

The replacement of diminutive suffixes in the New High German period

A time series analysis in word formation

Alfred Lameli

Albert-Ludwigs-Universität Freiburg

This article addresses the replacement of the diminutive form *-lein* by *-chen* as the leading suffix in written German during the New High German period. A large sample of diminutives from 1600–1900, retrieved from the DTA corpus, forms the basis of this investigation. The study aims to provide a detailed periodization of the replacement process with regard to both types and tokens. By using methods from computational linguistics and time series analysis, clear patterns of language variation and change are demonstrated; these patterns are to some extent interrelated and staggered throughout time. The study additionally shows that there are transfers between genres that coincide with the transition from *-lein* to *-chen*. This indicates that the replacement is due to a strengthening of the semantic effort of the diminutive suffix. Finally, information regarding the writers' origins is used to map the areal distribution of diminutive forms over time. The maps illustrate the importance of geography as a factor, particularly in the initial phase of the replacement; however, it seems to be without any impact by the end of the 19th century. In doing so, the study offers a particular framework for the analysis of word formation based on historical corpora.

Keywords: diminutive, word formation, morphology, language history, language variation, language change, corpus linguistics

1. Introduction

Almost every language uses diminutive forms; these forms express the root meaning but with a sense of smaller size (Jurafsky 1996: 534). Contemporary written Standard German is an interesting case as, for hundreds of years, it has preserved two main diminutive suffixes that are interchangeable both in their semantic

function and their syntactic construction (Wellmann 1975:131). Both suffixes *-lein* and *-chen* – which are often referred to as the liquid suffix and the guttural suffix, respectively – can be traced back to early Germanic, but there have been differences in their usage over time. During the period in which the modern supra-regional German written language came into existence, the Early New High German period (ENHG, ~1350–1650), *-lein* was predominant, while *-chen*, in contrast, was only rarely attested (Öhmann 1972; Stricker 2000:388). By the 18th century, *-lein* had been replaced by *-chen* along with a third suffix *-gen*, although *-gen* disappeared shortly thereafter (Schebben-Schmidt 1990; Wegera 2000). The *-lein* suffix continues to exist in the present day; however, nowadays, it is preferred within particular genres such as fairy tales, while *-chen* is currently the most common suffix (see Wellmann 1975:133, Fleischer & Barz 2012:233).

This shift in patterns of word formation offers insight into the complex dynamics of the process of language standardization (Davies & Langer 2006:224–241). The history of German diminutives has proved challenging for many studies, which have come at the problem from a variety of perspectives (e.g., Polzin 1901; Pfennig 1904; Wrede 1908; Öhmann 1972; Ettinger 1980; Wegera 1982; Nieuwenhuis 1985; Tiefenbach 1987; Schebben-Schmidt 1990; Dressler & Barbaresi 1994; Wegera 2000; Wegera & Solms 2002; Schmuck 2009; Ehlers 2011; Ott 2011; Elmentaler 2013; Elspaß 2015; Edelhoff 2016; Lameli 2018). One methodological problem is the fact that diminutives appear with a reasonably high frequency in spoken language but are less frequent in written language. Following Seebold (1983:1250), it is thus rather simple to describe their patterns of use in regional dialects, but difficult to define their use in the history of the German written language.

Early studies on diminutives tackled a variety of questions: they tried to verify the Latin origins of German diminutives (Polzin 1901), analyzed the diminutive usage of individual authors (Pfennig 1904), examined the geographical distribution of diminutives (Wrede 1908), and looked into the productivity of the *-chen* suffix (Gürtler 1909a, 1909b). Some authors have indicated that grammarians and renowned writers have influenced diminutive use (Öhmann 1972), while others have shown both similarities and differences between German and other languages in their use of diminutives (e.g., Ettinger 1980; Dressler & Barbaresi 1994; for a typological discussion of German diminutives that examines the question of language universals, see Nieuwenhuis 1985 and Jurafsky 1996).

Recent studies have been more explicitly designed as corpus analyses, but, in principle, tend to have similar aims. Müller (1993) and Stricker (2000) analyzed word formation processes of individual writers; Wegera (1982) discussed both the regional distribution and the development of the suffixes within the ENHG period, while Schebben-Schmidt (1990) focused on the regional distribution of the suf-

fixes in the first and second half of the 18th century; Wegera & Solms (2002) provided information about the proportion of use of the suffix *-chen* in the period 1650–1700; Elspaß (2015) compared suffixes from the 17th to the 19th century in the contexts of different degrees of formality. However, even though the body of research indicates that this is a well-documented situation, Wegera (2000:43) argued that the history of the *-lein/-chen* replacement is rather unclear, as the focus of the early studies is sometimes too narrow, while the more recent ones are more general. This lack of clarity is particularly the case in the 18th century, when the actual replacement took place. There is information on different patterns of diminutive use in the first half of the century versus the second half (Schebben-Schmidt 1990), but nothing is known about the actual process of replacement. Thus, it becomes clear why Wegera (2000:43) identified the lack of any comprehensive spatio-temporal documentation. Until today, it has never been carried out.

The present article addresses this necessity via an in-depth investigation of diminutive suffixes over a time span of 300 years, primarily during the New High German period (NHG, ~1650-). By using techniques from computational linguistics and time series analysis, a detailed periodization of diminutive replacement will be provided. The main question of this paper is the evolution of the suffix replacement. This question will be addressed by highlighting (i) the interrelation amongst the three suffixes and (ii) the associations among their use, literary domains, and the regional origin of the diminutive users.

The remainder of this article is structured as follows. Section 2 provides general findings from previous studies and outlines the consequences of the present analysis. Section 3 presents the corpus being used and outlines the study's methodology. Section 4 presents the results by (i) documenting the historical process, (ii) providing further insights into the data by highlighting differences amongst genres, and (iii) analyzing the geographical distribution of the data. Section 5 summarizes the findings and provides a conclusion.

2. Background

2.1 General development

In terms of the history of the diminutives in the 17th century, the following summary can be given. According to Wegera & Solms (2002), the proportion of *-chen* in the period 1650–1700 amounted to 6.8% of all diminutive forms. This result was based on the Bonner Frühneuhochdeutschkorpus (FnhdC; see Lenders & Wegera 1982; Wegera 1982), which is a collection of excerpts from 40 ENHG texts of different genres. However, in the newspaper collection of the German

Manchester Corpus (GerManC), Elspaß (2015) found that *-chen* and *-gen* do not occur within this same period, which indicates a difference of use according to genre. Gürtler (1909a, 1909b), in contrast, examined an unspecified selection from different authors and genres in the ENHG period. His main focus was the propagation of *-chen*, for which he found a clear increase between 1620 and 1660, though he did not give any indication of proportion. According to Polzin (1901:107) this increase of *-chen* might have occurred at the same time as a decrease in use of diminutives as a whole. Due to the rather unclear periodization in more recent corpus studies, however, there is no detailed information about this.

Based on his observations, Gürtler (1909a:33f.) indicated that there was a strong increase of *-chen* within the period 1730–1750, to the point that this suffix was used exclusively by around 1770. Öhmann (1972) disagreed with this assumption and identified several authors and grammarians who continued to use *-chen* into the subsequent period. On the other hand, Schebben-Schmidt (1990) indicated that there had been a transition in use from the first half to the second half of the 18th century in a corpus of 18 prose texts as well. Her main finding was that there had been a nearly complete loss of *-lein* in all texts of the Central German and the Southern German language area (see Figure 12 in the Appendix for an overview of the German language area). In Northern Germany *-lein* was, in contrast, maintained alongside *-chen*, simply with lower frequency. This is in line with Elspaß (2015), who found evidence of the existence of *-gen* and *-chen* in the newspaper corpus only from 1700 onward, alongside a complete loss of *-lein* during the second half of the 18th century.

In addition, Stricker (2000) provided a comprehensive insight into diminutive use in Goethe's œuvre that roughly covered the period between 1770 and 1830. Converting her individual counts into proportions, there were 14% *-lein* tokens and 86% *-chen* tokens; in contrast, her data show that there were 19% *-lein* types and 81% *-chen* types (Stricker 2000:230, 379). This contrasts with Wellmann's (1975:133) findings regarding diminutive use in selected writings of Kleist, Hölderlin, and Goethe in the years around 1800. Wellmann explored the individual book indices in detail, providing an analysis of types. He found a proportion of 69% *-chen* suffixes (90 hits) to 31% *-lein* suffixes (31 hits), indicating, like Stricker's analysis did, that *-chen* was dominant, but that *-lein* was not completely lost but rather quite vital. This is due to a renewal of uses of *-lein* that Pfennig (1904) had also found in the work of Schiller and contemporaries. It continued into the writings of Romanticism, in which *-lein* was used to produce a somewhat archaic or more folk-like sound that was suited, e.g., for distinguishing characters (Öhmann 1972:565, Pfennig 1904:15, Stricker 2000:250). This also reflected the situation on the eve of the 20th century. According to Wellmann (1975:131) the ratio between *-chen* and *-lein*, at that time, was approximately 4:1.

2.2 *-gen* suffix

The early literature saw *-gen* as a variant of *-chen*. For example, when Gürtler spoke about the guttural suffix, he usually referred to both *-chen* and *-gen* under the primacy of *-chen*; at times, he made reference to the regional background of *-gen* from Western Central German. While Ettinger (1980: 61) defined *-gen* as a sort of hypercorrection, Wegera (2000), in contrast, argued for the originality of the *-gen* suffix. Gürtler (1909a) concluded from his material that *-gen* was of higher frequency than *-chen* was around 1660, at least for writers from the East Central German region. Elspaß (2015: 400) could not confirm this finding. As mentioned above, there were neither *-gen* nor *-chen* variants found in newspapers from the 17th century. However, there were *-gen* variants in more informal writings of non-specialists (e.g., in private letters). Obviously, there was a difference in use amongst different varieties.

For the subsequent period, Schebben-Schmidt (1990: 314) found that there was a further clear dominance of *-gen* until 1750 in most regions. The only exception to this dominance was found in the writings of Low German, Saxonian, and Thuringian writers, as well as in those of Central Bavarian writers. Interestingly, this is not in line with the findings of Elspaß (2015: 401), who saw the relation between these two suffixes as rather balanced at this time. As there is no information about the quantitative properties of Schebben-Schmidt's corpus, a more detailed assessment of these results is not possible. Considering, however, that Schebben-Schmidt used 18 texts, there is very likely not much more data than what is provided by the GerManC newspaper corpus. The data set that Elspaß evaluated relied on 15 newspapers (Elspaß 2015: 396). It is thus likely that these results represent a difference in use according to genre. Finally, Elspaß (2015: 401, 2005: 344) found some regional *-gen* variants in another corpus of private letters from the 19th century.

2.3 Language authorities

More generally, a rather prominent question is what impact language authorities – i.e., grammarians or institutions of the 17th and 18th century – had on the replacement process (see Davies & Langer 2006; Öhmann 1972; Wegera 2000 for an overview). Davies & Langer (2006: 231–241) demonstrated that, beginning in the 16th century onward, grammarians weighed in on the correctness and use of diminutives. One area of focus was – stylistic acceptability aside – the question of euphony. Considering that there was a great deal of advice and many justifications for particular uses of diminutives, be they uses of *-lein* or of *-chen*, the impact of language authorities seems to be rather high. It is, indeed, easy to find coincidences

between codification and application. For example, Schebben-Schmidt (1990: 315) concluded that the choice of suffixes is mainly due to the offset consonant of the base lexeme: if the diminutive follows a base ending with /l/, *-chen* is preferred (e.g., *Seelchen* ‘soul.DIM’, rather than *Seellein*). Therefore, *-chen* serves in this data both as a morphological boundary marker and also as an object for euphonic appreciation (see Plank 1981; Stricker 2000: 227f.). In addition, *-lein* might be preferred for euphonic reasons after /ç/ or /x/ (Öhmann 1972: 557). In this data, the same applies in the use of epenthetic vowels (*-eli-*) or larger segments (*-el-*), as seen in *Bächelchen* ‘creek<el> DIM’ instead of *Bächchen* (Gürtler 1909a: 35). This corresponds with principles formulated by, e.g., Gottsched (1762) and others during the 18th century. Consequently, early literature has seen grammarians’ judgments as the crucial factor in the standardization process.

On the other hand, it is obvious that grammarians’ suggestions have been, on the whole, inconsistent (Elspaß 2015: 403) and sometimes hard to perform in practice, which is why in many cases the diminutives were used interchangeably (Gürtler 1909a: 36). With regard to this, Pfennig (1904: 8–9) also found that writers, especially from the Upper German region, used *-chen*, which they considered to be more prestigious for euphonic reasons. As Gürtler (1909a) showed, even prominent institutions such as the Silesian School of Poets (*Schlesische Dichterschule*) or the Fruitbearing Society (*Fruchtbringende Gesellschaft*) lacked any consistent pattern of use for diminutive suffixes. What this indicates is that, to a large extent, individual preference determined suffix use (see, e.g., Pfennig 1904 for a discussion on Schiller’s changing diminutive use across his lifetime).

Eckert (1986: 306), in addition, has made it clear that this holds not only for German diminutives, but also for French ones. She emphasized that the French grammarians’ specifications must essentially be seen as a reaction to already-established regional practices or stereotypes. This is true of the German situation as well (consider, e.g., the subtitle of Gottsched’s (1762) influential book, which referred to the example of prestigious writers). Voeste (1999) illustrated with her analysis of German adjective inflection that this might have been a part of larger trend. Davies & Langer (2006: 241) thus concluded that the standardization of diminutive suffixes was – contrary to more traditional assumptions – “not substantially influenced by grammarians.”

