R. Köhler (ed.) (2009), *Issues in quantitative linguistics: 1-9.*
Lüdenscheid: RAM

Measuring Morphological Productivity

Olga Pustylnikov[1]
Karina Schneider-Wiejowski[2]

Abstract

The present article addresses the phenomenon of productivity in derivational morphology from the point of view of its quantification. Research in morphological productivity was mainly done for the English language, and there is still a lack for German, although, it is a very interesting subject. We evaluate three quantitative approaches proposed in the literature to measure productivity of German noun suffixes. In addition, we apply a decomposition algorithm used in a multi-agent simulation (Pustylnikov, 2009) to identify productive suffixes. We compare three different types of corpora in German: two newspaper corpora from different periods of time, and one corpus of spontaneous speech. Empirical studies exploring productivity mostly focus on written data, although new word formations are expected to occur to a larger extent in spoken language. In this article, we investigate the differences in results based on spoken and written corpora.

1 Introduction

Predominant approaches in the area of language simulations (e.g., Steels, 2005; Vogt, 2003) deal with random vocabularies, while focusing on the emergence of meaning-form relations or compositionality (Kirby, 2007). Other use game theoretic assumptions modeling human communication strategies (Jäger, 2008), but still with random language input. This is done in order to simulate the emergence of language controlling different parameters. However, we might be interested in evolution of language that already has achieved a particular state in development. (Pustylnikov, 2009) have presented a simulation model that analyzes natural language input (e.g., German) rather than random words, and takes this language as the base for communication between the agents. The simulation model was initially designed to examine the evolution of suffixes in word formation. That is, the use of a particular suffix when it comes to create a new word is assumed to depend on the language use

[1] Texttechnology Group, SFB 673, University Bielefeld.
E-mail: Olga.pustylnikov@techfak.uni-bielefeld.de
[2] Linguistics and Literature, SFB 673, University Bielefeld
E-mail: karina.wiejowski@uni-bielefeld.de

in general (Tomasello, 2005), and on the lexicon a person has acquired in particular (Bybee, 1988: chapter 7, 119–141.). In this paper, we use the language decomposition algorithm of this model to study productivity of suffixes.

Suffixes that have the same function (e.g., to derive an adjective from noun) are supposed to compete during the evolution of language. For example, a suffix that is preferred to derive an adjective from verb is likely to be reused in future word formations (e.g., *ease* > *eas-y*). These affixes are called productive affixes (Baayen, 1991).

A lot of work in respect to morphological productivity was done synchronically, so that productivity of several affixes was compared within a single period of time (see e.g., Prell, (1991), Habermann (1994) and Stricker (2000)). There was some effort to combine synchronic and diachronic aspects of productivity (Munske, 2002). (Bauer, 2001) emphasizes:

> "A second problem for word-formation with the distinction between synchrony and diachrony is that it is frequently the case that a diachronic event is the evidence for a synchronic state. (Bauer, 2001:27)

Thus, it is reasonable to consider both perspectives when we understand productivity as a gradual process, and aim to measure its development. (Cowie et al., 2002:418) also emphasize the dynamics in change of word formation as the main aspect, and speak of *diachronic productivity*.

In this article, we present a combined approach to exploring morphological productivity in German. We examine productivity *diachronically* within the register of newspaper texts comparing corpora of 17th-19th century German to a 20th newspaper corpus. Further, we consider a spoken corpus of German in order to test the divergence in productivity values for spoken and written data. Since new word formations are more expected to happen spontaneously in speech, rather than in written language we evaluate this difference here. The evaluation for synchronic and diachronic productivity is made using four quantitative measures: two introduced by Harald Baayen (Baayen, 1991; Baayen,1992), one proposed by Kreyer (2009), and one based on the simulation model discussed here.

The paper is organized as follows: Section 2 discusses the literature on morphological productivity. Section 3 describes the multi-agent simulation model allowing to study the productivity of a particular language. The corpora used are presented in Section 4. In Section 5, we describe quantitative measures applied here to measure productivity. The obtained results are discussed in Section 6. Finally, the conclusions are drawn in Section 7.

2 Related Work

The phenomenon of morphological productivity has long been discussed theoretically in linguistic literature (see e.g., Schultink (1961), Plag (1999). And it has long been discussed as an insolvable mystery of derivational morphology (see Aronoff (1976:36)). But, the idea of morphological productivity is older than Aronoffs work, and the first remark on an aspect that is important to describe this phenomenon is made by Willmanns (1899) who describes derivation types by *vitality* and *persistency*. The same idea is expressed by Kruisinga (1932:22) who speaks of *living suffixes* and so called *dead suffixes*. An important step and often quoted statement comes from Schultink (1961):

> "We see productivity as a morphological phenomenon as the possibility for language users to coin unintentionally an in principle unlimited number of new formations, by using the morphological procedure that lies behind the form-meaning correspondence of some known words." (In: Evert & Luedeling (2001:167))

Three important aspects are mentioned in this statement: unintentionality, unlimitedness and regularity. All criteria given by Schultink (1961) can be seen as interdependent criteria. Unintentionality is in opposition to creativity. Words constructed by creativity are easily recognized as new words, this cannot be said at all for words produced unintentionally. The aspect of unlimitedness is a general property of natural language, and the productivity of a language is supported by the use of some regular affixes. In contemporary German, for example, it is possible to create many new adjectives with the suffix -*mäßig*: bananen-mäßig (banana like), kaffee-mäßig (coffee like), and in principle almost all can be -*mäßig*. Thus, the aspect of regularity is a very important one for derivation and word formation.

