

Morphological Productivity of Adjective Formation in German – A Diachronic Corpus Study Using the CLARIN-D Infrastructure

Erhard Hinrichs

Seminar für Sprachwissenschaft
Eberhard Karls Universität Tübingen, Germany
erhard.hinrichs@uni-tuebingen.de

Abstract

This paper demonstrates how large digital collections of diachronic and synchronic corpus data can shed new light on traditional research questions in morphology and diachronic linguistics. More specifically, the question of productivity of derivational affixes, a long-standing issue in the study of morphology, can be addressed in a much more fine-grained manner, especially from a diachronic perspective. As its empirical basis, the study makes use of two linguistically annotated corpus collections whose individual texts include high-quality metadata information about the date of origin, text type, and text size. These corpus collections are available as part of the CLARIN-D research infrastructure and are accompanied by powerful query and concordancing tools that support searching for morpho-syntactic patterns and visualization of query results.

1 Introduction

The availability of digital resources has led to an empirical turn and a renaissance of corpus-based research in the field of linguistics. The study of morphology has been no exception, uncovering, inter alia, the importance of frequency effects for language processing and linguistic change. A closely related, long-standing research issue in the field of morphology concerns the productivity of morphological processes, particularly of derivational morphology and word formation. Two closely related questions have been addressed here: (i) to identify the number of distinct lemmas that can combine with a given derivational affix at a given time (the synchronic perspective), (ii) to track the number of lemmas that can combine with a given derivational affix over time (the diachronic perspective).¹

The present paper examines the productivity of the derivational affix *-bar* in German that forms adjectives that in contemporary German express the modality of possibility or of disposition. The reason for focussing on *-bar* adjectives is motivated by remarkable shifts in productivity over time along both dimensions identified above. The significance of corpus-based research for estimating the productivity of *-bar* adjectives in German from a synchronic perspective was previously shown by Riehemann (1993). The present study covers the entire Modern High German period, which extends from the 17th century to the present.

The empirical basis for this study is provided by the Mannheimer Zeitungskorpus (henceforth: MZK) and by the Deutsches Textarchiv (henceforth: DTA; www.dta.de) (Geyken et al., 2011), which are both available as part of the CLARIN-D research infrastructure at the CLARIN Centers at the Berlin-Brandenburg Academy of Sciences (BBAW) and at the Institut für deutsche Sprache (IDS), respectively. The DTA corpus collection provides extensive coverage of the entire Modern High German period and offers accompanying query and visualization tools that are indispensable for the kind of study reported on here. The DTA contains German texts ranging from 1610 to 1900. The texts have been digitized and transliterated, using a high-precision double-keying method. The archive is still under construction. The version used for the present study dates from April 26, 2017 and consists of 146,681,395 lexical tokens with 1,025,417,920 Unicode characters that are taken from 595,929 digitized pages and 2,628 different

¹For a more comprehensive overview of the study of morphological productivity and the importance of corpus evidence therein see (Baayen, 2009).

published works. The texts represent different genres, including novels and other literary works, scientific and journalistic texts. The MZK, used by Riehemann (1993) for her corpus study of *-bar* adjectives, consists of newspaper articles written between 1985 and 1988 and contains 10,766,759 word tokens. The corpus is available as part of the Deutsches Referenzkorpus at the CLARIN-D Center at the Institut der deutschen Sprache in Mannheim.

From a diachronic perspective, *-bar* adjectives in German undergo remarkable shifts in productivity over the course of the Modern German period, as shown in Figure 1. The chart in Figure 1 is generated as a result of the query shown in (1).

