1}{N}$ ). The idea behind this measure is that productive patterns produce many  
30 low frequency types; the higher the amount of particular low-frequency types, namely hapax  
31 legomena, in relation to all tokens of the pattern, the higher its productivity. As mentioned  
32 above, this is an elegant solution to the problem that the newness of words is not a directly  
33 accessible category in synchronic corpora:  $\mathcal{P}$  is the probability that after sampling N tokens,  
34 the next token will be a hapax. However,  $\mathcal{P}$  as a statistical estimate is problematic because it  
35 is derived from a simple urn model. In this model, we sample from a fixed and unchanging

1 population. Yet words are not evenly distributed within and across texts (I would like to  
2 thank Harald Baayen for this point).  
3  
4 What is more, this approach is unnecessarily complex when we are dealing with diachronic  
5 data. As an analogy, consider marine biology. Using data about big fish (heavier than one  
6 kilogram), it is possible to predict the number of species below this weight – e.g. 431 species  
7 between 1 and 10 grams in the North Sea (Reuman et al. 2014). Now imagine we knew the  
8 whereabouts of every fish in a confined area of the North Sea; would we still use the indirect  
9 measure? We would of course use the actual data. This analogy is not as far off the mark as  
10 it might seem. We have a sophisticated model that allows us to make an educated guess  
11 about unknown properties (Baayen's  $\mathcal{P}$ ). But once we have more direct access to these  
12 properties, the model is superfluous.

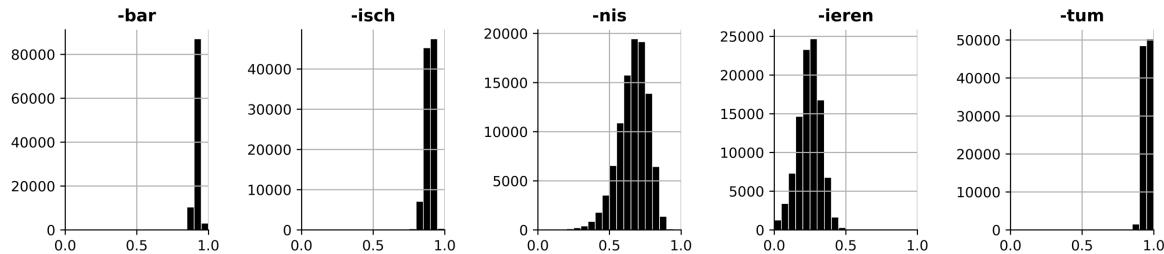
13  
14 What we can do, though, is use the actual data to test the goodness of the assumption  
15 behind  $\mathcal{P}$ . As hinted at above, the core idea is that hapaxes are indicative of neologisms. To  
16 test this postulate, we simply go through every suffix and every decade and determine a) the  
17 number of neologisms and b) the number of hapaxes in that decade. Figure 10 plots the  
18 means values (black: new words; grey: hapax legomena) for each suffix:



19  
20 *Figure 10: The number of neologisms (black) and the number of hapaxes (grey) per pattern per decade for the*  
21 *five derivational suffixes (-bar, -isch, -nis, -ieren, and tum).*

22  
23 There clearly is a correlation between both measures. This correlation is very close for some  
24 patterns (e.g. **-tum**), and rather loose for others (e.g. **-ieren**). Of course, the values plotted in  
25 figure 10 are the mean values over all the 100.000 simulations; they only allow for an  
26 overview of the data. But in each simulation, we have a time series of values for new words  
27 and one for hapaxes, and we can measure the correlation between both. That means we

1 have 100.000 correlation coefficients for each suffix. These values are plotted as histograms  
2 in figure 11:

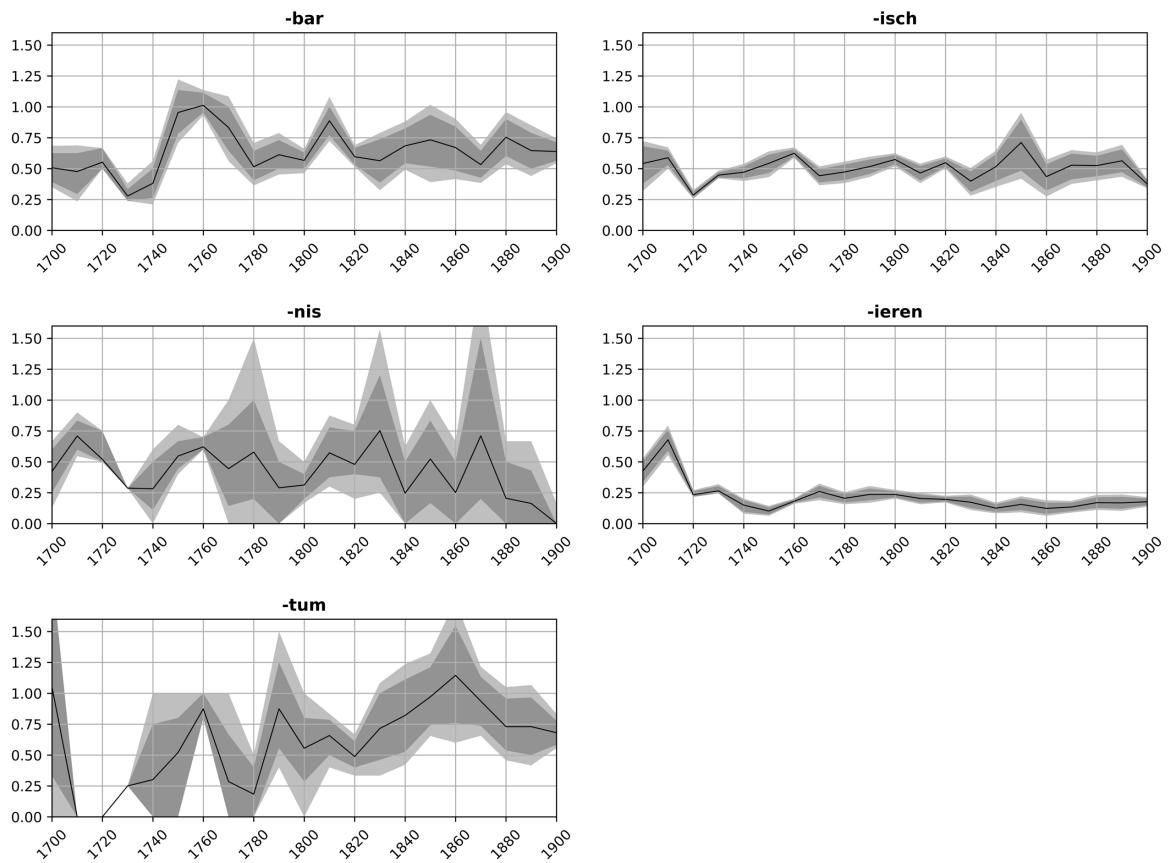


3  
4 *Figure 11: Histograms of correlation coefficients (Pearson's r) between hapaxes and new words for the 100,000*  
5 *simulations and the five derivational suffixes (-bar, -isch, -nis, -ieren, and tum).*

6  
7 For *-bar*, *-isch*, and *-tum*, the correlation is rather strong (it is consistently above 0.7),  
8 with *-tum* showing the highest degree of correlation in the simulations. The correlations  
9 for *-nis* are lower, but most of them are still above 0.5. The pattern with the consistently  
10 weakest correlations is *-ieren*. For some patterns, we can conclude, the number of hapaxes  
11 correlates closely with the number of new words – for some, this is much less the case.

12  
13 Yet even in cases where the correlation coefficients are consistently high, the actual values  
14 can differ considerably. Consider *-tum* and *-isch* in figure 10 above: For both patterns,  
15 hapaxes and new words correlate very closely (cf. Fig. 11), but the ratio between both values  
16 is very different. While the numbers of hapaxes and new words are roughly equal for *-tum*,  
17 the number of hapaxes is around twice the number of new words for *-isch*.