Aronoff (1976) makes an interesting statement in respect to morphological productivity. He points out that it is necessary to develop a method that allows estimating the number of all *possible words* formed by a word formation rule (WFR), and not just the ones given in sampled text:

> "There is a simple way to take such restrictions into account: we count up the number of words which we feel could occur as the output of a given WFR (which we can do by counting the number of possible bases for that rule), count up the number of actually occurring words formed by that rule, take a ratio of the two, and compare this with the same ratio for another WFR. In fact, by this method we could arrive at a simple index of productivity for every WFR: the ratio of possible to actual words." (Aronoff (1976:36))

This is the point where quantitative approaches come into play. Mostly all ideas and definitions made by the mentioned authors above are aspects that can be summarized as qualitative aspects of morphological productivity except for the idea of Aronoff that has not been formalized yet.

For Baayen (1992), there are a couple of aspects that can activate but also weaken a word formation process. The aim of Baayens research is to calculate the **probability** of finding new words in a sampled text that are formed by morphological process. Therefore, Baayen (1992) develops methods to measure morphological productivity in a quantitative way. Some of these measurements will be discussed in Section 5, and applied in our study.

Some other new findings of morphological productivity in a qualitative way are made by Plag (1999), Hay (2001) and Bauer (2001). All this work is based on the English language.

For German there is still less work done for productivity in derivational morphology. There are some single studies that try to detect consolidated findings. Scherer (2005) makes a corpus study for 400 years (1600-2000), and tries to find out whether the noun building suffix *-er*, which should be productive in contemporary German, has changed during time. She finds out that word formation processes are subject to diachronic change and that they can be measured in terms of productivity. Another quantitative study is made by Luedeling & Evert. (2004) who investigate a special suffix from German that is normally used for medical description of words, but can also be used in other word formations: *-itis*. Another study comes from Schneider-Wiejowski (2009) who investigates four German suffixes (*-nis*, *-heit/-keit/-igkeit*[3], *-ung* and *-sal*) in a diachronic way in a German variety spoken in Switzerland. It can be shown that there is a morphological change during the 20th century in this variety. Some of these suffixes become more productive whereas other suffixes lose their efficiency of building new words.

The present study tests some of the productive suffixes from Schneider-Wiejowski (2009) in a German newspaper genre *diachronically* (17-20th century), and *synchronically* in a 20th century spoken language corpus.

[3]The suffixes *-heit/-keit/-igkeit* are allomorphs, and therefore combined to the same suffix category.

3 Simulation Framework

3.1 Game Architecture

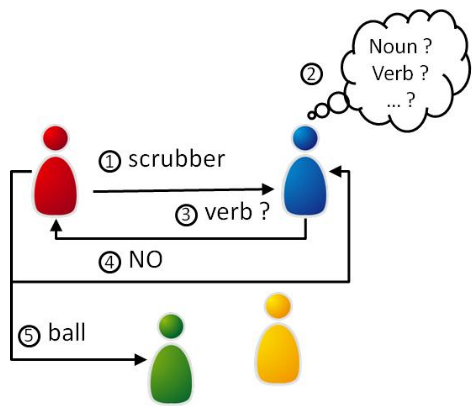

Figure 1. A possible game round in the first stage: An adult utters a word to the child (1), the child makes a guess (2-3), the adult values the answer of the child (4). If the answer was wrong, he picks the next agent and speaks to him (5), otherwise the child speaks to his neighbor. In the second stage, the answer of the adult is always YES, so that children speak more on their own.

For the purpose of the present work, we evaluate the decomposition module from the Morphological Derivation Game (MDG) introduced in Pustylnikov (2009). This model grounds on the theoretical assumptions of morphological processing Frauenfelder & Schreuder (1992), and was developed to test morphological processing mechanisms in a multi-agent simulation.

The inter- and intra-generational model of MDG consists of two types of agents adults: *A* and children *C*. In a dialogue situation (i.e., a round of the game), there is one adult who “moderates” the course of the game selecting a *C* at random and talking to it. The adult brings new words into play asking the child to guess the part-of-speech (POS) of a particular word. It might resemble the following communication situation:

a) Did you see the *scrubber*?
b) *Scrubber*?

Obviously the child (b) does not know the meaning of the word *scrubber* yet, and has to guess it. Literature on first language acquisition states that children use all the material present in the speech of the adult (Bybee, 1988) to uncover the meaning. In MDG, the agents are endowed with a ‘human-like’ decomposition module that segments the words, similar to what children are supposed to do Dressler & Karpf (1995), and identifies suffixes characteristic for a part-of-speech. The decomposition module operates on the acquired words, and filters out the most probable derivation suffixes of the input language. Pustylnikov (2009) could show that the decomposition module succeeds to identify real derivation suffixes in English and German. In case of the random language[4], children negotiate new derivation suffixes that are different from those identified by the adult (since no significant suffixes are present in the input data). In this paper, we investigate the suffixes segmented from German input data for their productivity (see Sec. 5). That is, we ask: do the suffixes identified by the decomposition procedure correlate with empirical findings on productive suffixes for German? Does material from 17th-19th century German differ from the 20th century with respect to productive suffixes?