(1) $\$p=\{ADJA,ADJD\}$ with $\$l=*\{bar\}$ #less_by_date[1600,1900]

The query is expressed in the DDC formalism (Jurish et al., 2014) that was developed for the DTA and for the DTA query web application available at <http://kaskade.dwds.de/dstar/dta/>. The query in (1) searches for tokens with part-of-speech tags ADJA or ADJD, the part-of-speech tags for attributive and predicative adjectives contained in the Stuttgart-Tübingen tagset (STTS, Schiller et al. (1995)), which is a de-facto standard for automatic morpho-syntactic annotation and adopted by the DTA. The inclusion of charts such as Figure 1 is facilitated by the export and download functions provided as part of the DTA query web application. The tracking of linguistic patterns over time further pre-supposes reliable and fine-grained metadata about the textual sources. The necessary reliability is ensured by the data collection policy for the DTA which adheres to the principle that only first editions of individual texts are included in the DTA collection so as to ensure highly reliable metadata. The DDC query language supports filtering of queries by specific time intervals such as 1600 to 1900 as the time period covered by the DTA.

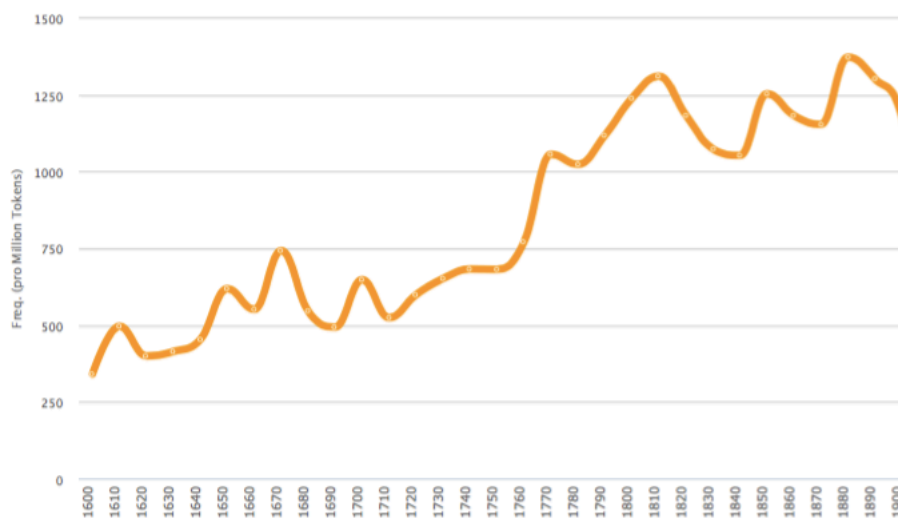


Figure 1: DTA data visualization of *-bar* adjectives

Figure 1 clearly shows an overall increase in diachronic morphological productivity of *-bar* adjectives over the period of Modern High German until the beginning of the 20th century. On average, this type of adjective occurs 558.61 times per 1 million tokens in the DTA over the period of 1600 to 1900. Its frequency climbs from an average of 341.19 per 1 Mio. tokens for the first decade of 1600 to an average of 1,373.26 for the last decade of the 19th century.²

One interesting question concerning the increase in productivity of *-bar* adjectives is whether this is true for other derivational affixes of adjective formation in German as well, such as *-lich*, as in *lös-lich* 'soluble', or *-ig*, as in *eck-ig* 'square'.

As a case in point, Figure 2 shows that the average number of occurrences per 1 Mio. tokens for

²The downward tail on the right edge of Figure 1 and of Figure 2 for the time stamp 1900 is due to a zero count for any decades that are out of range for a given query.

adjectives with the suffix *-lich* remains relatively stable over the entire period covered by the DTA, starting with an average of 10,402.06 per 1 Mio. tokens for the first decade of 1600 to an average of 9,810.13 for the last decade of the 19th century. In the full paper, similar charts will be shown for other derivational affixes whose morphological productivity remains stable throughout the Modern German Period. They are omitted here, due to lack of space. Thus, the change in productivity for *-bar* adjectives distinguishes this suffix from other derivational suffixes of adjectives. This in turn then calls for a principled explanation as to what may have caused this change in productivity.