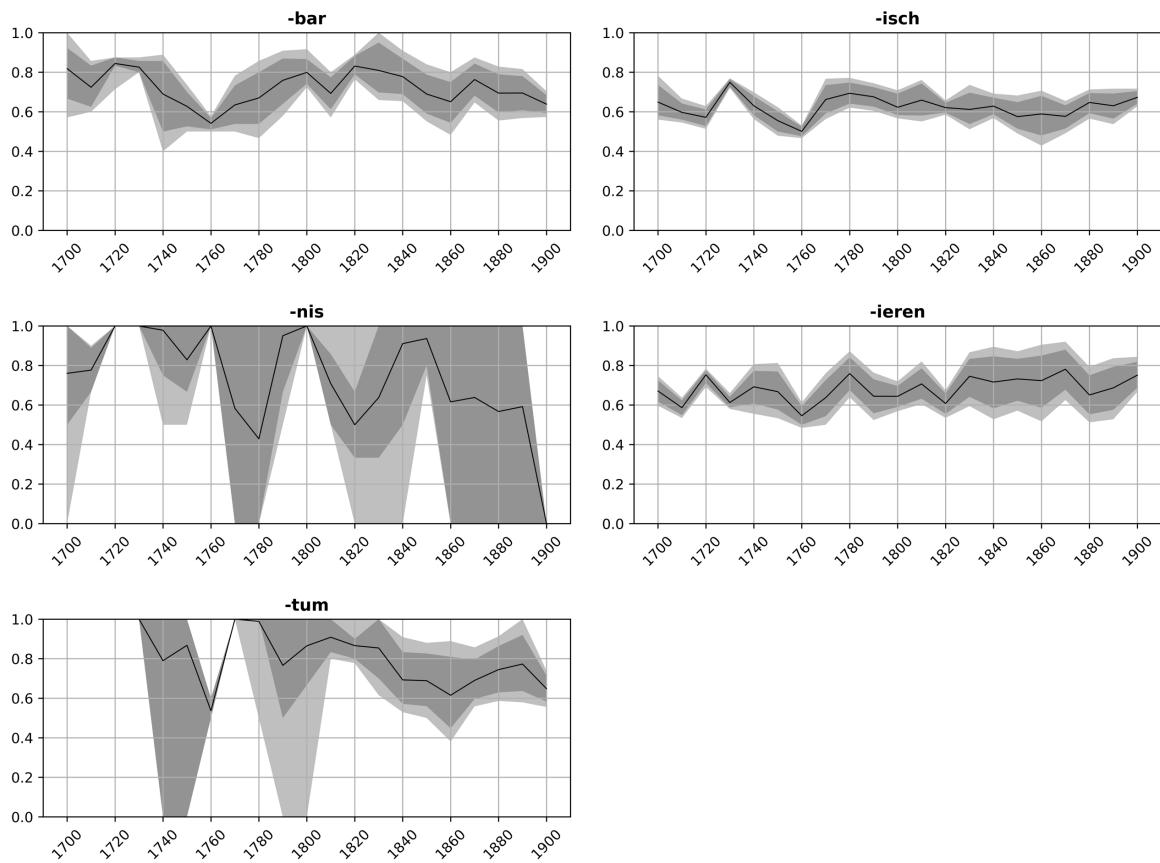
18  
19 This holds for the other patterns as well. Figure 12 plots the quotient of both ( $\frac{\text{neologisms}}{\text{hapaxes}}$ ) for  
20 each of the suffixes:  
21



1  
2 *Figure 12: The number of neologisms divided by the number of hapaxes per pattern per decade for the five*  
3 *derivational suffixes (-bar, -isch, -nis, -ieren, and -tum); light grey area: 99% confidence intervals, dark grey*  
4 *area: 95% confidence intervals.*

5  
6 The ratio of hapaxes and neologisms is remarkably stable over time for some of the patterns  
7 (e.g. **-isch** and **-ieren**). For **-tum**, however, and across suffixes, it varies considerably, from a  
8 maximum of 1 (**-tum**, 1850) to 0.13 (**-ieren**, 1750). If hapaxes are indicative of neologisms,  
9 they are so in a rather vague way.