3.2 Decomposition Module

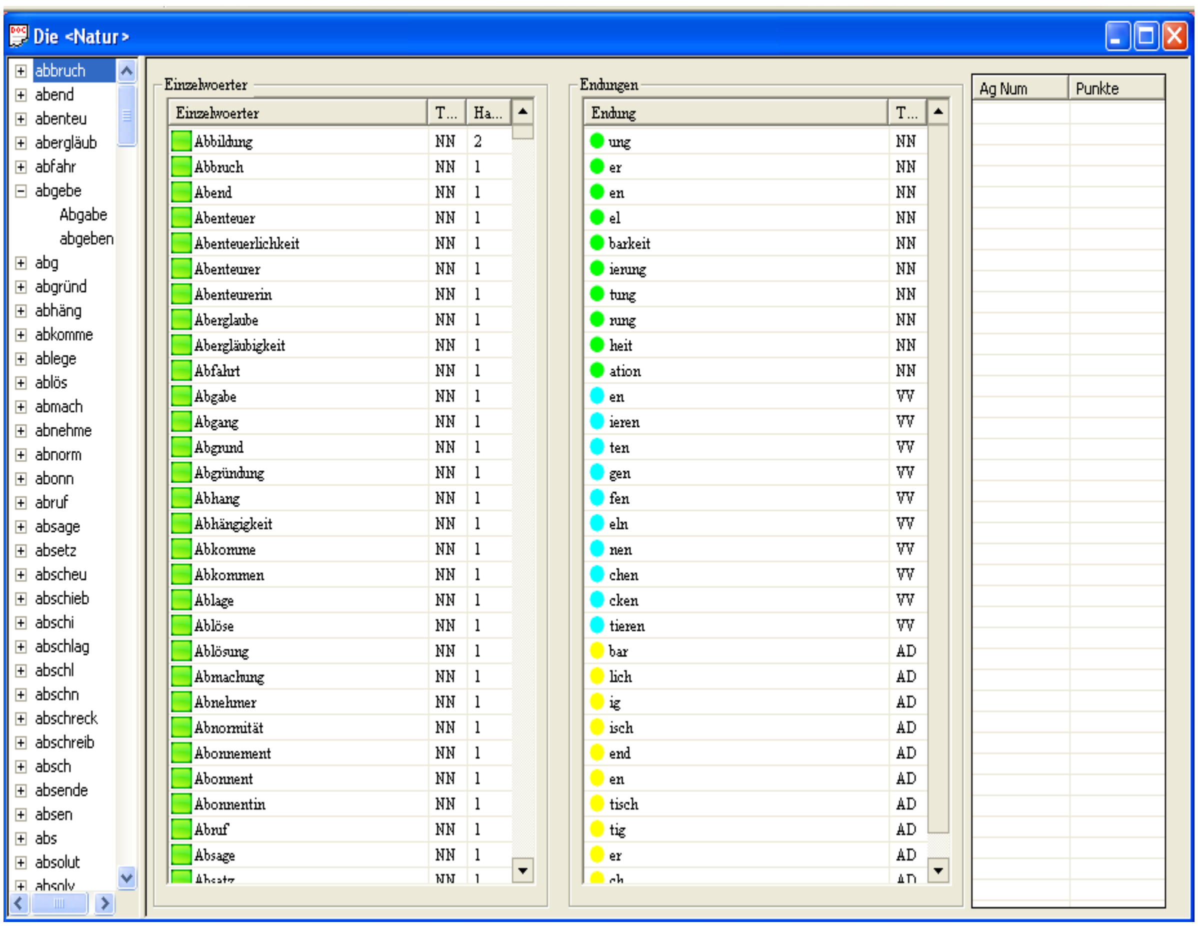

Figure 2. A snapshot of MDG after parsing a German newspaper corpus.

[4] By a random language, we mean here randomly generated strings from the Latin alphabet without any internal word structure.

The decomposition module is based on two sources: *lexicon* and *word-family* file. Here, we concentrate on the suffix induction from lexicon, which is the main mechanism in MDG to induce suffixes from raw data.[5] Figure 2 shows a screenshot of the MDG after parsing the German newspaper corpus *Süddeutsche Zeitung* (see Section 4). The first tree view displays the word families, the second the single words, the third shows suffixes of nouns, verbs and adjectives identified by decomposition, and ranked according to their relevance (estimated by the below algorithm). The last column contains the agent scores, which is empty at the beginning. Our hypothesis is that productive suffixes can be identified in language solely observing the structure of the words. The below algorithm, the same for adults and children, tries to find those suffixes that are best suited for word formation analyzing words from the lexicon. The adults' lexicon represents a corpus of texts (e.g., newspaper corpus) tagged with POS information. Adults process the whole lexicon, and have a greater chance to find the regularities of suffix use (Figure 2 shows the knowledge of the adult). Child agents acquire this knowledge partially, analyzing the words uttered by the adult during the game. Both types of agents use the following algorithm:

1. For each word class *c* and each word *w* encountered by the agent all possible suffixes $\{s_1,\ldots,s_n\} \in w$ are extracted and stored together with their frequencies F_i (for $F_i > 20\%$).[6]
2. If $F_1 \sim F_2$ according to a similarity threshold *sim* (.1 in this step) the shorter suffix is discarded. For example, if '*-isch*' and '*-ch*' have the same (or similar) frequency, then '*-ch*' is discarded.[7]
3. Step 2 is repeated comparing suffixes of *w* to all suffixes of all other words with a more sensitive value of *sim* (i.e., increasing the similarity threshold to .0000001 so that only suffixes with nearly the same frequencies are compared).
4. For each remaining suffix *s* a list of suffixes that have no substrings in common escape for *s* is constructed. That is, a list for the suffix *-ung* contains *{-tung, -lung, -kung, ...}*.

---

[5] In the case of word families, derivational suffixes are induced decomposing the words of a word family (e.g., *Geograph* > *geography* > *geographical*) into stem and possible suffixes. See Pustylnikov (2009) for details.

[6] 20% are determined heuristically in order to filter out inappropriate suffixes.