2.4 Genre and geography

Factors influencing diminutive use that consistently recur in almost all studies are genre and geography (see Appendix, Figure 12). Öhmann (1972: 556) summarized how, during the 16th century, *-chen* was used in the Central German region and in Northern Germany and was preferred in nonfiction, deeds, and court writings.

Fiction and religious writings, in contrast, tended to show *-lein*. By the 18th century, however, *-lein* had also become rare in fiction (Öhmann 1972: 558). Wegera & Solms (2002: 165) reported a preference of *-chen* within sermons, which seems to be rather contradictory to Öhmann's findings. However, these variants were stylistic ones that served a particular rhetorical purpose, as the authors mentioned. The same was true in the case of the abovementioned increase of *-lein* suffixes during the 19th century, which seem to have been preferred up to the current day in particular contexts (Fleischer & Barz 2012: 233).

Connected to the above-mentioned factors are differences in register and style. Pfennig (1904: 9) described how Schiller's father Johann Caspar differentiated types of diminutive according to stylistic levels. Stricker (2000: 385–387) reported similar observations about Goethe's diminutive use. As mentioned above, Elspaß (2015: 401) found a higher proportion of diminutives within less formal texts.

The geographical factor is reflected in all studies. "Geographical factor," in this body of research, refers either to the origin of writers being studied or to their orientation towards the prestigious writing of a particular region (for discussions on diminutives in dialects see Wrede 1908; Tiefenbach 1987). Given the dominance of *-lein* around 1600, it is obvious that this suffix went beyond the Upper German distribution area of dialects (Öhmann 1972). At the same time, *-chen* seems to have stemmed from the East Central German region, whose cultural center is Meißen. According to data from Schebben-Schmidt (1990) and Elspaß (2015), this was one of the first regions, if not the very first one, in which *-lein* had been lost; it had disappeared by 1750.

As mentioned above, *-gen* originated in the Western Central German area. From there it seems to have moved towards the east during the 15th and 16th centuries. After some time, it finally reached Upper German writers, who had until then had access only to *-lein* variants from their dialects. In the data of Schebben-Schmidt (1990) and Elspaß (2015), *-gen* dominated the Upper German and West Central German literature, but was also found with lower frequency in literature from Thuringia as well as in Northern literature; however, it was completely absent in texts from Saxony.

The geographical factor seems to have been important until around 1750. Schebben-Schmidt (1990: 315) concluded that the distribution of *-lein* and *-chen*, which had been primarily regional during the ENHG period, was no longer regionally-determined at this time (for more recent periods of time, see also Wellmann 1975: 131).

Setting this aside, some authors have proposed a north-south diminutive divide, with a higher frequency of diminutive use in the spoken language of southern German regions (e.g., Wrede 1908: 73, Schirmunski 1958: 235–237). There is,

however, no empirical data on this. With regard to the written language, neither Schebben-Schmidt (1990) nor Elspaß (2015) have provided any clear picture of this situation.

2.5 Some considerations for this study

On the one hand, the aforementioned studies have offered an impression of the general development of diminutives. On the other hand, the existing studies are rather difficult to compare to one another. For example, Gürtler's (1909a) finding of the increased usage of *-chen* between 1620 and 1660 – which is, for the author, a period in its own right – intersects the usual period boundary of the more recent studies, which is why the whole process is separated into different samples (e.g., 1600–1650 vs. 1651–1700). This makes comparison nearly impossible. Given the rather small proportion of *-chen* that Wegera & Solms (2002) were able to find, it is even unclear whether it really existed at that time (see also Wegera 1982:212). The present study thus aims to carry out a time-delineated evaluation of a large corpus that will also be able to capture fine-grained differences in time. Additionally, moving averages over time will be explored to find more general trends.

Furthermore, the existing corpus-based studies have usually evaluated periods of 50 years each, with the boundaries being more or less arbitrary. But what makes, e.g., a period from 1700–1750 more worthy of examination than a period of, say, 1715–1765? Again, Gürtler's observation provides a good example of the uncertainty that results from pre-defined period boundaries. The present study thus comes without any pre-determined periodization and instead tries to derive information regarding significant change points from the data in a bottom-up approach. In addition, even though the actual change obviously happened during the 18th century, in order to adequately place the results in time, both the preceding and the subsequent centuries are considered.

More recent studies have relied on very small data sets. For example, the GermanC newspaper corpus, which has been used in the most recent studies in the field (Elspaß 2015), reveals ca. 70 hits over 150 years; this is, according to Elspaß (2015:397), rather small for grammatical analysis. The present study thus relies on what is currently the most comprehensive corpus of the NHG period – namely, the Deutsches Textarchiv (DTA). From this corpus, a total of more than 110,000 diminutive hits between the years 1600 and 1900 has been retrieved. By doing this, a much larger sample can be explored in this study than the samples that have been used in the past, allowing a more precise identification of the actual transition from one suffix to the other.

Ettinger (1980:82) has criticized the fact that existing studies have mostly focused on tokens, not on types. Indeed, excluding Wellmann's (1975) and

Stricker's (2000) analyses, to date there has been no study on German diminutives that has considered the frequency of types. To the best of my knowledge, a study that considers a particular type-token ratio does not exist. This is unfortunate, as establishing the ratio of types and tokens may illustrate information about the diversity of language use (e.g., Wimmer 2005). Consequently, by exploring both types and tokens, this study sheds new light on the *-lein/-chen* replacement. This in turn helps to document the historical processes more precisely.

3. Materials and methods

3.1 The DTA corpus

The DTA corpus is a representative corpus of written NHG from the 15th to the 20th century (see Haaf & Thomas 2016). It thus perfectly covers the time of the *-lein/-chen* replacement. At the time of analysis (January 2017), the DTA contained 2,594 texts with a total of 142 million words. The corpus is partitioned in four main genres: fiction; nonfiction; scientific literature; and newspapers. However, text from newspapers is not continuously present in the corpus, which is why this genre has been excluded from this analysis. The other genres are roughly uniform in their frequency of appearance. Furthermore, texts before 1600 are at present only rarely documented; this is why they are excluded as well. The same applies to texts after 1900, which are also excluded. Altogether, this leads to a corpus of 2,441 texts.

The diminutives are not included in the POS tags of the DTA, which is why they needed to be searched individually. This required that certain considerations needed to be taken into account. For example, *-chen* – which is the more problematic suffix – shows a segmental identity with the suffix *-en* (NOM/DAT/GEN/ACC.PL) after <ch> and <sch>, which represent /x, ç/ and /ʃ/ respectively (e.g., [^]*Ursachen* 'cause.PL', [^]*Leichen* 'cadaver.PL', [^]*Wünschen* 'wishes.PL.PL').¹ It further coincides, e.g., with <schen> ending nouns (e.g., *Groschen* 'penny') and inflection of <sche> ending nouns (e.g., *Flasche-n*, 'bottle.NOM/GEN/DAT/ACC.PL'). There are also cases of absolute homography ([^]*Eichen* 'oak tree.PL' but also 'egg.DIM.SG/PL'), which make the search more complicated.

Another problem arises with the extant DTA POS tags (STTS), which, in principle, could be used to facilitate the search. However, in some cases they are set incorrectly, which is why they cannot be used. Take, for example, the *-isch* suffix, which is typically used to derive adjectives from nouns and formally coincides under certain circumstances with *-chen*. Unfortunately, many of these

1. The carat (^) is used to indicate mismatches.

adjectives are tagged as nouns in the DTA (*dem* ^*Englischen Handel* 'the.DAT.SG English.DAT.SG trading.DAT.SG'). Thus restricting the search to nouns using the POS leads to highly disorganized data.

Instead of that, another approach is taken. Given that there is a strand of literature that focuses on periods before 1600, the idea is to connect the corpus to this body of literature by taking the most comprehensive compilation of diminutives of the preceding period as a sub-corpus. This compilation, provided by Gürtler (1909b), contains ca. 300 *-chen* diminutives from the ENHG period. The very few loanwords and names that can be found in this corpus were removed from the list, and all of the simplicia and heads of compounds act as the initial input for the present study.

The choice to exclude names needs some explanation. The diminution of proper names is more complex than the diminution of appellatives, as there are particular synthetic means for the expression of emotive relations that need to be looked at separately in a discussion on diminutives (e.g., *Anna* vs. *Ann-i* vs. *Anna-lein*). Furthermore, family names and toponyms in particular show many idiomatized diminutive forms (e.g., the family name *Schmidtgen* 'smith.DIM') with a very strong regional influence (see Schmuck 2009 regarding the northern *-ing* diminutive). Consequently, diminished names require a separate methodological approach. In this way, exclusion of names is in line with Dressler & Barbaresi (1994: 84). A more practical reason for the exclusion of names is due to the fact that the available DTA POS tags for names (NE) are to a large extent either incomplete or inaccurate, which is why an individual corpus for both historical and more recent names and name variants must be consulted. However, there is no such corpus currently available.

A second set of data aims to illustrate the period following the intended time span, i.e. from 1900 onward. For that period, the compilation had to be constructed from scratch. This study used the Digitales Wörterbuch der deutschen Sprache (DWDS) as a corpus. The DWDS offers a data service called dlexDB (<http://www.dlexdb.de>). First, all *-chen* and *-lein* ending nouns were extracted from the DWDS. Due to the aforementioned idiosyncrasies in the data, this was especially unsatisfying for the *-chen* suffix. However, a large amount of non-diminutive *-chen* ending forms were able to be excluded via a syllable bound disambiguation of the pattern <schen> (e.g., *Gro-schen* vs. *Mäus-chen*, 'penny' vs. 'mouse.DIM'). As dlexDB – in contrast to the DTA – is sensitive to syllable structures, disambiguating these could be rather easily taken care of in the second step. After this process, instead of ca. 450,000 *-chen* (mis)matches in the DTA, there were ca. 9,000 matches in the DWDS, most of them compounds of only some hundred nouns. However, these matches had to be verified and, where necessary, eliminated. This verification was assisted by R programming, but was in essence

conducted manually. The most frequent bases were then compared to the historical list derived from Gürtler (1909a) and added if not already given. Finally, the compilation was supplemented with examples from the prescriptive linguistic discussions of the time (e.g., Schottel 1641: 328–329). Through this process, a list of, at present, 460 bases was compiled to serve as the starting point for the actual DTA data exploration.

3.2 Data exploration

Data were gathered via web scraping. To do this, orthographic and phonological variation had to be considered. For *-lein*, this was, e.g., <elein, elin, lin, lein>; for *-chen*, <ichen, ichin, chin, chen>. The same was also true for the bases, for which other orthographic variation also had to be taken into account: e.g., the contemporary standard German grapheme <t> varied with historical allographs <t, dt, tt>; <u, v> varied with <u, v, w>; <ß, ss> varied with <s, ß, sz, ss>; <ä> varied with <ä, ae, e> and, depending on the particular lexeme, <a>. Furthermore, inflection had to be considered. To some extent, this variation was already kept by the DTA, which is why it was possible, in principle, to only search for a single type of word. The results, however, were not satisfying. An example is provided by a search of $[[\text{Blatt}]_{\text{LEIN}}]$ ‘leaf.DIM_{LEIN}’. This was done via the regular expression in (1).

- (1) $[[\text{Blatt}]_{\text{LEIN}}] \rightarrow /^{[Bb]l(a|ä|ae|e)(t|tt|dt|d)(el|l)(ei|i|y)n(s?|e?|en?)\$/$

Among others, the following *-lein* diminutives showed up:

- (2) *bletlein* (Deckhardt 1611), *Blätlin* (Andreä 1616), *Blädtlein* (Silber 1618), *Blättlin* (Wild 1631), *Blättlein* (Corvinus 1715), *Blätlein* (Gladov 1727), *Blättelein* (Grimm 1812)

Altogether, this search returned a total of 894 hits. In contrast, the search for the word type “Blättlein”, which also takes into account orthographic variation and inflection (Jurish 2013), revealed only 808 hits. By controlling these hits, it became evident that the search conducted by the DTA considered, e.g., allographs <t, tt, dt>, but not umlaut variation <e>, the *-lin* variant, or Schwa epenthesis *-elein*. The problem is that these phenomena are typical for the early phase of the sample period. Their exclusion leads to a selection bias. For that reason, the regular expressions search is preferred over the search for word type commonly suggested in the DTA. Consequently, the regular expression terms must be adjusted for every lexeme.

The web scraping made use of the DTA interface (<http://kaskade.dwds.de>), which enables the gathering of raw data per annum. After correcting the information, a total of 114,087 diminutives (tokens, no compounds) for 460 bases were

returned. Table 1 summarizes both absolute and relative frequencies for the four relevant suffixes extracted from the DTA. Over the whole period *-chen* is obviously the most frequent suffix, followed by *-lein*. Decidedly lower is the frequency for *-gen* and another suffix, *-elchen*, that can be interpreted either as a linkage of *-lein* and *-chen* or as a combination of *-chen* with a linking element *-el* (Davies & Langer 2006: 238–239). In Stricker's corpus this *-elchen* suffix accounted for 0.40% of all documented diminutives (Stricker 2000: 230). This is comparable to the present corpus, where *-elchen* accounted for 0.59% of diminutives.