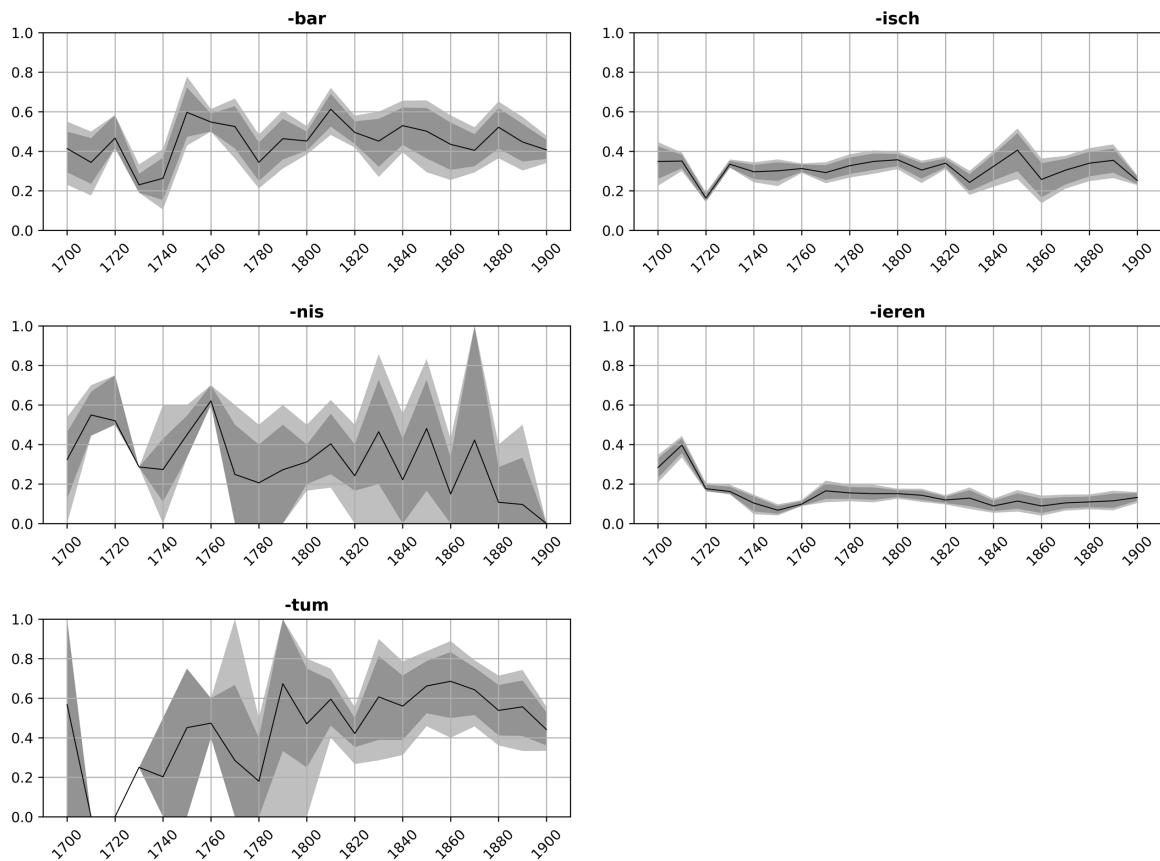
10  
11 Let us turn to a related but more specific aspect of the relation between hapaxes and  
12 neologisms. The basic assumption underlying  $\mathcal{P}$  is that neologisms are more likely to be  
13 found among lower frequency words. With the data at hand, we can actually determine the  
14 rate of hapaxes among the neologisms for any given suffix and time span (Figure 13).



1  
2 *Figure 13: The ratio of new hapaxes among all new words per pattern per decade for the five derivational*  
3 *suffixes (-bar, -isch, -nis, -ieren, and -tum); light grey area: 99% confidence intervals, dark grey area: 95%*  
4 *confidence intervals.*

5  
6 Most of the new words from each pattern are hapaxes. The ratio is relatively stable across  
7 patterns and time periods; it lies between 0.6 and 0.8 (the marginally productive pattern *-nis*  
8 is the only exception). This is in line with Baayen & Renouf's (1996: 75) observation that  
9 "neologisms [...] typically occur among the lowest frequency types". Thus if we are looking  
10 for new words, and we want to use frequency classes of words as a proxy, then hapaxes are  
11 the best choice (if 60% to 80% of all new words are hapaxes, there is not much room for *dis*  
12 *legomena* etc.).