[7] This is done in order to avoid duplicates. Since all possible suffixes are analyzed separately, the suffix *-ch* needs not to be considered twice - as a separate suffix, and as a suffix within -*isch*.

In sum, a suffix is said to be productive according to the above algorithm if it is frequent (frequency above the threshold), and also preceded by a large number of different substrings (i.e., the length of the suffix-list in Step 4).

4 Corpora

Table 1
Sizes of 3 German Corpora: 17th-19th, 20th century newspaper corpus *Süddeutsche Zeitung* (2004) and a spontaneous speech corpus (German) Hinrichs & Kuebler (2005).

Corpus	**GerManC**	**SZ**	**SpokenC**
Size	110,448	975,526	362,795

In the present study, we use three different types of data: two corpora from the newspaper genre, and one corpus of spontaneous speech. The *GerManC* project[8] represents a historical corpus of German from 1650 to 1800. The aim of this project is to compile a representative corpus of written German for these time steps. This period is a period of language change because, on the one hand, the modern standard was formed during it, and on the other hand, competing regional norms were finally eliminated. Currently only newspaper samples are available, but in future the project plans to add other genres too. The newspaper corpus consists of 2000 word samples from five regions (North German, West Central German, East Central German, West Upper German, East Upper German) within three periods of fifty years (1650-1700, 1701-1750 and 1751-1800). For each region three samples were taken for each period so that 110,448 words are available to work with. We test the total amount of data available for the period from 17th-19th century to compare with the newspaper corpus from the 20th century.

The 20th century newspaper corpus is extracted from the 10 years release of the German newspaper *Süddeutsche Zeitung* (2004). We consider a sample of the total corpus that comprises 975,526 tokens (henceforth abbreviated with SZ). The *Süddeutsche Zeitung* corpus was used in recent studies to evaluate text classification techniques (see, e.g., Mehler et al. (2007)).

Finally, we measure the degree of productivity in the *Tübinger* corpus of spontaneous speech Hinrichs & Kuebler (2005). This corpus comprises 362,795 tokens of spoken dialogue, and is used here as a baseline to evaluate the productivity measures. Productivity is assumed to be the probability of using an affix when it

[8] http://www.llc.manchester.ac.uk/research/projects/germanc/

comes to generate a new word (Baayen 1991). Although, neologisms are expected to appear in spoken language, the quantitative measures proposed in the literature are widely applied to written data. In this paper, we aim to test them for a spoken corpus, too. The reason is that large divergence in results between spoken and written data could indicate a weakness of a particular productivity measure.

5 Productivity Measures

Some German suffixes like -*heit* (Schön-*heit* ’beauty’, Gesund-*heit* ’health’) or -*ung* (Digitalisier-*ung* ’digitalization’, Computerisier-*ung* ’computerization’) are very productive because it is possible to create many (new) words using them. But there are also suffixes that do not account for word formation at all. Although, there are existing lexicalized words like Ereig-*nis* or Hinder-*nis* composed with the suffix -*nis*, this suffix would not produce any new word because it is completely unproductive. Derivational productivity is gradual, and some attempts to measure productivity were made in the past. In this section, we test some of these approaches.

5.1 Productivity *P* in the narrow sense

One very popular measure is the one proposed by Baayen (1991):

(1)
$$P = \frac{n_1}{N}.$$

This measure calculates the proportion of single tokens (i.e., hapax legomena) of words ending with, for example, -*ung* divided by the number of all tokens ending with -*ung* (*N*) in the corpus. By calculating the ratio of hapaxes, the types which only have one token, Baayen gets an estimate of the probability of becoming novel forms with a given affix. As noted by Baayen, and also remarked elsewhere (Bauer, 2001) this measure is dependent on the size of the corpus. This property does not allow for direct comparisons of the values obtained from corpora of different size. *P* is according to Baayen best suited as an index of productive vs. unproductive affixes.

Table 2 shows the values of *P* for the three corpora. Obviously, the values cannot be compared directly due to different corpus sizes. However, we can compare the productivity of the suffixes within a corpus ranking them according to the values of *P*, and then examine whether this ranking changes among the corpora (i.e. we can perform a rank test). In the 17th-19th century corpus *GerManC -nis* is most productive followed by *-heit/-keit*, *-ung* and *-er*, which have similar values.

Table 2
Noun suffixes in GerManC, SZ and SpokenC according to *P*. The values are obtained for the total number of tokens.

	GerManC			SZ			SpokenC		
suffixes	n_1	N	P	n_1	N	P	n_1	N	P
-nis	4	10	.4	55	1357	.04	55	1357	.04
-ung	156	864	.18	1016	20415	.049	1016	20415	.049
-er	129	803	.162	1061	22980	.046	1061	22980	.046
-heit/-keit	79	281	.281	234	2480	.094	234	2480	.094
token	110,448			975,526			362,795		

Table 3
Noun suffixes in GerManC and SZsmall measured with *P* for corpus samples of equal size.

	GerManC			SZsmall		
	n_1	N	P	n_1	N	P
-nis	4	10	.4	25	121	.02
-ung	156	864	.18	357	2086	.171
-er	129	803	.162	361	2607	.138
-heit/-keit	79	281	.281	83	262	.316
token	110,448			110,448		

According to *P*, *-nis* seems to be more productive but when we look at the number of tokens (10) ending with *-nis*, and the number of hapaxes (4) the high value of .4 is explained by a small difference between the two, rather than, by a high productivity of *-nis*. Considering the other two corpora, SZ and SpokenC, *-nis* appears to be less productive. The sample of SZ of the size of GerManC (Table 3) shows also smaller values for *P*.