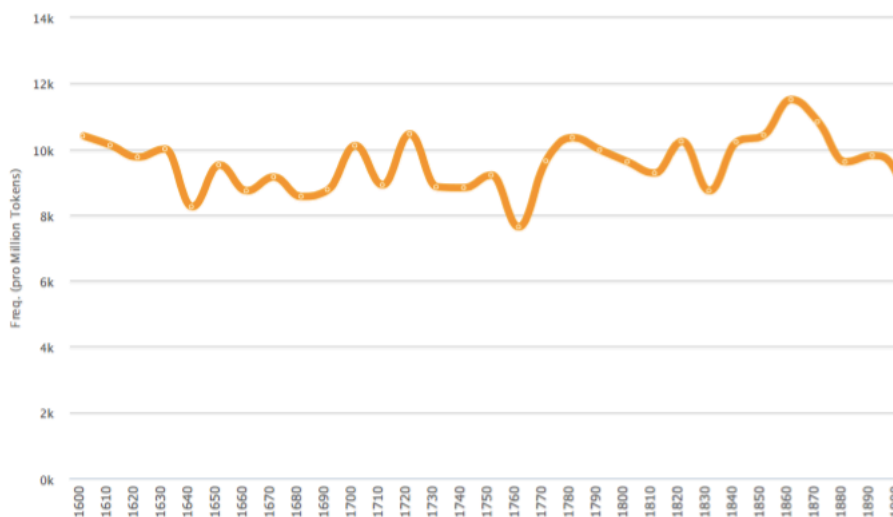


Figure 2: DTA data visualization of *-lich* adjectives

It turns out that a comparison of the DTA corpus collection and the MZK examined by Riehemann holds part of the answer to this puzzle. Table 1 contrasts the ten most frequent *-bar* adjectives occurring in the DTA during the 1st half of 17th century with the top ten *-bar* adjectives in the MZK.³

1600 - 1649		1985-1988	
Count	lemma	Count	lemma
816	<i>sonderbar</i> 'strange'	1698	<i>offenbar</i> 'obvious'
568	<i>fruchtbar</i> 'fertile'	1081	<i>unmittelbar</i> 'immediate'
532	<i>offenbar</i> 'obvious'	850	<i>sichtbar</i> 'visible'
394	<i>ehrbar</i> 'honorable'	811	<i>scheinbar</i> 'seemingly'
108	<i>sichtbar</i> 'visible'	595	<i>denkbar</i> 'thinkable'
235	<i>wunderbar</i> 'wonderful'	498	<i>vergleichbar</i> 'comparable'
214	<i>dankbar</i> 'thankful'	465	<i>erkennbar</i> 'recognizable'
185	<i>unsichtbar</i> 'thankful'	325	<i>spürbar</i> 'noticeable'
157	<i>unfehlbar</i> 'infallible'	312	<i>wunderbar</i> 'wonderful'
112	<i>unwandelbar</i> 'immutable'	300	<i>absehbar</i> 'foreseeable'

Table 1: Ten most frequent *-bar* adjectives for the 1st half of the 17th century (source: DTA) versus the 20th century (source: MZK).

Among the 303 most common *-bar* adjectives, which occur at least five times in the MZK, only fourteen such adjectives do not have a verbal base. Cf. Riehemann (1993), p. 77. Interestingly, four of these fourteen adjectives are at the top of the top ten list for the MZK shown in the right column of Table 1. By contrast, among the 101 distinct *-bar* adjectives which occur at least once in the DTA between 1600 and 1649, only 12 are unequivocally derived from verbs, and for an additional 17 the *-bar* suffix combines

³The DTA data in Table 1 have been compiled with the help of the concordancing function available as part of the DDC query language.

with a form that can be either analyzed as a verbal stem or as a so-called zero-derived nominal. Such instances include *streitbar* 'argumentative', where *streit* can either be the zero-derived nominal *streit* 'disagreement' or the verbal stem *streit-* 'to argue'.