Table 1. Occurrences of diminutives in the DTA corpus

	Absolute frequency	Relative frequency (per million tokens)
<i>-lein</i>	49,796	348.44
<i>-chen</i>	58,407	408.70
<i>-gen</i>	5,209	36.45
<i>-elchen</i>	675	4.72
[Total N = 114,087]		

Certain diminutives have become lexicalized over time. This is seen in the present corpus for *Fräulein* ('woman.DIM' > 'honored dame' > 'mademoiselle'), *Bisschen* ('bite.DIM' > 'a smidge' ADV), *Mädchen* ('maid.DIM' > 'girl'), and *Märchen* ('narration.DIM' > 'fairy tale'), which were certainly diminutives around 1600, but which had become lexemes without diminishing connotation by around 1900 (Erben 2006: 95). For the sake of consistency, the data distribution without these lexicalized items is reported. Also excluded from further discussion are the mostly phonologically-motivated *-elchen* diminutives that are obviously marginal. Not affected by this decision are those words with an *-el* ending base (e.g., *Vogel* 'bird'). In doing so, a total of 50,119 hits serve as the basis for further analysis. These hits were found in 910 texts (i.e., 37.28% of all DTA texts) written by 525 authors.

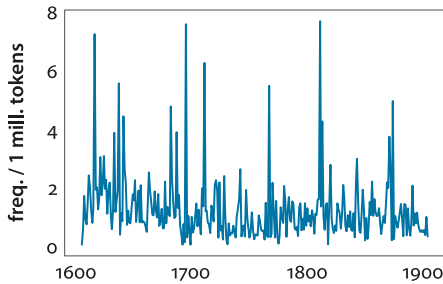
To carry out a preliminary investigation of the corpus' significance, the data were compared to Wegera & Solms's (2002: 164) quantification of the FnhdC; this study features, up to date, the largest quantification of diminutive suffixes. Among all diminutives in their sample, they found 6.8% *-chen* suffixes between 1650 and 1700 (other suffixes were not reported). Limiting the sample of the present corpus to the same period, there are 5.2% *-chen* suffixes, which is fairly comparable to the counts in the FnhdC. The DTA corpus hence offers a suitable link to other studies.

4. Diminutives in the New High German period (1600–1900)

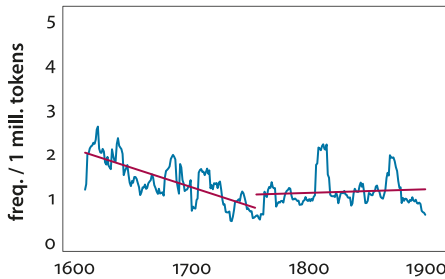
4.1 Frequency of diminutives

4.1.1 Overall pattern of tokens

Figure 1 shows the occurrences of tokens between 1600 and 1900. Figure 1a depicts the raw signal, with a mean of 1.2 diminutives per 1 million tokens; this qualifies the phenomenon as a rather rare one in the DTA corpus. Figure 1b shows, in contrast, the simple moving average (SMA) over periods of 10 years each. SMA serves to reduce the random fluctuations (drifts) that appear, such as if a certain lexeme is used disproportionately often in a single book. SMA is used here and in the following analyses with a moving window of 10 years' length for the estimation of trends.



a. Occurrences over time



b. Simple moving average of order 10

Figure 1. Diminutives within the DTA corpus (tokens)

In this case, within the estimated distribution, two individual phases can be identified. The first one lasts from 1600 to roughly 1750. The second phase lasts from roughly 1750 to 1900. These two phases are represented as gray lines in Figure 1b. There is a statistically significant difference between both the variances and the means of these groups ($F(1,289)=17.986, p < .001$; $t(270.38)=5.295$,

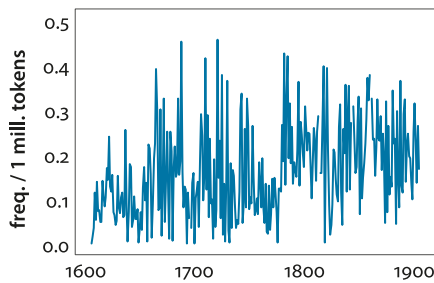
$p < .001$). Testing the two groups separately reveals the first component to be non-stationary with a clear downward trend ($R^2(144) = .572$, $p < .001$) while the second component is stationary without any significant trend (1600–1749: $KPSS(146) = 2.937$, $p < .01$ vs 1750–1899 $KPSS(145) = 0.286$, $p > .1$).² In other words, the frequency of diminutives is significantly reduced until 1750 and, from then onward, stable over time. This confirms Polzin's (1901:107) assumption about a general decrease of diminutive use during the 17th century. The second phase is characterized by the fact that, even if deviations from the mean exist in forms such as individual spikes in diminutive use, the distribution always returns to the mean.

4.1.2 Overall pattern of types

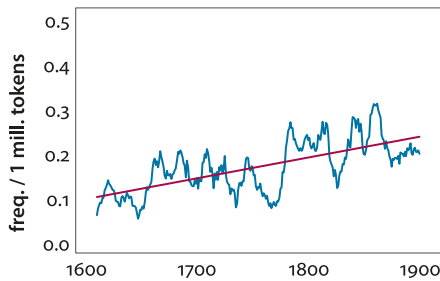
As discussed above, the literature usually makes no distinction between tokens and types. Adding to the already-performed exploration of tokens, Figure 2 illustrates a particular pattern for the time series of types. The estimated distribution is non-stationary ($KPSS(291) = 3.546$, $p < .01$) with a significant linear trend upwards ($R^2(289) = .455$, $p < .001$). A significant correlation between the distributions of types and tokens is not found ($r(289) = -.099$, $p = .090$). With regards to Polzin's (1901:107) assumption about the decrease in diminutive use, there is now a more differentiated picture. While types decrease during the first half of the 17th century, confirming Polzin's hypothesis, they increase around 1650 to a much higher level.

Figure 3 provides a sharper picture of the relationship between types and tokens by depicting the cumulated type-token ratio (TTR) for all diminutives per year. In this context, TTR is calculated as the ratio of the number of tokens to the number of types for all authors per year. There are four phases: (i) a relatively stable situation until the middle of the 17th century (1647); (ii) an increase in the TTR until the middle of the 18th century (1762); (iii) a short period of stabilization (1776); (iv) a second sequence of TTR growth. Both increases indicate that the development of diminutive use is going in a new direction. Considering that TTR is traditionally interpreted as a measure of lexical diversity (e.g., Wimmer 2005), phases (ii) and (iv) indicate a broadening of access to the lexicon. This does not necessarily imply higher productivity of the suffixes in the sense that they are now used for new words. Instead, Figure 3 indicates that there is a greater use of diminutives for different concepts, regardless of whether they are new concepts or not. However, as will be shown later, the diminutive suffixes have in fact expanded

2. The concept of "stationary data" is crucial for the analysis of time series. Stationary data exists without trend and periodicity. Non-stationary data is data that, in contrast, has a significant trend or periodicity over time. To determine whether a time series is stationary or not, one can perform the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS, Kwiatkowski et al. 1992).



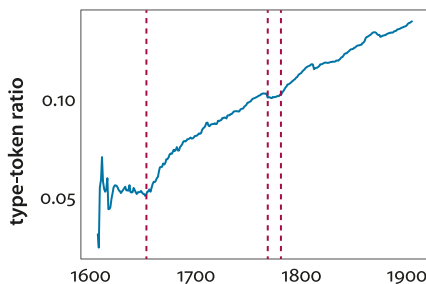
a. Occurrences over time



b. Simple moving average of order 10

Figure 2. Diminutives within the DTA corpus (types)

to new semantic sectors and pragmatic contexts, which is a type of domain-specific productivity.³ This enlargement had obviously already begun in the middle

**Figure 3.** Cumulated type-token ratio of diminutives in the DTA corpus

3. Regarding morphological productivity, there are more fine-grained and effective measures available which are typically based on the frequency of hapax legomena (e.g., Baayen 2005; Gaeta & Ricca 2006). As this is a controlled data collection and not a randomly sampled one, it is problematic to perform measures that are based on hapax legomena. Moreover, as this

of the 17th century when the decrease of tokens was still ongoing. As will be shown later, this coincides with an increase of diminutive use in nonfiction.

4.2 Individual suffixes

Differentiation of tokens

Figure 4 shows the occurrences of the suffixes *-lein*, *-chen*, and *-gen* over time. Figure 4a depicts the time series of the tokens, while Figure 4b shows the results of a change point analysis. The individual suffixes show very clear trends. While *-lein* has a more or less continuous downward trend, *-chen* follows a steep S-shaped upward trend, which is the typical curve for language change (see Labov 1994: 65–66, Blythe & Croft 2012). In addition, *-gen* demonstrates a quick rise and a step-by-step fall, which identifies it as a temporary word-formation particular to the 18th century.

Furthermore, it becomes clear that the downward trend of *-lein* in Figure 4a is, in its distribution up to 1730–1740, quasi-identical with the first part of the distribution in Figure 1b. However, this is a rather trivial point, as *-lein* is used nearly exclusively over the course of the 17th century. As expected, *-chen* becomes the dominant diminutive form beginning in the middle of the 18th century (Schebben-Schmidt 1990), even if it does not reach the frequency *-lein* had in the beginning of the 17th century when it was the predominant suffix. Consequently, Figure 4 makes clear that phases (ii) and (iv), depicted in Figure 3, reflect contradictory situations. Phase (ii) is mainly driven by the increase of individual *-lein* types that are, at the same time, decreasing as tokens (see Figure 2). That is, during the process of reduction, the *-lein* suffix is at least temporarily applied in fresh contexts. In contrast, phase (iv) reflects to the proliferation of *-chen*, which is a more concise process.

A more concrete phasing is provided in Figure 4b. Blue lines are standardized frequencies from Figure 4a. Red lines depict the results of a multiple change point analysis. For the detection of change points, the binary segmentation algorithm (Edwards & Cavalli-Sforza 1965) provided by Killick & Eckley (2014) was used. Similar to the process of analysis of the patterns seen above, this approach was based upon changes in variance. The analysis is therefore not primarily looking for individual jumps in the data, but for time sequences of similar vari-

corpus is a selection of types that, to a large extent, are already known from earlier periods, the interpretation of hapax legomena in terms of their status as innovative forms is inappropriate. Consequently, statements concerning the productivity of the suffixes are beyond the scope of this study. However, it is appropriate to focus on the lexical diversity, which is the reason why the TTR measure is used.

ability that are different from subsequent sequences with differing variability. The maximum number of change points was restricted to 10 points. In no case, however, was this maximum reached.

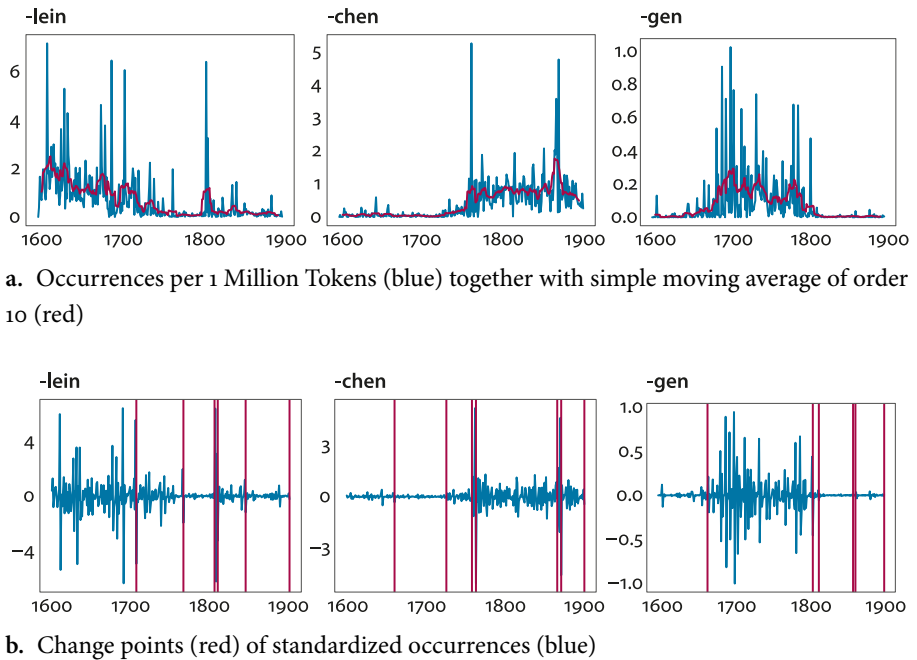


Figure 4. Suffixes *-lein*, *-chen*, *-gen* from 1600–1900 within the DTA corpus (tokens)

The first change point for *-lein* occurs in the year 1706.⁴ At this point, the dominance of *-lein* is clearly reduced. It is followed by a period with further reduction until 1765. Two further change points come shortly after 1800 (1805, 1809). Most interestingly, this change indicates a slight renewal of *-lein*. As will be shown later, this coincides with a separation of genres. Another change point follows in 1843 with a further increase in the occurrence of *-lein*. The last change point is an artifact indicating the end of the time series.