We can interpret this result as a shift in productivity with respect to *-nis*. However, obviously the values of *P* are biased by a small number of tokens, and cannot be accepted as a true indicator of a productivity shift for *-nis* (although, such a shift might exist in language).

5.2 Hapax-conditioned Productivity *P**

Table 4
Noun suffixes in GerManC, SZ and SpokenC according to *P**.

	GerManC	*P**	SZ	*P**	SpokenC	*P**
hapaxes	1354		11352		730	
-er	129	.09	1061	.09	52	.07
-ung	156	.11	1016	.09	64	.08
-heit/-keit	52	.04	234	.02	9	.01
-nis	4	.003	55	.004	7	.009

Table 5
Noun suffixes in GerManC and SZ measured with *P** in two corpus samples of equal size

	GerManC	*P**	SZ	*P**
hapaxes	1354		3789	
-er	129	.09	361	.09
-ung	156	.11	357	.09
-heit/-keit	52	.04	83	.02
-nis	4	.003	25	.006

To overcome the drawbacks of *P*, Baayen (1992) proposes another measure *P** that allows to compare affixes - within and across corpora. This measure compares hapaxes of a particular morphological category (e.g., suffix *-ung*) to all hapaxes of a part of speech (POS) (e.g., nouns) in the corpus, and asks about the contribution of this affix to all singular word formations.

(2) $$P^* = \frac{n_1}{h_t}$$

In the above formula, n_1 denotes the number of hapaxes formed by a particular word formation rule for a part of speech *t*. The quantity h_t denotes the total number of hapaxes in the corpus. Bauer (2001) doubts whether this index really measures the right thing, asking: “What proportion of new coinages use affix X?” That is, *P** assumes that productive formations are within the total amount of hapax legomena in the corpus. Bauer (2001) contrasts *P** with *P* stating that *P* asks: “What proportion of words using affix X are new coinages?” However, the assumption that hapax legomena are really new formations is also implied by *P*, and can be doubted, especially when we deal with written corpora. That is, words might occur only once in a corpus indicating an unproductive and lexicalized formation process. We are not in the position to rule out the problem of appropriate corpus selection, but we consider a spoken corpus to additionally verify the expressiveness of the measures.

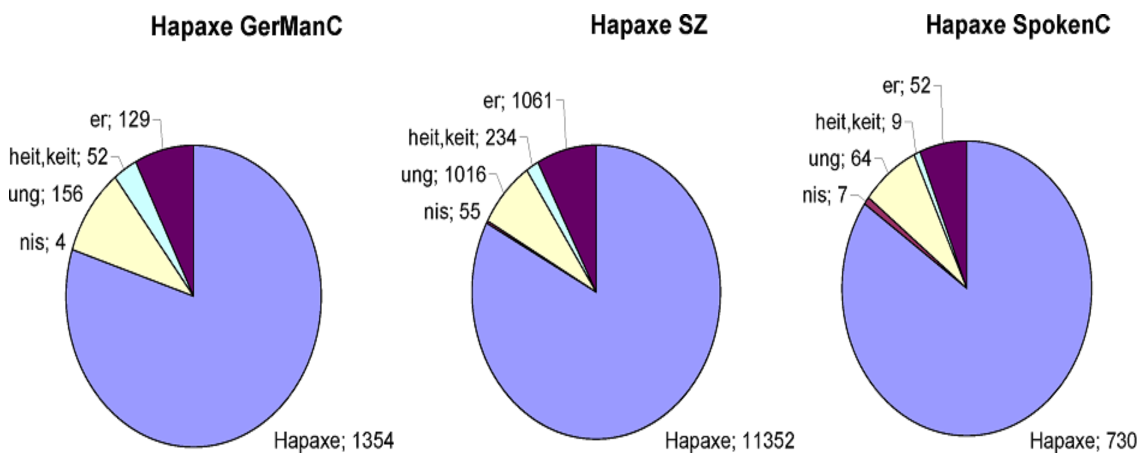

Figure 3. Noun hapaxes in GerManC, SZ and SpokenC.

Finally, Baayen assigns different functions to both measures and recommends applying them as complementary. This is done in the present paper.

The values of *P** (Tables 4-5, Figure 3) look much different from what is shown in Table 2 for *P*. According to *P**, *-ung* and *-er* have almost the same productivity, followed by *-heit/-keit* and *-nis*. The picture is consistent for different periods of time (GerManC and SZ), as well as for the spoken corpus.

5.3 Productivity based on Type Frequency

The first two measures presented in this section assume that the number of hapaxes in the corpus is an indicator of productivity. Kreyer (2009) argues that it is reasonable to investigate not only hapax legomena but the total distribution of types formed by a particular word formation rule. He proposes a measure to account for this task:

(3) $$P' = \frac{n_1}{\sqrt{1}} + \frac{n_2}{\sqrt{2}} + ... + \frac{n_x}{\sqrt{X}} = \sum_{i=1}^{X} \frac{n_i}{\sqrt{i}},$$

with n_i being the number of types that occur i times in the corpus. X is the total number of different frequencies of types i formed by the word formation rule. P' takes its maximal value if $n_1 = X$, that is, if there are only hapaxes in the corpus. Its minimal value is achieved for $n_X = 1$ if $X > 1$. That is, for example, if there is one type in the corpus ($n_X = 1$), and this type occurs 1000 times ($X = 1000$). Thus, P' takes the total frequency distribution of types into account giving less weight to highly-frequent types. A high value of P' indicates high productivity.