13  
14 From this does not follow, however, that hapaxes are a particularly *good* proxy for new  
15 words. There may be more or less hapaxes for each pattern that are not new but simply  
16 rare. This is the case in the DTA corpus, as figure 14 shows:



1  
2 *Figure 14: The ratio of new hapaxes among all hapaxes per pattern per decade for the five derivational suffixes*  
3 *(-bar, -isch, -nis, -ieren, and -tum); light grey area: 99% confidence intervals, dark grey area: 95% confidence*  
4 *intervals.*

5  
6 Thus, while it is true for the data used here that most of the new words are hapaxes, the  
7 reverse does not hold for all patterns and all time periods: Among the hapaxes, the ratio of  
8 new words varies quite drastically, between 0.6 for *-tum* in the 19<sup>th</sup> century and below 0.2  
9 for *-ieren*. For this pattern, hapaxes are not very indicative of new words because most of  
10 the hapaxes (>80%) are simply rare but recurring words. Something similar holds for *-isch*,  
11 where the amount of new hapaxes lies below 40%, and for *-bar* and *-nis*, where it is below  
12 60%. What is more, while the amount is stable over time for some patterns (*-bar*, *-isch*, *-*  
13 *ieren*), it varies for others (*-nis* and *-tum*). These differences over time and across patterns  
14 are an obstacle to the comparability of different  $\mathcal{P}$  values.

15  
16 Note that this may be an effect of the corpus size: It is possible that the rare hapaxes  
17 for *-ieren* in Figure 14 would not be hapaxes in a larger corpus, leading to a higher ratio of  
18 new hapaxes among all hapaxes. But a change in corpus size would probably also affect the  
19 near-perfect correlation between hapaxes and new words for *-tum* (in Figure 10 above).  
20 Different corpus sizes can amplify or mute certain metrics of patterns, and the precise  
21 change also depends on the characteristics of the pattern itself. This is in some respects  
22 analogous to room modes in acoustics: Depending on the size, shape and surfaces of a room,  
23 certain frequencies are amplified. To understand (and predict) the interaction, we need a  
24 good model of sound waves and the way they are reflected or absorbed by certain surfaces.  
25 Likewise, to understand the interaction between word formation patterns, their metrics, and  
26 corpus size, we need a good model. To build this model, we need to investigate the

1 relationship between corpus size, hapaxes, types, tokens, and new words more thoroughly.  
2 Unfortunately, that is beyond the scope of this paper.  
3  
4 We can tentatively conclude that for different word formation patterns at different points in  
5 time, hapaxes vary with regard to the number of neologisms they are supposed to  
6 represent. Hapax-based measures should thus be used with caution in diachronic  
7 investigations. The  $P_{\text{neo}}$  measure is directly interpretable, and it measures productivity more  
8 directly, as I have just demonstrated. It is also more independent of the corpus size, again in  
9 contrast to Baayen's  $\mathcal{P}$ : After all,  $P_{\text{neo}}$  is type-based, and  $\mathcal{P}$  is token-based. Any change in the  
10 size of the sample will affect the token-based measure naturally more.<sup>13</sup>

11

## 12 7. Summary

13 In this paper, I have shown how a measure based on the notion of new words (which in turn  
14 are a central part of definitions of productivity) can be employed in investigations of  
15 diachronic productivity. I have presented a method that deals with the problem of varying  
16 sub-corpus sizes (randomly resample the sub-corpora up to a predefined size), and to the  
17 problem of old words appearing as new at the start of the corpus (take an earlier corpus and  
18 determine a point in time when almost all old words have registered).

19

20 The measures proposed in this paper are based on first occurrences of words in a large  
21 diachronic corpus. These measures are a more direct proxy to new words than hapax-based  
22 measures – they measures what we want to measure, which is not always true for hapax-  
23 based measures – and they are also more stable when sample sizes vary. The basic  
24 assumption of hapax-based measures (that hapaxes are indicative of new words) does not  
25 hold to the same degree for all patterns and all time periods; the correlation between new  
26 words in the corpus and hapaxes varies considerably.