The first change point for *-chen* is in 1660. Until then, there is clear variability in the data. Subsequently, the distribution is more uniform until the second change point in 1725, at which point *-chen* use begins to rapidly increase over circa 20 years. Between 1757 and 1762, there are two further change points, after which

4. It should be emphasized that the explored year specification applies only for the present data corpus. It can be assumed, however, that in other corpora with a similar amount of data, the change point is roughly around 1706. In addition, this is true for all further year specifications.

the use of *-chen* is stabilized in a longer phase, during which it has a higher frequency of use. This lasts until 1867, at which point two change points are once more aligned. Afterwards, the frequency of *-chen* is reduced. Unlike for *-lein*, there is no coincidence between the individual change points; the only exceptions are the change points in 1757/1762, which coincide fairly well with the change point found for *-lein* in 1765. This near-coincidence is remarkable, as it indicates the preliminary peak of *-chen* propagation, as well as the preliminary low level of *-lein* reduction.

Regarding previous research, it is most notable that the period between the second and the third/fourth change point for *-chen* (= 1726–1757/62) perfectly aligns with the prediction in Gürtler (1909a: 33f.). In addition, the previous finding of relations changing from the first to the second half of the 18th century (Schebben-Schmidt 1990) is confirmed. Additionally, the renewal of *-lein* has been established in the literature. Taken as a whole, however, this renewal is not as intense as it is usually assumed. Davies & Langer (2006: 240) put it aptly in classifying it as a “minor hiccough.”

Furthermore, special mention should be made of the fact that use of *-gen* clearly begins to occur before the beginning of the *-chen* propagation. The propagation of *-gen* began as early as 1666. Even if the frequency of its use remained rather small, it still significantly exceeded the amount of *-chen* use at this time. Focusing on the change points at which the use of *-gen* and *-chen* began to increase, i.e., 1666–1725, a *t*-test between the two suffixes reveals a significant difference in the mean amount of use ($t(59.434) = 4.032, p < .001$). This was true, however, only for this period. Two change points shortly after 1800 (1804, 1812) indicate, finally, that the suffix was nearly lost, although there was more use around 1858 as indicated by two other change points.

This raises the question as to what extent the three suffixes are interrelated. Considering that (i) *-gen* had appeared earlier and (ii) there was a slight difference between *-lein* minimum use and *-chen* peak use, there might have been a delayed positive correlation between *-chen* and *-gen* and a delayed negative correlation between *-lein* and *-chen*. In order to verify this, a sample cross correlation analysis (CCF) was performed. The basic idea behind CCF is to produce a set of correlations between two time series that is suitable for exploring not only an overall sample correlation, but also possible delays in correlation. For the current time series of 300 time units (years), cross correlations for 300 maximal lags had to be checked. Negative coefficients therefore refer to correlations before a certain time; positive coefficients refer to correlations after a certain time. Table 2 reports the results at lag 0, which represents the overall correlation.

Table 2 highlights a weak negative correlation between *-lein* and *-chen*, which is the expected result given the contrast between the trends in Figure 4a. Explor-

Table 2. Correlations between individual time series at lag 0 (tokens)

	<i>-lein</i>	<i>-chen</i>	<i>-gen</i>
<i>-lein</i>	1		
<i>-chen</i>	-.235 ^{***}	1	
<i>-gen</i>	.050	-.101	1

*** = $p < .001$; ** = $p < .01$; * = $p < .05$

ing the remaining lags, there is an even stronger correlation of $r(298) = -.307$, $p < .001$ at lag -5 , indicating that *-chen* leads *-lein* with a delay of five years. This means that there is a significant tendency for changes in *-chen* use to be followed by changes in *-lein* use five years later. This is almost exactly the pattern yielded by the change point analysis, which returned a delay of three to eight years. The data show no similar relationship amongst the other suffixes. The assumed delayed correlation between the frequencies of *-chen* and *-gen* is not statistically present in the analysis, which provides evidence that *-gen* is an original suffix (Wegegera 2000).

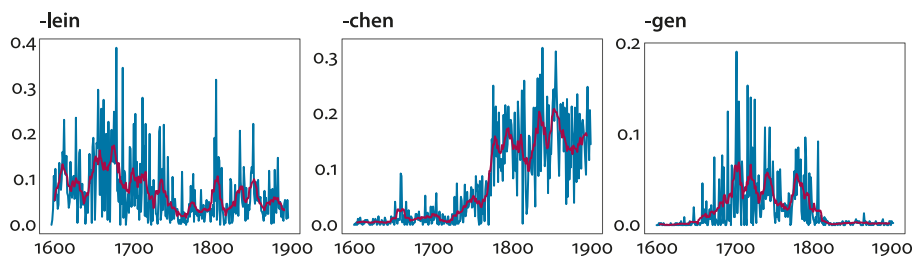
4.2.2 Differentiation of types

Figure 5a illustrates the distribution of types. At first glance, it is evident that there are patterns in the distribution of types similar to the patterns in the distribution of tokens, although they differ in detail. A correlation analysis between types and tokens reveals a moderate correlation between types and tokens for *-lein* ($r(298) = .591$, $p < .001$) and for *-chen* ($r(298) = .659$, $p < .001$), while there is a strong correlation for *-gen* ($r(298) = .820$, $p < .001$). In the CCF, there is no delay in the data. The highest correlations are for all suffixes at lag 0.

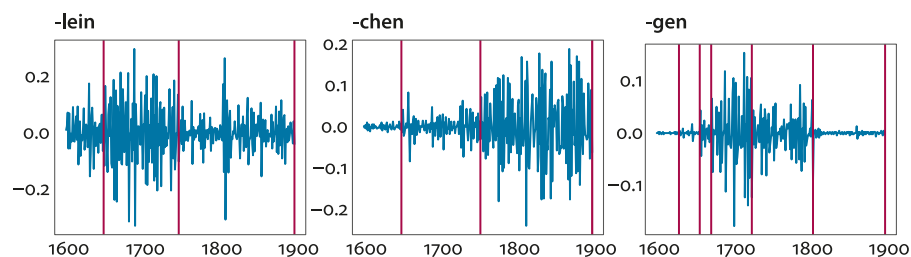
Furthermore, there are almost identical change points for *-lein* (1650 vs. 1747) and *-chen* (1650 vs. 1752). Obviously, between 1650 and, say, 1750, there was a period in which the *-lein* suffix was used for successively fewer and fewer lexemes; meanwhile, *-chen* was largely stable, though with some amplitudes and a considerable increase at the end of this period. Interestingly, there were many fluctuations in the *-lein* distribution. Particularly between 1640 and 1700, there was a considerable increase in types that was not reflected in the distribution of tokens. It seems that the authors, even as the overall use of this suffix was decreasing, were seeking new contexts for its application.

As Table 3 indicates, the two distributions were not significantly correlated at that time. There was only a significant correlation between *-lein* and *-gen*, which was merely due to some similarities during the 18th century.

Nevertheless, the *-lein* expansion around 1640 was a more general process. This is underlined by the fact that there was an identical change point for *-chen* in 1650 and similar change points in 1629 and 1656 for *-gen*. As seen above, this was



a. Occurrences per 1 Million Tokens (blue) together with simple moving average of order 10 (red)



b. Change points (red) of standardized occurrences (blue)

Figure 5. Suffixes *-lein*, *-chen*, *-gen* from 1600–1900 within the DTA corpus (types)

Table 3. Correlations between individual time series at lag 0 (types)

	<i>-lein</i>	<i>-chen</i>	<i>-gen</i>
<i>-lein</i>	1		
<i>-chen</i>	-.058	1	
<i>-gen</i>	.259***	-.111	1

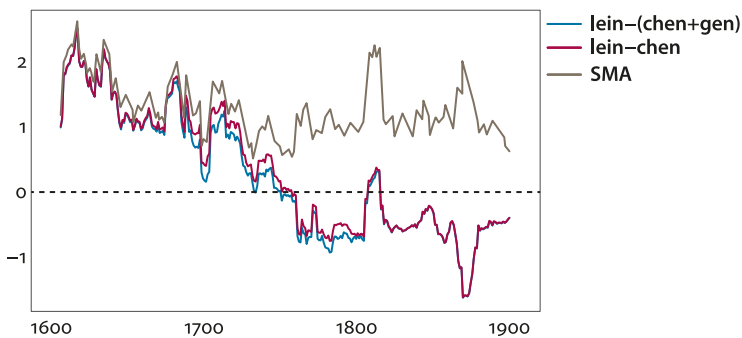
*** = $p < .001$; ** = $p < .01$; * = $p < .05$

approximately the starting point of *-gen*'s rather short existence, while for *-chen* it was something like a short prelude before the acceleration in use one century later. As for between the two *-chen* change points in 1650 and 1752, once more there was a significantly higher mean for use of *-gen* than there was for use of *-chen* ($t(64.772) = 4.199, p < .001$).

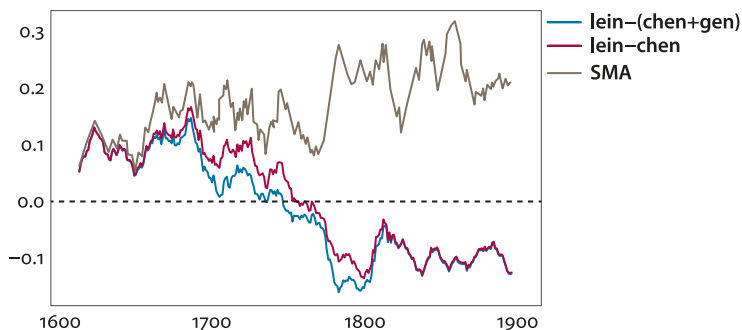
In light of previous studies, these findings are of particular interest. As mentioned above, Gürtler (1909a) identified an increase of guttural suffixes between 1620 and 1660 that had not been confirmed by more recent studies. The change points in 1629 and 1656 for *-gen*, found in this analysis, were fairly similar to those identified by Gürtler. Furthermore, there was a very short phase of increased *-chen* types between 1647 and 1663 that also lined up with Gürtler's findings. The fact that this had not been found in other studies is due to, first, the recent literature's primary orientation towards tokens and, second, the common tendency to prede-

fine long-lasting periods. This increase must be specified as a result of an increase of types. Whether it is a transition period of its own, as Gürtler claims, is another question. It rather seems to be an early spike in use, as there are other such spikes, although smaller ones.

Figure 6 summarizes this situation. Figure 6a represents tokens, Figure 6b types. For every year, the difference between the SMA data for *-lein* and the SMA data for *-chen* and *-gen* is presented. The red line shows the difference between the main suffixes *-lein* and *-chen* ($\text{Diff} = \text{SMA}_{\text{lein}} - \text{SMA}_{\text{chen}}$). The blue line shows the difference between *-lein* and *-chen/gen* ($\text{Diff}^* = \text{SMA}_{\text{lein}} - \text{SMA}_{\text{chen}} - \text{SMA}_{\text{gen}}$). Positive values indicate prevalence of *-lein* and negative values indicate prevalence of *-chen* and *-gen*, respectively. The horizontal line is the zero difference indicating a balanced proportion of the suffixes, while the beige line is overall SMA of types and tokens.



a. Types; positive values indicate dominance of *-lein*, negative values indicate dominance of *-chen/gen*



b. Tokens; positive values indicate dominance of *-lein*, negative values indicate dominance of *-chen/gen*

Figure 6. Difference of moving averages

Figure 6a (tokens) shows a downward trend with a somewhat stable situation in the second part of the period under discussion. Figure 6b (types) shows, in contrast, an S-shaped curve. The estimated year of transition is 1754 for tokens. There are two estimated years of transitions for types: 1746 for the relationship between the main suffixes *-lein* and *-chen* and 1738 for the relationship amongst all three suffixes. In light of all of this, the particularity of the *-gen* interlude also becomes obvious. Around 1700 this suffix is not only more frequent but also has a wider range of application.

4.2.3 Periodization of the replacement

Based on the previous findings, the process of replacement can be specified more clearly. In Table 4, six subsequent periods are defined on the basis of the change points analysis of tokens. For each period, the proportions of suffixes within these periods are reported. While tokens refer to the frequency of usage, types refer to the variety of usage. Period 1 and Period 2 refer to the phase of *-lein* dominance; Period 4 and Period 5, in turn, refer to the phase of *-chen* dominance.

Table 4. Periods of diminutive replacement according to change point analysis of DTA data

Period	Duration in years	Proportion of tokens			Proportion of types			
		<i>-lein</i>	<i>-gen</i>	<i>-chen</i>	<i>-lein</i>	<i>-gen</i>	<i>-chen</i>	
1	1600–1660	61	90% (4,019)	1% (43)	9% (412)	67% (225)	6% (21)	26% (88)
2	1661–1725	65	87% (9,360)	11% (1,157)	3% (293)	48% (267)	33% (184)	19% (107)
3	1726–1762	37	40% (1,671)	14% (597)	46% (1,929)	35% (167)	28% (136)	37% (177)
4	1763–1812	50	20% (2,384)	9% (1,090)	71% (8,542)	33% (228)	23% (155)	44% (302)
5	1813–1843	31	23% (1,667)	<1% (16)	77% (5,685)	36% (152)	1% (6)	63% (265)
6	1844–1900	57	19% (2,192)	<1% (29)	80% (9,030)	40% (209)	2% (10)	58% (304)
			[N=50,119]			[N=3,003]		

Period 1 starts from the documentation beginning in 1600 and describes the linguistic situation at the final stage of the ENHG period. In Period 1, *-lein* was the most frequent suffix. However, in this period, there was already a relatively high amount of *-chen* types. In other words, there was already a high variety of *-chen* usage (types) in this period, even though the frequency of usage (tokens) was rather small.