Table 6
Values of P' measured for GerManC, SZsmall, SZ and SpokenC.

suffixes	GerManC	SZsmall	SZ	SpokenC
-nis	.8309	.6959	.4939	.8189
-ung	.7660	.7397	.5725	.6446
-er	.7360	.7330	.5886	.6649
-heit/-keit	.7774	.8525	.7446	.7030

Results for our data are shown in Table 6. With respect to *-nis* the results are not plausible, which can be explained by a small number of different types in the GerManC and SpokenC. Figure 4 illustrates the frequency rank of *-nis* in different corpora on a log-log plot. There are only two different type frequencies (hapaxes and dis legomena) in the two corpora. While P' weights upper ranks higher, a small frequency spectrum of types results in higher values of P'. In contrast, the other two corpora (SZsmall and SZ) have a larger frequency spectrum of types, which lowers the overall result of P'. So, at least for *-nis* we cannot make any judgments about productivity according to P'.

5.4 Simulation Rank

Simulation Rank (SR) is an index that results from empirical ranking of suffixes identified by the decomposition module of MDG. According to the filtering algorithm presented in Section 3 the SR ranks the suffixes with respect to their probability to be selected for word production. Ten suffixes[9] are selected at most, and the three best ranked are those used for production. On the one hand, we evaluate the MDG's potential in identifying productive suffixes, on the other hand, we compare the ranking of SR to the results of productivity measures described above.

Table 7
Noun suffixes in GerManC, SZ and SpokenC according to *P*.

SR	GerManC	SZ	SpokenC
1	-er	-er	-er
2	-ung	-ung	-en
3	-en	-en	-ung

The decomposition module of MDG identifies only *-er* and *-ung* as appropriate for word production. Suffixes *-nis* and *-heit/-keit* are not present in the best-of-ranking. This finding is in line with the results obtained by *P**, and in line with the literature about these suffixes. Suffix *-en* is additionally ranked within the three best suffixes identified by the algorithm. The results are consistent for different time periods (17th-19th vs. 20th century) and genres (newspaper vs. spoken language).

6 Discussion

In this paper we have evaluated language decomposition techniques applied to measure morphological productivity. The simulation based productivity measurement represents a frequency based approach to productivity. This approach identifies productive suffixes in a language, and can be used to test productivity assumptions for particular suffixes.

Measuring the productivity of certain German suffixes with established methods for measuring productivity could show that at first glance it looks like *-nis* should be very productive from 1650 to 1800, whereas the other three noun forming suffixes *-er*, *-ung* and *-heit*/*-keit* seem to be less productive. The suffix *-nis* generates the highest value of *P* in the corpus of written speech.

For the SZ corpus one can say that all values except for the value of *P* for *-heit*/*-keit* are similar to each other. The allomorphs of *-heit* show a very high value of *P* (0.094). For the spoken corpus we observed according to *P* that *-ung* is the most productive suffix followed by *-nis*, whereby *-er* and *-heit*/*-keit* are not productive at all.

[9] "Ten" is a filtering parameter, which can be varied to include less relevant suffixes.

This finding does not confirm the assumptions that are made in the linguistic literature. Fleischer & Barz (1995) as well as Lohde (2006) say that all suffixes explored here should be very productive for forming new words in contemporary German except for -*nis*. The suffix-*nis* is declared as unproductive. Therefore, we have to question our findings and find explanations to state this. First, we can think about the value of *P* itself and what it should express. All words that occur only once in a given sample should be hapax legomena in the sense of Baayen. This means, hapax legomena are put on the same level with newly created words. But there is no guarantee for making this implication. There are many reasons why one word only occurs once in a sampled text:

- The word is new and was created by word formation process.
- The word is very old and rarely used.
- The word is not new, and it is still used in a language but it only occurs once by chance.

The other thing one has to consider is that *P* should be interpreted as a constrained value dependent on corpus size. And in our first experiment the total number of tokens is not the same for all corpora. The corpus from the 17th century is a very small one, and it consists only of approximately 100,000 tokens whereas the SZ corpus is nine times larger.

Therefore, we decided to select a sample of the SZ corpus of about 100,000 tokens and to look at *P* again. After shortening the available token size, all values change. The new calculation shows that -*heit*/-*keit* should be the most productive one of the four suffixes followed by -*ung* and -*er*. -*nis* is unproductive. This conclusion comes closer to our intuition about the productivity of -*nis*, however, -*ung* and -*er* are more productive than -*heit*/-*keit*, which is not confirmed by the results.

If we look at the results of the second experiment in which hapax-conditioned productivity was examined we can come to another conclusion. Table 4 show that -*nis* is less productive than the other three suffixes independent of time period or genre. Suffixes -*er* and -*ung* should be very productive, and -*heit* can still be interpreted as a productive suffix. And these values do not change after shortening the number of examined tokens from the SZ corpus (Tab. 5). Amazingly there is the same structure for all ratings and these values do not depend on the size of the corpus. That means, according to *P** there is no shift in productivity in the period of time examined here. On the one hand, this result shows the stability of *P** in contrast to *P*. On the other hand, it is probably not sufficient to consider hapaxes only, and we are better off to look at the full frequency distribution of words formed by an affix.

To account for this, we used the measure of Kreyer (2009) that combines all type frequencies in a single measure giving to less frequent types (i.e., hapaxes etc.) higher weights. The results (Tab. 6) for *-nis* are biased by the small number of types, and cannot be treated as representative. The overall results for *-ung* and *-er* seem to be plausible, since both suffixes have similar values in all the corpora, which is in line with our expectations. The suffix *-heit/-keit*, however, has better values, than *-ung* and *-er*, which is understandable, too. These allomorphs are declared as productive in German, and it is possible to create new words using them. In sum, it seems plausible to look at the type frequency spectrum as Kreyer does, however the significance of the results with respect to productivity needs further testing.