27

28 Armed with the two measures based on new words,  $V_{\text{neo}}$  and  $P_{\text{neo}}$ , I tracked changes in the  
29 productivity of five German suffixes.  $V_{\text{neo}}$ , the number of new words, can be used to  
30 compare patterns with respect to their contribution to the expansion of the vocabulary:  
31 From this perspective, *-isch* plays a larger role than *-tum* because it contributes more new  
32 words to the overall lexicon of the language. But if we look at  $P_{\text{neo}}$ , which can be regarded as  
33 the renewal rate of the category, *-tum* is more productive than *-isch*, because relative to all  
34 types in the category, *-tum* produces more new words. It is this latter measure that is more  
35 directly linked to our intuitive notion of productivity, I would argue.

36

37 The large diachronic corpora that have become available lately provide us with a richness of  
38 data that I have not done justice to. Of all the conceivable parameters of a pattern, I have  
39 only used the first attestation of word formation products. Accordingly, the end of this paper  
40 is more of a starting line than a finishing line.<sup>14</sup> For example, we could take the last

---

<sup>13</sup> Of course, this is not to say that  $P_{\text{neo}}$  is independent of the sample size: the more words we sample per decade, the higher the likelihood of finding new words. That is why the number of tokens per decade should be reported (see formula (1) above). The assumption is simply that  $P_{\text{neo}}$  is *more stable* than  $\mathcal{P}$  if we vary the sample size.

<sup>14</sup> And if we think bigger, we can also ask for the renewal rates of larger categories, or of the vocabulary as a whole. Where do new words come from? What relative amounts do derivation, compounding, conversion, loan words, acronyms, etc. have? A prerequisite for this kind of investigation is a manually cleaned corpus, though, which is almost impossible considering the corpus size necessary for the investigation.

1 attestation into account and reconstruct the life-span of word-formation products (cf.  
2 Neuhaus 1971, 1973).  
3  
4 We could also think of a more sophisticated model of what it means for a word to be ‘alive’.  
5 We could investigate the speed, or the burstiness, with which new words spread through  
6 texts. Although not immediately relevant to productivity, this is connected to an interesting  
7 question: What makes a new word successful? What are the (structural, pragmatic, social)  
8 conditions for it to become lexicalized and used extensively (cf. e.g. Kerremans et al. 2012)?  
9

10 References

- 11 Anderson, Karen Elizabeth (2000). *Productivity in English nominal and adjectival derivation, 1100–2000*. Ph.D. thesis. University of Western Australia.
- 12 Anshen, Frank & Mark Aronoff (1989). Morphological productivity, word frequency and the Oxford English Dictionary. In Fasold, Ralph. W. & Deborah Schiffrin (eds), *Language Change and Variation*. Amsterdam: Benjamins, p. 197–202.
- 13 Anshen, Frank & Mark Aronoff (1999). Using Dictionaries to Study the Mental Lexicon. *Brain and Language* 68, p. 16–26
- 14 Aronoff, Mark (1976). *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.
- 15 Aronoff, Mark (1983). Potential words, actual words, productivity and frequency.  
Proceedings of the 13th international congress of linguists, p. 163–171
- 16 Baayen, R. Harald (1989). A corpus-based approach to morphological productivity. *Statistical Analysis and Psycholinguistic Interpretation*. PhD thesis, Free University, Amsterdam.
- 17 Baayen, R. Harald (1992). Quantitative aspects of morphological productivity. In Booij, Geerd & Jaap van Marle (eds), *Yearbook of Morphology 1991*. Dordrecht: Kluwer Academic Publishers, p. 109–149.
- 18 Baayen, R. Harald (1993). On frequency, transparency, and productivity. In Booij, Geerd & Jaap van Marle (eds), *Yearbook of Morphology 1992*. Dordrecht: Kluwer Academic Publishers, p. 181–208.
- 19 Baayen, R. Harald (2001). *Word Frequency Distributions*. Kluwer Academic Publishers, Dordrecht.
- 20 Baayen, R. Harald (2009). Corpus linguistics in morphology: morphological productivity. In Lüdeling, Anke & Merja Kytö (eds.) *Corpus Linguistics. An international handbook*. Berlin: de Gruyter, p. 900–919
- 21 Baayen, R. Harald & Rochelle Lieber (1991). Productivity and English derivation: a corpus-based study. *Linguistics* 29.5, p. 801–843.
- 22 Baayen, R. Harald & Renouf, Antoinette (1996). Chronicling The Times: Productive Lexical Innovations in an English Newspaper. *Language*, 72, p. 69–96.
- 23 Baroni, Marco & Stefan Evert (2005). Testing the extrapolation quality of word frequency models. In Danielsson, Pernilla & Martijn Wagenmakers (eds.), *Proceedings of Corpus Linguistics 2005* (<http://www.birmingham.ac.uk/research/activity/corpus/publications/conference-archives/2005-conf-e-journal.aspx>, accessed 15<sup>th</sup> October 2017)
- 24 Bauer, Laurie (2001). *Morphological Productivity*. Cambridge: Cambridge University Press.
- 25 Berg, Kristian (2017). *Die Morphologie der Schrift im Deutschen und Englischen*. Unpublished habilitation thesis, University of Oldenburg.
- 26 Berg, Kristian (submitted). Produktivität, Lexikalisierung und Rückbau. -nis im Neuhochdeutschen. Submitted to *Beiträge zur Geschichte der deutschen Sprache und Literatur*.