Period 2 starts with the first change point of *-chen* in 1660; this change point represents a reduction of *-chen* frequency. In contrast, *-gen* became more frequent

at this point. Most noticeable is the types frequency; every third type is a *-gen* suffix. The *-lein* suffix was, in contrast, slightly reduced in frequency.

Period 3 starts with the next *-chen* change point in 1725; this was when the actual *-chen* propagation started. This period describes the core of the development until the change point in 1762, at which point the *-chen* increase ended. Interestingly, there was at this time an almost balanced frequency between *-lein* and *-chen*. Nevertheless, in this period, the estimated transitions from *-lein* to *-chen* (as in Figure 6) occurred. There was the highest proportion of *-gen* in the DTA history at this time.

In Period 4, the transition was finished and *-chen* had become the most frequent suffix. The *-gen* suffix was successively reduced. For *-lein*, there was a situation similar to the situation for *-chen* in Period 1: a small percentage of tokens, but a clearly higher number of types. The reduction of *-lein* in the second half of the 18th century was, first of all, a reduction in the frequency of usage (tokens), not a reduction in the variety of application (types). Most notably, the amount of *-lein* types (33%) and the sum of *-gen* and *-chen* types (67%) is strongly in line with the findings of Wellmann (1975:133) regarding the writings of renowned authors around 1800 (*-lein*=31%, *-chen*=69%, as drawn from book indices).

Period 5 starts with the change point in 1812. In this period, *-gen* was almost lost, while a slight renewal of *-lein* took place. For *-chen* there was an increase of types.

Period 6 starts with the change point in 1843 and lasts until the end of documentation in 1900. As a whole, this period saw a stabilization of the situation of Period 5.

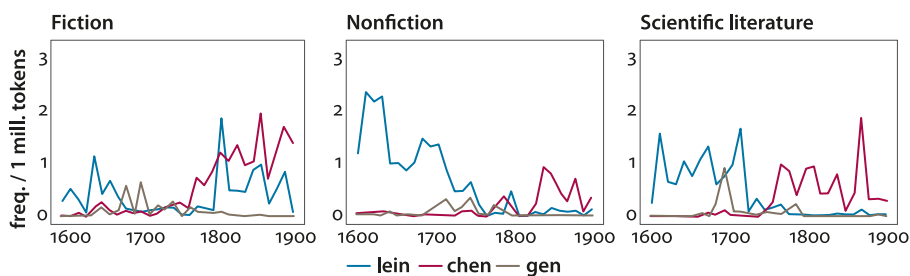
4.3 Genres

This section aims to further differentiate the replacement process. In order to obtain more information on the domains of usage, the frequency of diminutive suffixes within particular genres is analyzed. First, three superordinate categories are separated from one another: fiction, i.e., literature that serves as amusement and entertainment, such as novels, narrations, and drama; nonfiction, i.e., literature without explicit literary objectives or demands, such as specialists' literature, reports, guidebooks, and devotional literature; and scientific literature from both humanities and natural sciences. Needless to say, this is a rather rough categorization that, however, is suited to providing information about general strands of writing. Moreover, this classification is essential, as the DTA documents only a few subgenres over the whole period – e.g., prose, theological writings, or poems. These subgenres are considered in a second approach. The individual subordinate

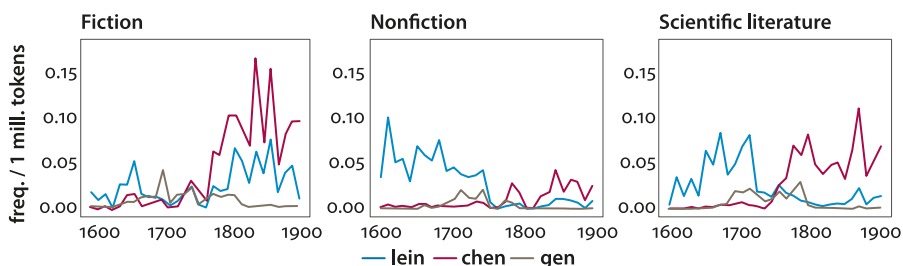
categories are defined by the DTA editors and available in the XML raw data. In the following, terms as such biology, edifying literature, lyric poetry, and so forth are drawn from the DTA terminology.

4.3.1 Overview

Due to the partitioning of data into three categories, an annual documentation is no longer possible. Thus, the data is re-sampled into periods of ten years. Figure 7 provides an overview of the distribution of the diminutive suffixes according to the three superordinate genres in time.



a. Suffixes within genres (Tokens)



b. Suffixes within genres (Types)

Figure 7. Distribution of suffixes *-lein*, *-chen*, *-gen* within the DTA superordinate genres fiction, nonfiction, scientific literature; intervals of ten years' length

The first result clear from Figure 7 is that there is a strong accordance between types and tokens. This is to say that within the genres, as expected, an increase of types is connected to an increase of tokens and vice versa. Second, there are different patterns of use for diminutives within the genres. While in more recent times, fiction has been the most preferred genre for diminutives, they are shown to be rather rare within nonfiction. This is in contrast to earlier periods, in which nonfiction was the most important domain for diminutives. This is especially true

for the first half of the 17th century. The most important turn is obviously around 1760, which is in accordance with the formerly discussed changes; in more detail, it is the end of Period 3, which is the period of the actual replacement. Therefore, there is no doubt that the replacement of *-lein* proceeded as a transfer between the genres, which indicates that the replacement was a stylistic one. This was no less than a fundamental pragmatic re-orientation of the DTA authors. Some particularities of usage will be highlighted in the following paragraphs.

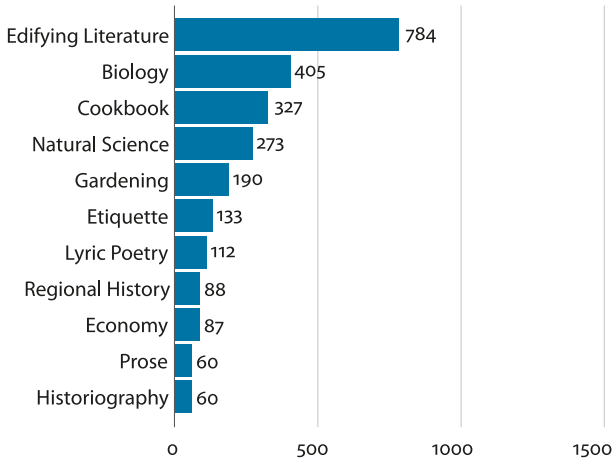
4.3.2 *-lein*

The situation for the *-lein* suffix is a very dynamic one. At the time at which *-lein* was almost exclusively documented (Period 1–2), it was mainly used in nonfiction and scientific literature. More precisely, until 1640, nonfiction was its predominant domain. The most important texts in this time for its use were funeral sermons, followed by rather general theological texts. Typical examples of its use include *Kindlein* ‘child.DIM’ and *Töchterlein* ‘daughter.DIM’, with reference to deceased persons; more general examples include *Sprüchlein* ‘saying.DIM’, *Wörtlein* ‘word.DIM’, *Büchlein* ‘book.DIM’, and *Stündlein* ‘hour.DIM’. From that point onward (Period 3), there was a clear reduction of *-lein* within these genres. It was subsequently the botanical literature that showed the most *-lein* diminutives (e.g., *Blümlein* ‘flower.DIM’, *Blättlein* ‘leaf.DIM’, *Tröpflein* ‘drop.DIM’, *Ästlein* ‘branch.DIM’). Regarding tokens, it is also evident that *-lein* was represented more or less consistently widely in scientific literature between Period 1 and Period 3. This is interesting from the perspective that – in contrast to fiction, where diminutives often express contextual dependent emotion – this is the genre in which diminutives are probably most strictly used with their prototypical meaning: the indication of smallness. Furthermore, regarding the types, it is evident that the number of *-lein* diminutives even increased within Period 2 and Period 3, which provides evidence argument that there was an orientation toward the more prototypical meaning before the replacement.

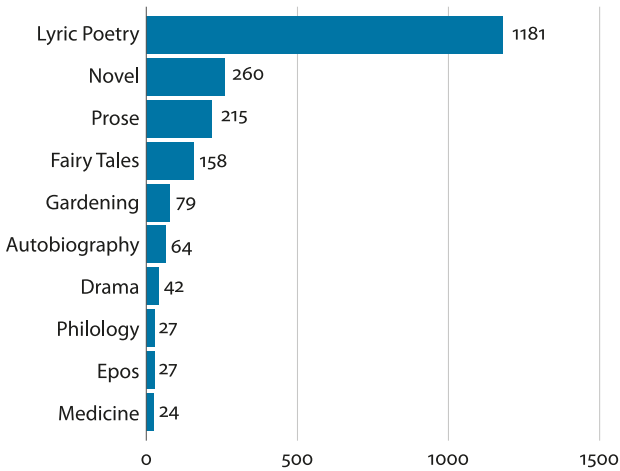
This process ended around 1760, at which point *-lein* was only sparsely documented. The most important domains were then – i.e., between 1700–1760 – edifying literature, biological literature, and cookbooks. With the propagation of *-gen* and *-chen*, a shift amongst the genres becomes obvious; particularly from 1800 onward (Period 5), *-lein* use was restricted to fiction (above all lyric poetry, novels, and novellas), while there were nearly no occurrences in nonfiction and scientific literature, which were the domains in which *-lein* previously prevailed. This finding is in line with the renewal of *-lein* shortly after 1800 (Figure 4).

This change becomes especially clear in Figure 8, which contrasts the 10 most popular subgenres for *-lein* use before and after the replacement. As the dominance of *-lein* finally came to an end around 1760, the final part of Period 3 (1762)

is taken to be the division border of Figure 8; the figure therefore compares the 50 years preceding that border (Phase A) to the 50 years that follow it (Phase B).



a. Tokens 1712–1762



b. Tokens 1763–1813

Figure 8. Top 10 subgenres of *-lein* representations during an interval of ± 50 years around the end of Period 3 (1762)

The figure demonstrates a dramatic increase of *-lein* in lyric poetry. In particular, this increase benefits from the jump around 1800 (see Figure 7). This jump is due to Arnim and Brentano's anthology "Des Knaben Wunderhorn" ("The Youth's

Magic Horn') from 1806 to 1808, which is a collection of romantic folk songs, fairy tales, and ballads.

Figure 7 further indicates a clear increase for types in the time up to 1770. This is, to a large extent, due to writings from the era of Storm and Stress (*Sturm und Drang*) and the Göttinger Hainbund; Pfennig (1904) and Öhmann (1972) also pointed to this. Authors from this era typically used *-lein* as an indicator of traditional topics, mostly within lyric poetry and novels. At that time, the *-lein* suffix was considered to be rather old-fashioned. This is interesting when compared to Schebben-Schmidt (1990:316), who found that there was no *-lein* increase in her data. However, the DTA data indicates that this renewal did indeed occur and, as seen above, correlates with the type analysis of Wellmann (1975:133). However, most prominently, *-lein* then became the typical diminutive suffix for the romantic literature of the 19th century.

4.3.3 *-chen*

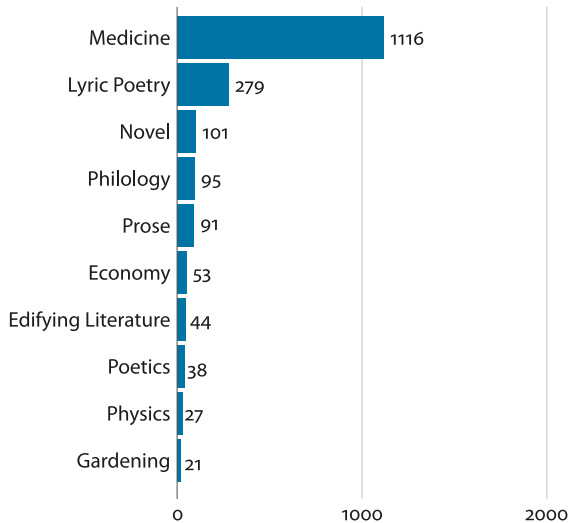
The first *-chen* suffixes arose around 1650 in fictional literature (Period 1), specifically in the subgenres of lyric poetry and prose. Fiction continued to be the leading genre for the propagation of this suffix one hundred years later, around 1750. This is not completely in line with the findings of Gürtler (1909a:17), who concluded that *-chen* was mainly reserved for fictional literature of the most popular styles (e.g., drama) before 1750.

Figure 9 indicates that in the phase before the replacement (Phase A), *-chen* was preferred in subgenres that were of secondary importance for *-lein* at the same time (Figure 8). This was especially true for medicinal literature, which was not represented in the genres that featured *-lein* during Phase A. The same applied for lyric poetry and novels during Phase A.