Other factors like phonological and semantic constraints for selecting a suffix might also influence productivity of an affix but a quantitative model capturing these kinds of information is still missing. Therefore, there is a research gap, and in future this question should be investigated more in depth, especially for German.

As the theoretical background it is possible to describe the phenomenon of productivity with the grammaticalization theory, which is discussed to a great extent in the literature (see e.g., Diewald (1997), Heine (2003), Hopper & Traugott (2003), Lehmann (1995)). Grammaticalization is a cycling process. That is, existing lexical items are worn down, but at the same time new grammatical affixes are created. Although, there is a controversial discussion about the question whether the grammaticalization theory can be used to explain the emergence of affixes and not - in the narrow sense - just for changes from syntax to clitics (see Nuebling (2006:72)), we assume that it can be used for derivational morphology. If grammaticalization is a cycling process, it should be also explained what happens with affixes that become unproductive. We assume that the German languages (and other languages too) have a *balanced derivation system*. If one affix becomes unproductive, another one will get more productive. If one suffix dies, another one is born. There are a few examples of present day German for affixoids, emerging affixes that will maybe convert to affixes once. One of such an affixoid is *höllen-* (*hell-*) that is very productive now, especially in spoken language: there can be a *Höllenhitze* ('a hell heat') or one can have a *Höllentag* ('a hell day').

7 Conclusion

In this article, we analyzed morphological productivity of four German suffixes using four different approaches. We evaluated three quantitative measures proposed by (Baayen, 1991; Baayen, 1992; Kreyer, 2009), and one frequency based approach used for language decomposition. The simulation based ranking allows us to identify productive suffixes in language and rank them according to their productivity. The results of the ranking are consistent with the literature on productivity of these suffixes. Results of SR and *P** when applied to different kinds of corpora show almost the same degrees of productivity, either for different time spans, or for different corpus genres. This finding indicates that either the degree of productivity for these suffixes has not changed much in this time span, or this variation is not captured by the measures. To answer this question, additional experiments with speakers' intuition about the use of suffixes, as well as more quantitative tests are needed. The measure *P* shows a variation among the different kinds of corpora, however, this variation is not confirmed by findings from the literature. In sum, although, quantitative approaches like those examined here succeed to distinguish productive suffixes from the unproductive ones, there is still much work to be done to grasp the dynamics driving productivity from synchronic and diachronic perspectives.

Acknowledgments

This research is supported by the German Research Foundation *Deutsche Forschungsgemeinschaft* (DFG) in the Collaborative Research Center 673 'Alignment in Communication'. We would also like to thank Alexander Mehler for fruitful discussions and comments.

References

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, Mass.: MIT Press

Baayen, H. (1991). Quantitative aspects of morphological productivity. In: J. M. Geert Booij (ed.), *Yearbook of Morphology: 109–149*. Dordrecht-Boston-London: Kluwer.

Baayen, H. (1992). On frequency, transparency, and productivity. *Yearbook of Morphology 1992*: *181–208*.

Bauer, L. (2001). *Morphological productivity*. Cambridge: Cambridge University Press.

Bybee, J.L. (1988). *Morphology as lexical organization*. Academic Press. (New York)

Cowie, C., Dalton-Puffer, C. (2002). Diachronic word-formation over time: Theoretical and methodological considerations. In: Javier Días Vera. *A*

Changing World of Words. Studies in English Historical Lexicography, Lexicology and Semantics: *410–437*. Rodopi.

Diewald, G. (1997). *Grammatikalisierung: Eine Einführung in Sein und Werden grammatischer Formen*. Tübingen: Niemeyer (= Germanistische Arbeitshefte 36).

Dressler, W.U., Karpf, A. (1995). The theoretical relevance of pre- and proto-morpholgy in language acquisition. *Yearbook of Morphology 1994: 99–122*. Springer. Dordrecht-Boston- London: Kluwer.

Evert, S., Lüdeling, A. (1991). Constraining psycholinguistic models of morphological processing and representation: the role of productivity. In: G. Booij and J. van Marle (eds.), *Yearbook of morphology 1991: 165–183.* Dordrecht: Kluwer.

Evert, S., Lüdeling, A. (2001). Measuring morphological productivity: is automatic preprocessing sufficient? In: P. Rayson, A. Wilson, T. McEnery, A. Hardie, and S. Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 Conference*: *167–175*. Lancaster: Peter Lang.

Fleischer, W., Barz, I. (1995). *Wortbildung der deutschen Gegenwartssprache*. Tübingen: Max Niemeyer Verlag.

Habermann, M. (1994). *Verbale Wortbildung um 1500. Eine historisch-synchrone Untersuchung anhand von Texten Albrecht Dürers, Heinrich Deichlers und Veit Dietrichs*. Berlin: de Gruyter,

Hay. J. (2001). Lexical frequency in morphology: Is everything relative? *Linguistics 39(4), 1041–1070.*

Heine, B. (2003). Grammaticalization. In: Joseph, B. D. & Janda, R. D. (Eds.), *The Handbook of Historical Linguistics: 575–601*. Oxford: Blackwell.

Hinrichs, E.W., Kübler, S. (2005). Treebank profiling of spoken and written german. In: *Proc. of the Fourth Workshop on Treebanks and Linguistic Theories (TLT)*, December 2005. Barcelona, Kluwer Academic Publishers.

Hopper, P.J., Traugott, E.C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.

Jäger, G. (2008). Evolutionary stability conditions for signaling games with costly signals. *Journal of Theoretical Biology 253, 131–141.*

Kirby, S. (2007). The evolution of meaning-space structure through iterated learning. In: C. Lyon, C. Nehaniv, A. Cangelosi (eds.), *Emergence of Communication and Languag*e: 253–268. Berlin: Springer Verlag..