- 1 Bolinger, Dwight (1948). On defining the morpheme. *Word* 4, p. 18–23.
- 2 Bolozkym Shmuel (1999). *Measuring productivity in word formation: the case of Israeli Hebrew*. Leiden: Brill.
- 3 Booij, Geert (2012). *The Grammar of Words. An Introduction to Linguistic Morphology*. Oxford: Oxford University Press.
- 4 Cowie, Claire (1999). *Diachronic word-formation: A corpus-based study of derived nominalizations in the history of English*. Ph.D. thesis, University of Cambridge.
- 5 Cowie, Claire & Christiane Dalton-Puffer (2002). Diachronic Word-formation: theoretical and methodological considerations. In: Diaz Vera, Javier E. (ed.), *A Changing World of Words: Studies in English Historical Semantics and Lexis*. The Netherlands: Rodopi, p. 410–436.
- 6 Dal, Georgette (2003). Productivité morphologique: définitions et notions connexes. *Langue française* 140, p. 3–23.
- 7 Dalton-Puffer, Christiane. 1996. *The French influence on Middle English morphology: A corpus-based study of derivation*. Berlin: de Gruyter.
- 8 Fleischer, Wolfgang & Irmhild Barz (2012). *Wortbildung der deutschen Gegenwartssprache*. Berlin: de Gruyter.
- 9 Flury, Robert (1964): *Struktur und Bedeutungsgeschichte des Adjektiv-Suffixes -bar*. Winterthur: Keller.
- 10 Gaeta, Livio & Davide Ricca (2006). Productivity in Italian word formation: A variable- corpus approach. *Linguistics* 44.1, p. 57–89.
- 11 Geyken, Alexander, 2007: The DWDS corpus: A reference corpus for the German language of the 20th century. In: Fellbaum, Christiane (ed.): *Collocations and Idioms: Linguistic, lexicographic, and computational aspects*. London: Bloomsbury, S. 23–41.
- 12 Hartmann, Stefan (2016). *Wortbildungswandel. Eine diachrone Studie zu deutschen Nominalisierungsmustern*. Berlin: de Gruyter.
- 13 Hartmann, Stefan (2018). Derivational morphology in flux: A case study of word-formation change in German. *Cognitive Linguistics* 29(1), p. 77-119..
- 14 Hilpert, Martin (2013). *Constructional change in English: Developments in allomorphy, word formation, and syntax*. Cambridge: Cambridge University Press.
- 15 Kaplan, E.L. & Paul Meier (1958). Nonparametric Estimation from Incomplete Observations. *Journal of the American Statistical Association* 53, p. 457–481.
- 16 Kempf, Luise (2016). *Adjektivsuffixe in Konkurrenz. Wortbildungswandel vom Frühneuhochdeutschen zum Neuhochdeutschen*. Berlin: De Gruyter.
- 17 Kerremans, Daphné, Susanne Stegmayr & Hans-Jörg Schmid (2012). The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. In: Allan, Kathryn & Justyna Robinson, (eds.), *Current methods in historical semantics*. Berlin: de Gruyter, p. 59-96.
- 18 Lenders, Winfried & Klaus-Peter Wegner (1982): *Maschinelle Auswertung sprachhistorischer Quellen. Ein Bericht zur computerunterstützten Analyse der Flexionsmorphologie des Frühneuhochdeutschen*. Tübingen: Niemeyer.
- 19 Lindsay, Mark & Mark Aronoff (2013). Natural Selection in Self-Organizing Morphological Systems. In: Montermini, Fabio, Gilles Boyé, & Jesse Tseng (eds.) *Morphology in Toulouse: Selected Proceedings of Décembrettes* 7. Munich: Lincom Europa, p. 133–153.
- 20 Lohde, Michael (2006). *Wortbildung des modernen Deutschen: ein Lehr- und Übungsbuch*. Tübingen: Narr.
- 21 Motsch, Wolfgang (2004). *Deutsche Wortbildung in Grundzügen*. Berlin: de Gruyter.