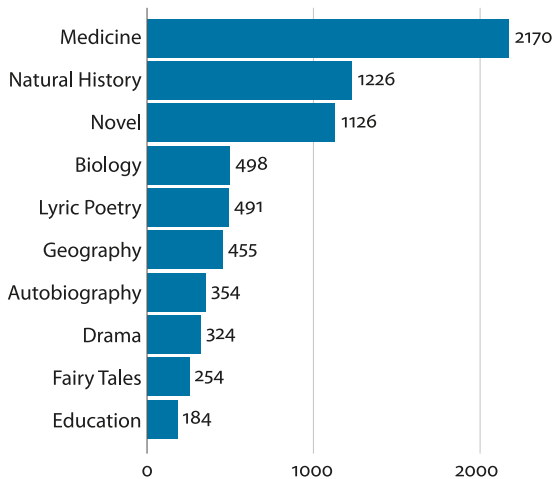
In Phase B, *-chen* can be seen to appear particularly frequently in scientific literature (above all medicine, but also natural history and biology). This contradicts Pfennig's (1904:32) assumption that diminutives were scarcest in this genre. It is, however, obvious that in the given contexts diminutives were used according to their base function (see above), which is to specify the smallness of a given subject, as can be seen in the following example:

- (3) In neuern Zeiten ist dieses Bläschen so oft, und in befruchteten Eychen bis in das dritte Monat der Schwangerschaft so gleichmäßig wahrgenommen worden [...]
 'In more recent times this small bubble ['bubble.DIM'] has been noticed so often and so consistently until the third month of pregnancy in fertilized eggs ['egg.PL.DIM'; i.e., egg cells] [...]' (Blumenbach 1789:348)

In fictional literature, *chen* is very frequent in Phase B, at least compared to *-lein*; the number of *-chen* suffixes in fiction (Figure 9b) exceeds the sum of all *-lein*



a. Tokens 1712–1762



b. Tokens 1763–1813

Figure 9. Top 10 subgenres of *-chen* representations during an interval of ± 50 years around the end of Period 3 (1762)

occurrences in Phase B (Figure 8b). However, as seen in Figure 8, this fact is mainly due to an extraordinarily high amount of *-lein* suffixes in lyric poetry.

While in Phase B, *-chen* is basically represented in all genres (see Figure 7, Figure 9), it is also clear that this suffix was comparatively infrequent in lyric poetry, which was the most significant domain for *-lein*. Considering that in Phase A *-chen* was already preferred in secondary *-lein* domains, the two suffixes were to some extent in complementary distribution before and after the replacement.

4.3.4 *-gen*

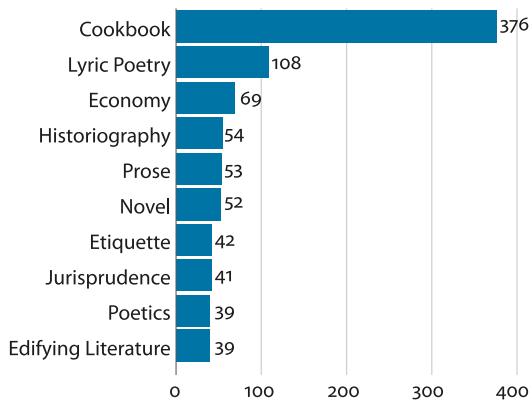
The *-gen* suffix appeared in fiction around 1650. Soon afterwards, shortly before 1700, the preference for *-gen* usage had intensified not only within fictional literature, but also in the realm of scientific literature. Interestingly, in fictional literature at this time, *-gen* was the leading suffix, but its popularity dropped shortly after. Apart from that, there were only marginal differences amongst the genres.

On the whole, the amount of *-gen* usage is too small for a more focused discussion. An exception is the clear preference for *-gen* in cookbooks during Phase A of Figure 10. These hits were all from a cookbook by Gottlieb Sigmund Corvinus from 1715. Corvinus, who was born and died in Leipzig, originated from the Upper Saxon area, which is why it is likely that these *-gen* variants, as per Gürtler (1909a), were reflexes from dialect. The next section will focus on such regional issues.

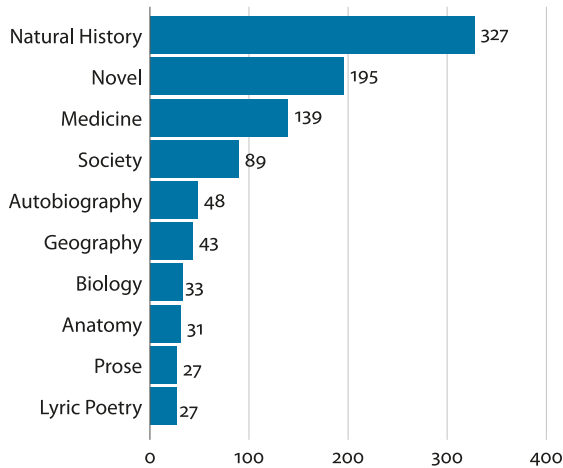
4.4 Geographic distribution

Examining the geographic distribution of diminutives is essential for describing the phenomenon. However, since the DTA was not explicitly developed for regional analyses, an examination of this factor is difficult to implement. The methodological issue is that there is no geographic information on DTA data available, which is why it must be reconstructed from scratch. Therefore, for this study, the authors from the DTA corpus first were identified; second, information on the birthplaces of these people was gathered. It is understood that a person's birthplace does not necessarily coincide with the place of their language acquisition. On the whole, however, after also examining other possible factors – e.g., authors' places of activity or the location of their publishers – birthplace has proven to be the best stand-in for the regional imprint of the DTA writers.

The necessary information was gathered from the servers of the German National Library (DNB). The DNB provides biographic information on historical figures that can be collected via a web interface. In a few cases, this was supplemented by information from other sources, such as Wikipedia. In total, 1,032 DTA authors between 1600 and 1900 were identified. The necessary information regard-



a. Tokens 1712–1762



b. Tokens 1763–1813

Figure 10. Top 10 subgenres of *-gen* representations during an interval of ± 50 years around the end of Period 3 (1762)

ing their places of birth was available for 854 (82.75%) of the authors. Amongst authors who employed diminutives, information about birthplace was found for 506 out of 525 persons (96.38%).

With this information, the DTA data was ready to be mapped. It appears that this is the first time such maps have been created. In contrast to previous studies that mapped data from corpora, the chosen approach in this study is innovative in that it indicates both writers who used diminutives and writers who did not.

In contrast to previous studies, which typically used time steps of 50 years from a century's beginning onward, the present study relies on the periodization provided by Table 4 for analysis. This leads to a further re-sampling of the DTA data into six subsequent periods, summarized in Table 5. Adding onto the information in Table 4, Table 5 reports information on the number of identified locations (together with the number of all locations), the number/percentage of locations with diminutive hits, and the number of diminutive suffixes within each particular period.

Table 5. Identified DTA Locations (i.e., birthplaces of 1,032 authors)

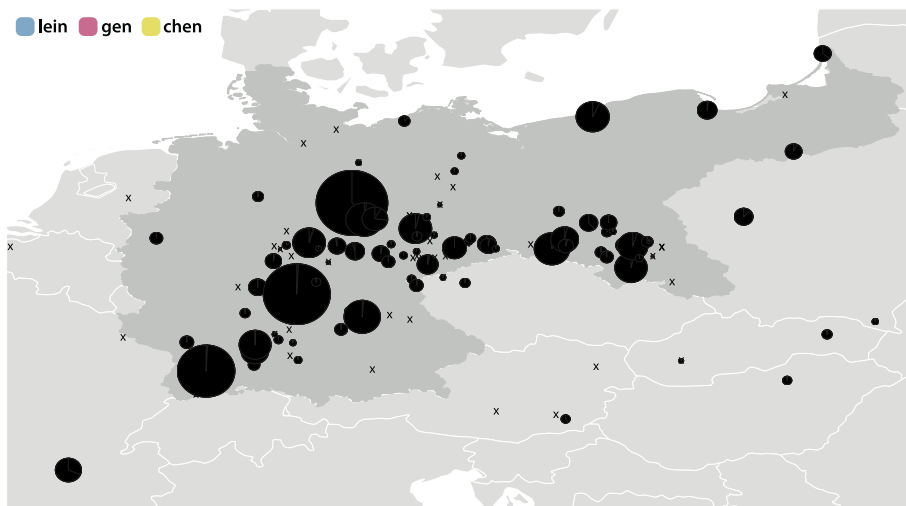
Period	Identified locations	Locations (with diminutive hits)	Diminutives (tokens)		
			<i>-lein</i>	<i>-gen</i>	<i>-chen</i>
1 1600–1660	146 (/246)	37% (90)	71% (134)	5% (10)	24% (45)
2 1661–1725	122 (/165)	39% (64)	51% (119)	20% (46)	30% (70)
3 1726–1762	62 (/77)	64% (49)	34% (66)	32% (62)	34% (65)
4 1763–1812	156 (/165)	59% (98)	32% (133)	19% (77)	49% (203)
5 1813–1843	100 (/108)	53% (57)	40% (83)	6% (12)	54% (113)
6 1844–1900	250 (/271)	38% (103)	38% (140)	5% (17)	57% (208)

Table 5 demonstrates that the regional coverage of the DTA data is rather unbalanced. Nevertheless, the individual samples are suited to depicting the dynamics amongst the periods of time. A second result made clear by Table 5 is that the geographical spread of diminutives changed over time. The locations with diminutive hits increased from Period 1 until Period 3 (64% coverage), then successively decreased afterwards. The highest jump was from Period 2 to Period 3, which was the period of *-chen* propagation.

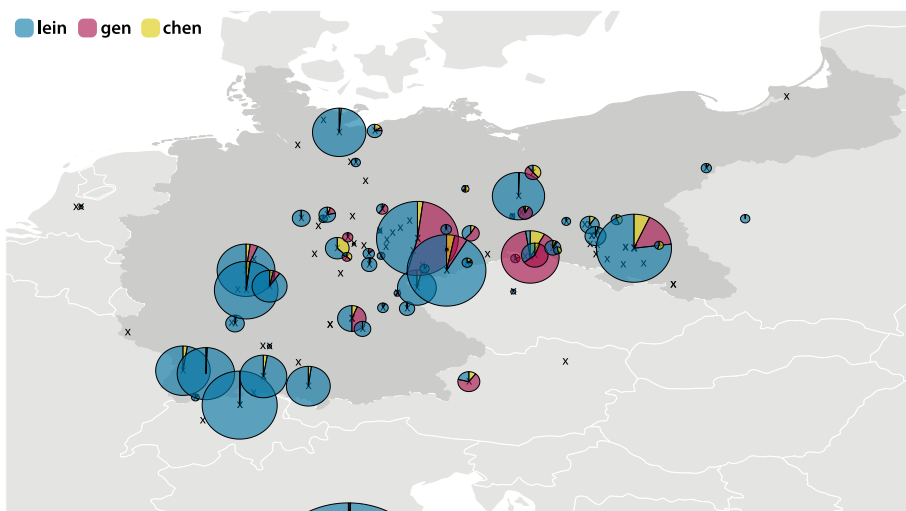
The maps in Figure 11 show the distribution of *-lein*, *-chen*, and *-gen* within these periods. Crosses (×) indicate the origin of all DTA writers within each particular period. By examining these maps, it becomes possible to place writers who use diminutives in relation to those who do not. The proportion of use of each suffix is indicated by simple pie charts for each location. The radius of each pie chart is proportional to the number of types at individual locations. For reference, the base map is provided as a combination of the German national grid around 1900 with a modern map of Europe. For a better orientation to areal issues, Figure 12 provides an overview of the German language area (see Appendix).

In more detail, the maps can be characterized as follows:

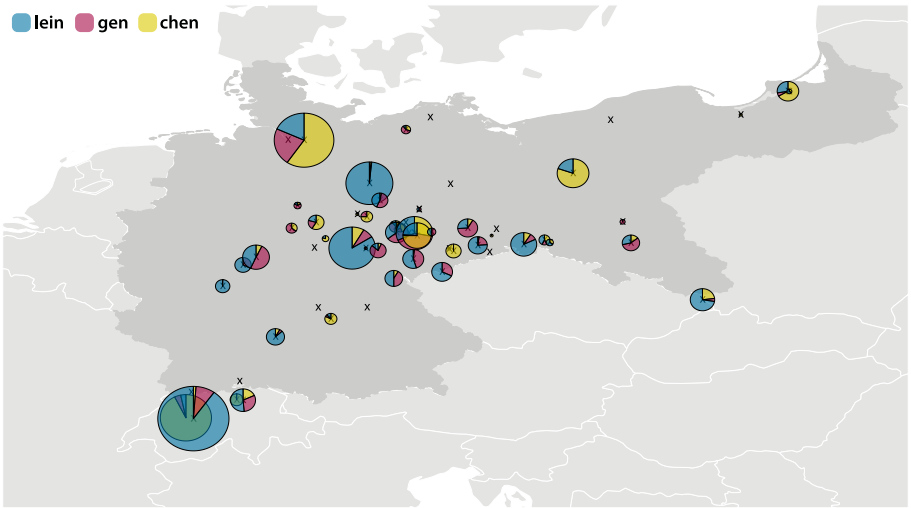
Map A (Period 1: 1600–1660) represents the situation before the initial propagation of *-gen* and *-chen* within the ENHG period. Obviously, not every writer used diminutives and as such, there were many locations where only very few



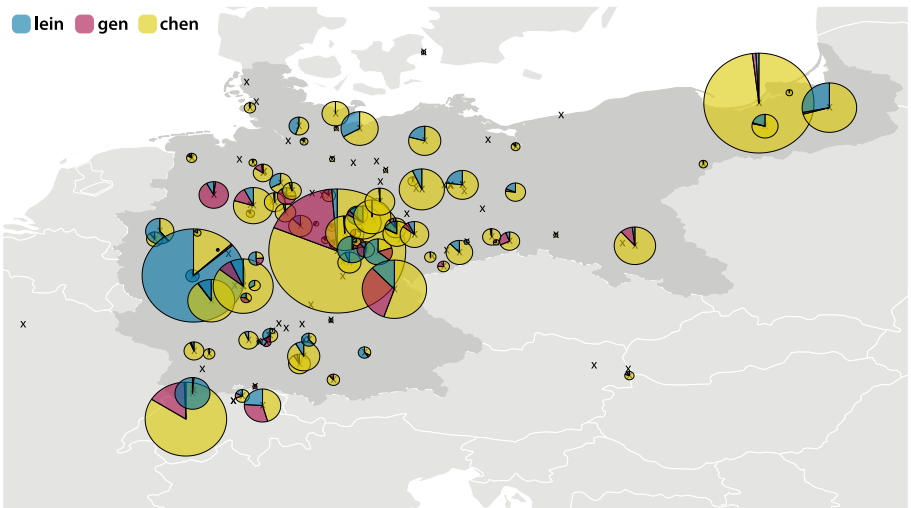
Map A. Period 1 (1600–1660)