Kruisinga, E. (1932). *A Handbook of Present-Day English, volume 5*. Groningen: Noordhoff.

Lehmann, C. (1995). *Thoughts on grammaticalization*. LINCOM studies in theoretical linguistics 1. LINCOM Europa,. München/Newcastle: Lincom (= Lincom Studies in Theoretical Linguistics 01).

Lohde, M. (2006). *Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch*. Tübingen: Narr.

Lüdeling, A., Evert, S. (2004). The emergence of productive non-medical *-itis*: corpus evidence and qualitative analysis. In: *Proceedings of the First International Conference on Linguistic Evidence: 351–370.* , Mouton de Gruyter, Berlin, New York.

Mehler, A., Geibel, P., Pustylnikov, O. (2007). Structural classifiers of text types: Towards a novel model of text representation. *Journal for Language Technology and Computational Linguistics (JLCL) 22(2), 51–66.*

Munske, H. (2002). Wortbildungswandel. In: Habermann et al. (ed.), *Historische Wortbildung des Deutschen: 23–40*. Tübingen: Niemeye.

Nübling, D., Dammel, A., Duke, J., Szczepaniak, R. (2006). *Historische Sprachwissenschaft des Deutschen. Eine Einführung in die Prinzipien des Sprachwandels*. Tübingen: Gunter Narr Verlag.

Plag, I. (1999). *Morphological productivity. Structural constraints in English derivation*. Berlin-New York: Mouton de Gruyter.

Prell, H.-P. (1991). *Die Ableitung von Verben aus Substantiven in biblischen und nichtbiblischen Texten des Frühneuhochdeutschen*. Frankfurt am Main: Lang.

Pustylnikov, O. (2009). Modeling learning of derivation morphology in a multi-agent simulation. In: *Proceedings of IEEE Africon 2009*. IEEE, September 2009. AFRICON 2009, Nairobi, IEEE.

Scherer, C. (2005). *Wortbildungswandel und Produktivität. Eine empirische Studie zur nominalen -er-Derivation im Deutschen*. Tübingen: Niemeyer.

Schneider-Wiejowski, K.. Sprachwandel anhand von Produktivitätsverschiebungen in der schweizerdeutschen Derivationsmorphologie. *Linguistik online 38, 2009.*

Schultink, H. (1961). Produktiviteit als morfologisch fenomeen. *Forum der Letteren 2, 110–125.*

Steels, L. (2005). The role of construction grammar in language grounding. http://www.isrl.uiuc.edu/ amag/langev/paper/steels_constructionGrammar.html, (Retrieved November 2005).

Stricker, S. (2000).. *Substantivbildung durch Suffixableitung um 1800 : untersucht an Personenbezeichnungen in der Sprache Goethes.* Heidelberg: Universitätsverlag Winter.

Süddeutscher Verlag. (2004). *Süddeutsche Zeitung 1994-2003.* 10 Jahre auf DVD. Süddeutscher Verlag. (München).

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press, March 2005.

Vogt, P. (2003). Iterated learning and grounding: From holistic to compositional languages. In: S. Kirby (ed.), *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI*, pages 76–86, Vienna, 2003.

Willmanns, W. (1899).. *Deutsche Grammatik. 2. Abteilung. Wortbildung*. Verlag von Karl J. Trübner, Strassburg, 2 edition.

Appendix

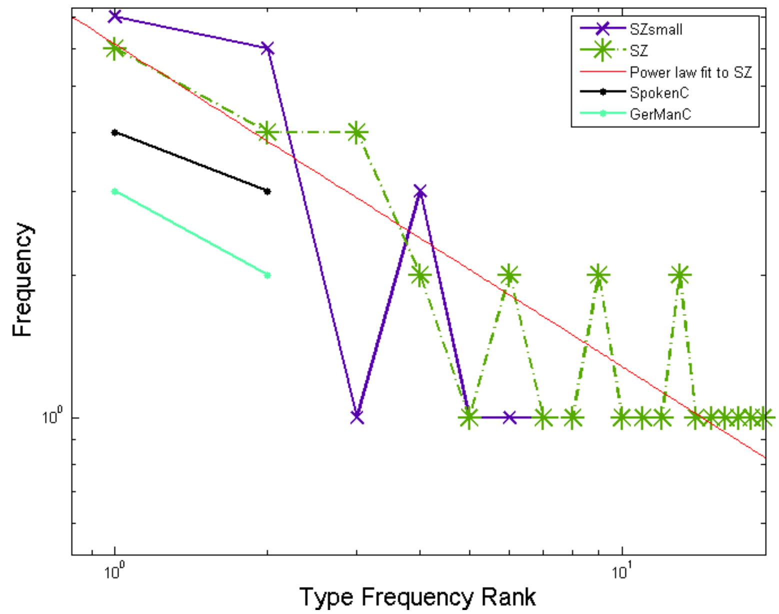

Figure 4: Frequency distributions of types formed with -*nis* in GerManC, SZsmall, SZ and SpokenC on a log-log plot (values from Table 8).

Type Frequency	GerManC	SZsmall	SZ	SPOK
1	3	7	6	3
2	2	6	4	4
3		1	4	
4		3	2	
5		1	1	
6		1	2	
7		1	1	
8			1	
9			2	
10			1	
11			1	
12			1	

13	2
14	1
15	1
16	1
17	1
18	1
19	1

Table 8: Frequencies of hapax, dis, ... *n*-legomena formed with -*nis* in GerManC, SZsmall, SZ and SpokenC.