- 1 Neuhaus, H. Joachim (1971). *Beschränkungen in der Grammatik der Wortableitungen im*  
2 *Englischen*. Ph.D. thesis, Universität Saarbrücken.
- 3 Neuhaus, H. Joachim (1973). Zur Theorie der Produktivität von Wortbildungssystemen. In  
4 Ten Cate, Abraham P. & Peter Jordens (eds.), *Linguistische Perspektiven: Referate des*  
5 *VII. Linguistischen Kolloquiums, Nijmegen, 26.-30. September 1972*. Tübingen:  
6 Niemeyer, p. 305–317.
- 7 Plag, Ingo (1999). *Morphological Productivity: Structural Constraints in English Derivation*.  
8 Berlin: de Gruyter.
- 9 Reuman, Daniel, Henrik Gislason, Carolyn Barnes, Frédéric Mélin, & Simon Jennings (2014).  
10 The marine diversity spectrum. *Journal of Animal Ecology* 83.4, p. 963–979.
- 11 Säily, Tanja (2016). Sociolinguistic variation in morphological productivity in eighteenth-  
12 century English. *Corpus Linguistics and Linguistic Theory* 12(1), p. 129–151.
- 13 Scherer, Carmen (2005). *Wortbildungswandel und Produktivität. Eine empirische Studie zur*  
14 *nominalen -er-Derivation im Deutschen*. Tübingen: Niemeyer.
- 15 Scherer, Carmen (2015). Change in productivity. In: Müller, Peter O., Ingeborg Ohnheiser,  
16 Susan Olsen, & Franz Rainer (eds.), *Word-Formation: An International Handbook of the*  
17 *Languages of Europe, Volume 3*. Berlin: de Gruyter, p. 1781–1793.
- 18 Schneider-Wiejowski (2011). *Produktivität in der deutschen Derivationsmorphologie*. Ph.D.  
19 thesis, Universität Bielefeld ([https://pub.uni-bielefeld.de/download/2473126/](https://pub.uni-bielefeld.de/download/2473126/2477032)  
20 2477032, last accessed 15 October 2017).
- 21 Schröder, Anne (2011). *On the Productivity of Verbal Prefixation in English: Synchronic and*  
22 *Diachronic Perspectives*. Tübingen: Narr.
- 23 Štichauer, P. (2009). Morphological productivity in diachrony: the case of the deverbal nouns  
24 in *-mento*, *-zione* and *-gione* in old Italian from the 13th to the 16th century. In  
25 Montermini, Fabio, Gilles Boyé, & Jesse Tseng (eds.), *Selected proceedings of the 6th*  
26 *Décembrettes*. Somerville: Cascadilla Proceedings Project, p. 138–147.
- 27 Taylor, John (2002). *Cognitive Grammar*. Oxford: Oxford University Press.
- 28 Tschentscher, Christhild (1956). Die Geschichte der Silbe “tum” im Deutschen.  
29 *Muttersprache* 72, p. 1–8, 39–47, 67–78.
- 30 Trips, Carola (2009). *Lexical Semantics and Diachronic Morphology: The Development of*  
31 *-hood, -dom and -ship in the History of English*. Tübingen: Niemeyer.