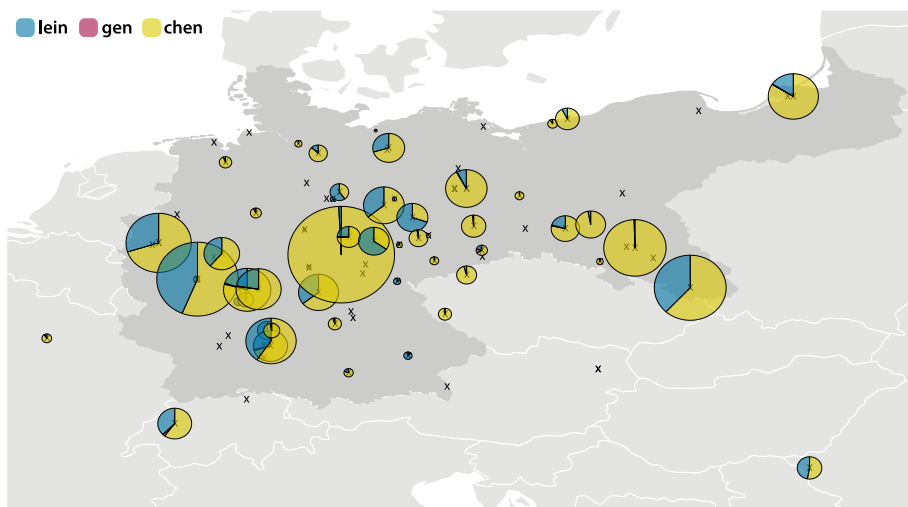
Map B. Period 2 (1661–1725)



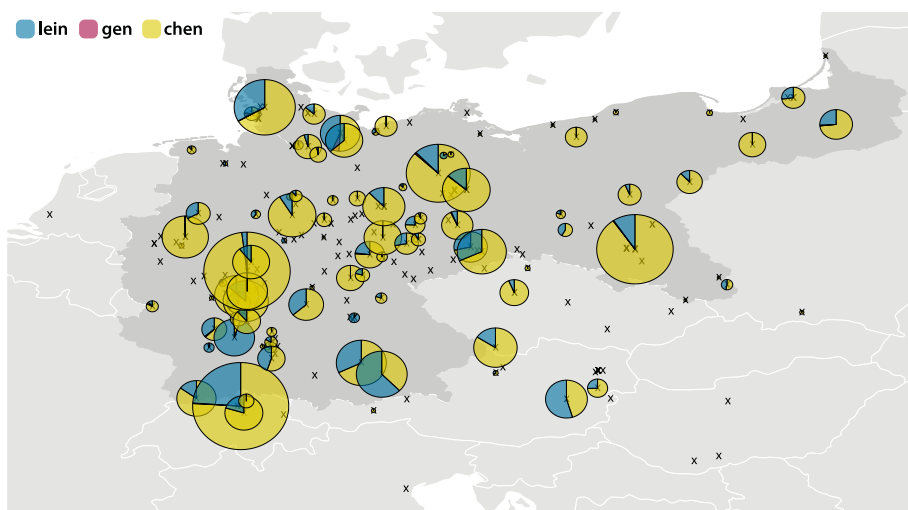
Map C. Period 3 (1726–1762)



Map D. Period 4 (1763–1812)



Map E. Period 5 (1813–1843)



Map F. Period 6 (1844–1900)

Figure 11. Geographical distribution of suffixes according to the empirically defined periodization^a

^a Crosses indicate the birthplaces of writers under investigation; pie charts show the distribution of suffixes per writer; the radius of pie charts is proportional to the number of types.

diminutives occurred. The core area was in the Central German region (Thuringia, Saxony, Silesia, Hesse) together with the western part of the Upper German region. The most frequent suffix was *-lein* (71%). Some occurrences of *-chen* were present in Thuringia and Silesia (24%). Even fewer occurrences of *-gen* could be found in Thuringia and Saxony (5%). This finding is striking, as there has been a clear expectation concerning a geographical distribution of diminutive suffixes by the end of the ENHG period (Schebben-Schmidt 1990:314). As a whole, this seems not to be the case. On the other hand, Wegera & Solms (2002:165) considered the fact that there might already have been an effect of registers (“Lekte”) in this time; this, however, cannot be proven at present. Nevertheless, this would explain why there is no clear geographic pattern to be seen.

Map B (Period 2: 1660–1725) shows an increase of *-gen* (20%). This was the actual initiation phase of the language change, as Table 4 likewise indicates. This suffix arose north of Saxony and Silesia, which confirms that the abovementioned *-gen* variants in Corvinus’ cookbook were likely to have been regional ones. But, on the whole, *-lein* was still dominant (51%), while *-chen* (30%) was represented primarily in the East Central German area but also as a secondary diminutive form for most authors. In some cases, these forms were phonologically motivated.

Map C (Period 3: 1726–1762) represents the steep increase in the S-curve of *-chen* propagation. It is striking that the *-chen* propagation is not particularly reflected on the map; all suffixes show a proportion of 32–34%. Moreover, it appears that the propagation was not geographically influenced, but rather affected by individual authors. As Figure 7 and Figure 9 show, these authors fundamentally worked adjacent to the bodies of scientific literature (medicine) and fiction (lyric poetry). Regarding *-gen*, which found its zenith of existence within this period, this suffix was represented in virtually every region, although it was mostly a secondary variant. Also, *-chen* spread into new regions, but not as intensively as *-gen* did. This, as a whole, led to the particular situation of Period 3, during which the highest geographical coverage was found: diminutives were present in 64% of the DTA locations at this time.

Within Map D (Period 4: 1763–1812), a total change in the situation can be seen. This was the termination phase of the period of language change, in which *-chen* was finally found in 49% of all regions alongside a considerable amount of *-gen* (19%). The latter were, however, outdated at this time (Pfennig 1904:3). There is no clear geographical pattern, save for the fact that there was no *-gen* variant in the northern area or in Brandenburg. At the same time, the regional coverage had clearly been reduced.

Map E (Period 5: 1813–1843) depicts the point at which the language change was completed. The *-gen* suffixes were almost completely lost (6%). At the same time, *-lein* and *-chen* were more frequent for the Central German writers, while

they were less frequent for the writers from Northern Germany or to those from Bavaria. Consequently, the *-lein* renewal (40%) was especially reflected in the Central German area.

Map F (Period 6: 1844–1900) shows the predominance of *-chen* (57%) along with its spread to further regions. There was no longer any geographical impact in this period.

It is convenient to compare the geographical distributions of Periods 4–6 with the map provided by Elspaß (2005: 343) regarding the diminutive suffixes of written “everyday language” of non-specialists during the 19th century. Doing so reveals that only in Period 4 (1763–1812) can there be seen similarities between the writings of these DTA authors and the writings of non-specialists, both of which featured considerable use of *-gen* variants. The highly standardized written language represented by the 19th century DTA authors (Period 5 and Period 6), however, did not follow this pattern.

All in all, the maps clarify that geography is an important factor not only in the initiation phases of the propagation of *-chen/gen* (Period 2), but also to some extent in later periods, although geography seems to have lost its impact by the end of the 19th century. At this time, diminutive use in the German written language had no relationship to the patterns of use in local dialects described by, e.g., Wrede (1908) or to the regionally-bound writing practice of non-specialists (Elspaß 2005).

Furthermore, the chosen approach makes it possible to test whether regional preferences have affected diminutive choice, as has been suggested already by Wrede (1908) for the spoken language. From this assumption, one would expect to find higher usage rates for diminutives in the central and southern parts of the German language area. Indeed, Figure 11 reveals some preferences for diminutive usage in the Central German area during Period 4 (Map D) and Period 5 (Map E). In addition, in Period 6 (Map F), at least in the Alemannic region, most authors used diminutives, which contrasts with the patterns in the central and northern regions. There is, however, no clear indication of a more general trend.

5. Conclusion

This article has analyzed the replacement of the German diminutive *-lein* by the suffixes *-gen* and *-chen* during the NHG period as provided by data from DTA. The chosen approach is specifically suited to represent the development of all three suffixes under discussion. There are clear patterns of language change that are different for (i) suffixes, (ii) genres, and (iii) regions. The results are consistent with

the literature, but go beyond the results of other studies with regards to some important points.

The process of replacement was extremely fast. Period 3, which was the actual phase of transition, only lasted 37 years, which equates to roughly one generation. The process can best be described by the S-shaped curve of *-chen* propagation that was found for both types and tokens. This is consistent with expectations from other studies on language change (Labov 1994: 65–66, Blythe & Croft 2012). At the same time, *-chen* and *-lein* in particular showed a significant interrelation expressed by a delayed negative correlation, according to which *-chen* led *-lein*. What is important in this is not so much the exact time specification of this delayed correlation (five years), but the finding of the delay as such. It indicates that there was a strong interrelation between these two suffixes. This was true neither for the relationship between *-lein* and *-gen* nor for the relationship between *-chen* and *-gen*. Moreover, *-gen* seemed to be rather discrete, which confirms Wegera's (2000) assumption. This is underlined by the fact that it was the *-gen* suffix that obviously initiated the replacement.

The whole process was accompanied by a reduction of diminutive use. This had already been hypothesized by Polzin (1901:107) but never proven empirically. Pure diminutive use is expressed by tokens. There were two phases regarding their development. Initially, there was a significant reduction of tokens up to the middle of the 18th century. From then onward, the number of tokens became stable over time. When Fuhrhop & Werner (2016:135) assumed that diminutives nowadays are less frequent than they had been in earlier times, this would fit, given the historic reduction in the present data. Furthermore, it is likely that post-Romanticism, diminutives were even more reduced in number.

This does not apply, however, to the variability of diminutive use, which is expressed by the number of types. In contrast to the number of tokens, there was a successive increase of types over time, which indicates that diminutives became more productive over time. The present study has shown that this was due to transfers between genres. During the first 150 years of the time span, diminutives – at that time, almost exclusively *-lein* diminutives – were reserved for theological writings, such as sermons or edifying literature. Considering the continuous decrease of suffixes in the theological literature (i.e. nonfiction in Figure 7), it seems that the rhetorical use of diminutives was considered outdated from the middle of the 17th century onward. This is contrary to botanic literature, in which diminutives served to specify the subjects under discussion. In such scientific contexts, however, the rise of the “modern” *-chen* suffix took place during the second 150 years of the time span. Moreover, from then onward, diminutives – i.e., all three diminutives, *-lein*, *-chen*, and *-gen* – were frequently represented in all genres.

The geographic approach taken here has been based on the origin of writers. For the first time, the historical stages of *-lein/-chen* replacement are documented using maps. Interestingly, there was no clear regional distribution of suffixes for the ENHG period (Period 1), which might be due to register variation. At the same time, this was the period with the lowest geographical coverage of diminutives. A regional pattern emerged only in Period 2 (1661–1725) with the increase of *-gen* suffixes. From Period 4 (1763–1812) onward, such a geographical pattern no longer existed, while the regional coverage, which found its peak during the time of the actual *-chen* propagation (Period 3), successively decreased. This confirmed Schebben-Schmidt's (1990) finding that geographical dependencies – in contrast to spoken language – were almost lost after 1750. Contrary to Schebben-Schmidt (1990) and Elspaß (2015), however, this analysis found that even in spite of the strong reduction in *-lein* use, this suffix was nevertheless present in all regions. Still, as the analysis of genres reveals, it was no longer represented in prose texts, but in lyric poetry, novellas, etc., which is why it has not been captured in corpora that do not cover such genres. Regarding the assumption that diminutives were more frequent in the southern part of the language area, no clear indication could be found that this was the case.

Furthermore, there was not only one, but multiple different – though smaller – processes of diminutive increase, some of which have been overlooked by other studies. Among these processes was an increase of *-chen* around 1660. These smaller processes contributed to the periodization that the change point analysis in this study empirically compiled. Examining these processes also helps us understand the actual process of language change in more detail. Finding that, e.g., during the phase of *-lein* reduction there was an increase of *-lein* types, it becomes obvious that the process of *-lein* reduction was not linear but rather a fairly complex development in which writers were looking for new domains of application. In a sense, this was successful, considering that *-lein* finally became a variant that is especially suited for, e.g., the expression of a more folk-like sound.