The shift in productivity from *-bar* primarily combining with nouns to a suffix attaching to verbs is due to a shift in meaning of this suffix. According to Flury (1964), it is the *-bar* suffix itself that has undergone a shift in meaning. Modern High German *-bar* is derived from Middle High German *bære* and Old High German *-bâri*, both meaning 'bear'; i.e., the original meaning of *dankbar* is that of someone bearing thanks. It is due to this earlier meaning of *bar* that it combines primarily with nouns, a situation still in evidence at the beginning of the Modern German Period, as witnessed in the DTA. This earlier meaning of *bar* has gradually been transformed into the modal meaning of the suffix *-bar* in contemporary German that signals disposition and/or ability. The shift in meaning that the *-bar* suffix has undergone diachronically is quite plausible. For example, *fruchtbarer Boden* 'fertile soil' originally referred to soil bearing fruit, and then gradually shifted to soil **capable of** bearing fruit. Likewise, *dankbar* 'thankful' shifted from bearing tokens of gratitude to having the disposition of expressing gratitude. The shift in meaning of *-bar* has resulted in a more abstract, and thus more generally applicable lexical semantics of this suffix. From this observation, the increase in productivity follows immediately.

The contemporary dispositional meaning of the suffix is, in turn, readily compatible with combining with a verbal rather than a nominal base. The beginnings of this shift from a nominal to a verbal suffix can be witnessed in the DTA data of the 17th century. As mentioned above, some of the *-bar* adjectives attested in the DTA for that period, such as *streitbar* and *dankbar*, can be either analyzed as a verbal stem or as a so-called zero-derived nominal, paving the ground for (re-)analyzing such forms as verbal stems, as the dispositional meaning of the suffix became more and more established.

Due to space limitations, this abstract can only sketch the findings of the synchronic productivity of the *-bar* suffix in contemporary German. The received wisdom, first summarized by Toman (1983), is that the *-bar* suffix combines with transitive verbs. However, Riehemann has pointed that the MZK contains counterexamples to this generalization such *oxidierbar* 'capable of oxidizing' and *verrottbar* 'decomposable' that are derived from the intransitive verbs *oxidieren* 'oxidize' and *verrotten* 'decompose'. It turns out that such examples are not a novel phenomenon, but are already attested in the DTA. In the full paper, this matter will be described in more detail.

References

- [Baayen2009] R. Harald Baayen. 2009. *Corpus Linguistics in Morphology: Morphological Productivity*. A. Lüdeling and M. Kyto eds.: *Corpus Linguistics. An International Handbook*. Berlin: Mouton De Gruyter, pp. 900-919.
- [Flury1964] Robert Flury 1964. *Struktur- und Bedeutungsgeschichte des Adjektiv-Suffixes -bar*. Winterthur: P. G. Keller.
- [Geyken et al.2011] Alexander Geyken, Susanne Haaf, Bryan Jurish, Matthias Schulz, Jakob Steinmann, Christian Thomas und Frank Wiegand. 2011. Das Deutsche Textarchiv: Vom historischen Korpus zum aktiven Archiv. S. Schomberg et al. eds. *Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland*. pp. 157-161.
- [Jurish et al.2014] Bryan Jurish, Christian Thomas, and Frank Wiegand. 2014. Querying the Deutsches Textarchiv. In: U. Kruschwitz, F. Hopfgartner, and C. Gurrin eds.: *Proceedings of the Workshop MindTheGap 2014: Beyond Single-Shot Text Queries: Bridging the Gap(s) between Research Communities* (co-located with iConference 2014, Berlin, 4. März, 2014), pp. 25–30.
- [Riehemann1993] Susanne Riehemann. 1993. Word Formation in Lexical Type Hierarchies – A Case Study of *bar*-Adjectives in German. Sfs Report Nr. 2-93. Seminar für Sprachwissenschaft, Eberhard-Karls-Universität Tübingen.
- [Schiller et al.1995] Anne Schiller, Simone Teufel, and Christine Thielen. 1995. *Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS*. Technical Report. Universität Stuttgart, Institut für maschinelle Sprachverarbeitung, and Seminar für Sprachwissenschaft, Universität Tübingen.
- [Toman1983] Jindrich Toman. 1987. *Wortsyntax: Eine Diskussion ausgewählter Probleme deutscher Wortbildung*. Tübingen: Niemeyer.