Considering the actual motivation of *-chen* propagation, the very clear substitution in scientific literature has to be emphasized once more. Interestingly, this was the domain in which diminutives were most restricted to their pragmatic base function, which is to indicate smallness. While *-lein* was often found in botanic literature or in natural sciences before 1760, it was no longer used in these genres after 1760. In contrast, prior to 1700, *-chen* occurred in scientific writing (medicine), where it was also most frequent, shortly after replacing *-lein*. Against this background, it seems that *-lein* had been weakened semantically by the middle of the 18th century (Periods 1–3). This is underlined by the clear decrease in rates of use of this suffix in nonfiction; writers obviously did not see a clear semantic benefit in using the available *-lein* suffix. This was possibly conditioned by the suc-

cessive dominance of the emotive meaning over the denotative meaning in many contexts (Dressler & Barbaresi 2001; Biały 2013). There was, however, still a need for the unambiguous indication of smallness. This is why the regional variants *-chen* and *-gen* came into play. As new written suffixes, they were more suited to express the prototypical meaning without or at least with fewer emotive connotations, which is why they were semantically more efficient than *-lein*. Finally, from Period 4 onward, *-chen* adopted the former pragmatic functions of *-lein*. From this perspective, it must be assumed that the actual replacement of *-lein* by *-chen* was due to the temporary strengthening of the semantic effort of the diminutive.

Acknowledgments

I am grateful to two anonymous reviewers and the editors of JHL for their valuable suggestions. I also thank Jürg Fleischer, Jeffrey Pheiff, Alexander Werth, and the audience of the 8th GGSG conference (Leipzig 2016) for helpful comments on earlier versions of this paper. Special thanks go to Damaris Nübling for constructive discussions on the topic. The usual disclaimers apply.

Corpora

Deutsches Textarchiv: Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache.

Herausgegeben von der Berlin-Brandenburgischen Akademie der Wissenschaften, Berlin 2018. 2007-present. Available at <http://www.deutschestextarchiv.de>.

Durrell, Martin, Silke Scheible, Anita Auer, David Denison & Núria Yañez Bouza.

2006-present. German Manchester Corpus. Available at <http://www.llc.manchester.ac.uk/research/projects/germanc>.

German National Library. Available at <http://www.dnb.de/>.

Wikipedia. Available at <https://www.wikipedia.de/>.

References

- Baayen, Harald R. 2005. Morphological Productivity. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski, eds., 243–255.
- Barbour, Stephen & Patrick Stevenson. 1990. *Variation in German: A Critical Approach to German Sociolinguistics*. Cambridge: Cambridge University Press.
- Biały, Paulina. 2013. On the Priority of Connotative over Denotative Meanings in Polish Diminutives. *Studies in Polish Linguistics* 8:1.1–13.
- Blumenbach, Johann Friedrich. 1789. *Anfangsgründe der Physiologie*. Wien: Wappler.
- Blythe, Richard A. & William Croft. 2012. S-curves and the Mechanism of Propagation in Language Change. *Language* 88:2.269–304. <https://doi.org/10.1353/lan.2012.0027>

- Davies, Winifred V. & Nils Langer. 2006. *The Making of Bad Language: Lay Linguistic Stigmatisations in German, Past and Present*. (= *VarioLingua*, 28.) Frankfurt/Main: Peter Lang.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin/New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110877052>
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 2001. Morphopragmatics of Diminutives and Augmentatives: On the Priority of Pragmatics over Semantics. *Perspectives on Semantics, Pragmatics, and Discourse* ed. by István Kenesei & Robert M. Harnish, 43–58. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.90.07dre>
- Eckert, Gabriele. 1986. *Sprachtypus und Geschichte: Untersuchungen zum typologischen Wandel des Französischen*. Tübingen: Narr.
- Edelhoff, Maïke. 2016. Von Blätterchen und Bäumchen: Die Entwicklung der Plural-Diminutive und Diminutiv Plurale im Deutschen und Luxemburgischen. *Wortbildung im Deutschen: Aktuelle Perspektiven* ed. by Elke Hentschel, 191–213. Tübingen: Narr.
- Edwards, A. W. F. & Luigi Luca Cavalli-Sforza. 1965. A Method for Cluster Analysis. *Biometrics* 21:2.362–375. <https://doi.org/10.2307/2528096>
- Ehlers, Klaas-Hinrich. 2011. „Dürfen wir essing?“ Beobachtungen zur Übernahme des mecklenburgischen *ing*-Diminutivs in das regionale Hochdeutsch. *Niederdeutsches Jahrbuch* 134.79–92.
- Elmentaler, Michael. 2013. Von *Zettelchen*, *Äpfelchen* und *Schäfchen*: Funktionale und areale Aspekte des Diminutivgebrauchs im Niederdeutschen und in der norddeutschen Alltagssprache. *Twenty-Nine Smiles for Alastair* ed. by Jarich Hoekstra, 21–37. Kiel: Christian-Albrechts-Universität.
- Elspaß, Stephan. 2005. *Sprachgeschichte von unten: Untersuchungen zum geschriebenen Alltagsdeutsch im 19. Jahrhundert*. Tübingen: Niemeyer. <https://doi.org/10.1515/9783110910568>
- Elspaß, Stephan. 2015. Grammatischer Wandel im (Mittel-) Neuhochdeutschen – von oben und von unten. Perspektiven einer Historischen Soziolinguistik des Deutschen. *Zeitschrift für Germanistische Linguistik* 43:3.387–420. <https://doi.org/10.1515/zgl-2015-0022>
- Erben, Johannes. 2006. *Einführung in die deutsche Wortbildungslehre. 5. durchgesehene und ergänzte Auflage*. Berlin: Schmidt.
- Ettinger, Stefan. 1980. *Form und Funktion in der Wortbildung. Die Diminutiv- und Augmentativmodifikationen im Lateinischen, Deutschen und Romanischen*. Tübingen: Niemeyer.
- Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache. 4. Auflage*. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783110256659>
- Fuhrhop, Nanna & Martina Werner. 2016. Die Zukunft der Derivation oder: Derivation 2.0. *Linguistik Online* 77:3.129–150.
- Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian Word Formation: A Variable-Corpus Approach. *Linguistics* 44:1.57–89. <https://doi.org/10.1515/LING.2006.003>
- Gottsched, Johann Christoph. 1762. *Vollständigere und Neuerläuterte Deutsche Sprachkunst. Nach den Mustern der besten Schriftsteller des vorigen und itzigen Jahrhunderts abgefasst und bey dieser fünften Auflage merklich verbessert*. Leipzig: Breitkopf.
- Gürtler, Hans. 1909a. *Das Deminutivsuffix -chen im Frühneuhochdeutschen*. Düsseldorf: Frembgen.

- Gürtler, Hans. 1909b. Materialien zur Geschichte der Diminutiva auf -chen im Frühneuhochdeutschen. *Zeitschrift für Deutsche Wortforschung* XI.181–210.
- Haaf, Susanne & Christian Thomas. 2016. Die historischen Korpora des Deutschen Textarchivs als Grundlage für sprachgeschichtliche Forschungen. *Sprachgeschichte des Deutschen: Positionierungen in Forschung, Studium, Schule* ed. by Holger Runow, Volker Harm & Levke Schiwiek, 217–234. Stuttgart: Hirzel.
- Jurafsky, Daniel. 1996. Universal Tendencies in the Semantics of the Diminutive. *Language* 72:3.533–578. <https://doi.org/10.2307/416278>
- Jurish, Brian. 2013. Canonicalizing the Deutsches Textarchiv. *Perspektiven einer corpusbasierten historischen Linguistik und Philologie* ed. by Ingelore Hafemann, 235–244. Berlin: BBAW.
- Killick, Rebecca & Idris A. Eckley. 2014. Changepoint: An R Package for Changepoint Analysis. *Journal of Statistical Software* 58:3.1–19. <https://doi.org/10.18637/jss.v058.i03>
- Köhler, Reinhard, Gabriel Altmann & Rajmund G. Piotrowski, eds. *Quantitative Linguistics: An International Handbook*, 243–255. Berlin/New York: De Gruyter.
- Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt & Yongcheol Shin. 1992. Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root. *Journal of Econometrics* 54:1–3.159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Labov, William. 1994. *Principles of Linguistic Change, Vol. 1: Internal Factors*. Oxford: Blackwell.
- Lameli, Alfred. 2018. Zur Diminuierung männlicher und weiblicher Verwandtschaftsbezeichnungen in der jüngeren Sprachgeschichte des Deutschen. *Sprachwandel im Deutschen* ed. by Luise Czajkowski, Sabrina Ulbrich & Christina Waldvogel. Berlin/Boston: De Gruyter. <https://doi.org/10.1515/9783110526585-001>
- Lenders, Winfried & Klaus-Peter Wegera. 1982. *Maschinelle Auswertung sprachhistorischer Quellen. Ein Bericht zur computerunterstützten Analyse der Flexionsmorphologie des Frühneuhochdeutschen*. Tübingen: Niemeyer.
- Müller, Peter O. 1993. *Substantiv-Derivation in den Schriften Albrecht Dürers. Ein Beitrag zur Methodik historisch-synchroner Wortbildungsanalysen*. Berlin/New York: De Gruyter. <https://doi.org/10.1515/9783110859072>
- Nieuwenhuis, Paul. 1985. Diminutives. University of Edinburgh PhD dissertation.
- Öhmann, Emil. 1972. Suffixstudien VIII: Die deutschen Diminutivsuffixe -lein und -chen. *Neuphilologische Mitteilungen* 73.555–567.
- Ott, Dennis. 2011. Diminutive-Formation in German: Spelling out the Classifier Analysis. *Journal of Comparative Germanic Linguistics* 14:1.1–46. <https://doi.org/10.1007/s10828-010-9040-x>
- Pfennig, Heinrich. 1904. *Das Deminutivum bei Schiller und seinen Zeitgenossen*. Straßburg: Trübner.
- Plank, Frans. 1981. *Morphologische (Ir-)Regularitäten: Aspekte der Wortstrukturtheorie*. Tübingen: Narr.
- Polzin, Albert. 1901. *Studien zur Geschichte des Deminutivums im Deutschen*. Straßburg: Trübner. <https://doi.org/10.1515/9783111348148>
- Schebben-Schmidt, Marietheres. 1990. Studien zur Diminution in der deutschen Schriftsprache des 18. Jahrhunderts. *Deutsche Sprachgeschichte: Grundlagen, Methoden, Perspektiven* ed. by Werner Besch, 313–321. Frankfurt/Main: Peter Lang.
- Schirmunski, Viktor. 1958. Verstärkte Wortformen in den deutschen Mundarten. *Zeitschrift für Mundartforschung* 26.225–238.

- Schmuck, Mirjam. 2009. Personennamen als Quelle der Grammatikalisierung: Der *ing* Diminutiv in Mecklenburg-Vorpommern. *Beiträge zur Namenforschung* 44:1.35–65.
- Schottel, Justus Georg. 1641. *Teutsche Sprachkunst*. Braunschweig: Gruber.
- Seebold, Elmar. 1983. Diminutivformen in den deutschen Dialekten. *Dialektologie: Ein Handbuch zur deutschen und allgemeinen Dialektforschung* ed. by Werner Besch, Ulrich Knoop, Wolfgang Putschke & Herbert E. Wiegand, 1250–1255. Berlin/New York: De Gruyter. <https://doi.org/10.1515/9783110203332.1250>
- Stricker, Stefanie. 2000. *Substantivbildung durch Suffixableitung um 1800: Untersucht an Personenbezeichnungen in der Sprache Goethes*. Heidelberg: Winter.
- Tiefenbach, Heinrich. 1987. -chen und -lein: Überlegungen zu Problemen des sprachgeographischen Befundes und seiner sprachhistorischen Deutung. *Zeitschrift für Dialektologie und Linguistik* 54.2–27.
- Voeste, Anja. 1999. *Varianz und Vertikalisierung: Zur Normierung der Adjektivdeklinaton in der ersten Hälfte des 18. Jahrhunderts*. Amsterdam/Atlanta: Rodopi.
- Wegera, Klaus-Peter. 1982. Möglichkeiten und Grenzen der philologischen Auswertung einer elektronisch gespeicherten Datei zum Frühneuhochdeutschen. Aufgezeigt am Beispiel der Diminutivsuffixe. *Sprachen und Computer: Festschrift zum 75. Geburtstag von Hans Eggers* ed. by Hans Fix, Anneli Rothkegel & Erwin Stegentritt, 207–222. Dudweiler: AQ-Verlag.
- Wegera, Klaus-Peter. 2000. „Gen, oder wie Herr Gottsched will, chen.“ Zur Geschichte eines Diminutivsuffixes. *Wortschatz und Orthographie in Geschichte und Gegenwart: Festschrift für Horst Haider Munske zum 65. Geburtstag* ed. by Mechthild Habermann, Peter O. Müller & Bernd Naumann, 43–58. Tübingen: Niemeyer.
- Wegera, Klaus-Peter & Hans-Joachim Solms. 2002. Wortbildung des Mittelhochdeutschen. Zur Methode und zum Stand ihrer Erforschung, dargestellt am Beispiel der Diminutive. *Historische Wortbildung des Deutschen* ed. by Mechthild Habermann, Peter O. Müller & Horst Haider Munske, 159–169. (= *Reihe Germanistische Linguistik*, 232.) Tübingen: Niemeyer. <https://doi.org/10.1515/9783110940756.159>
- Wellmann, Hans. 1975. *Deutsche Wortbildung: Typen und Tendenzen in der Gegenwartssprache. Eine Bestandsaufnahme des Instituts für deutsche Sprache Forschungsstelle Innsbruck. Zweiter Hauptteil: Das Substantiv*. Düsseldorf: Schwann.
- Wimmer, Gejza. 2005. The Type-Token Relation. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski, eds., 361–368.
- Wrede, Ferdinand. 1908. *Die Diminutiva im Deutschen*. Marburg: Elwert.

Appendix

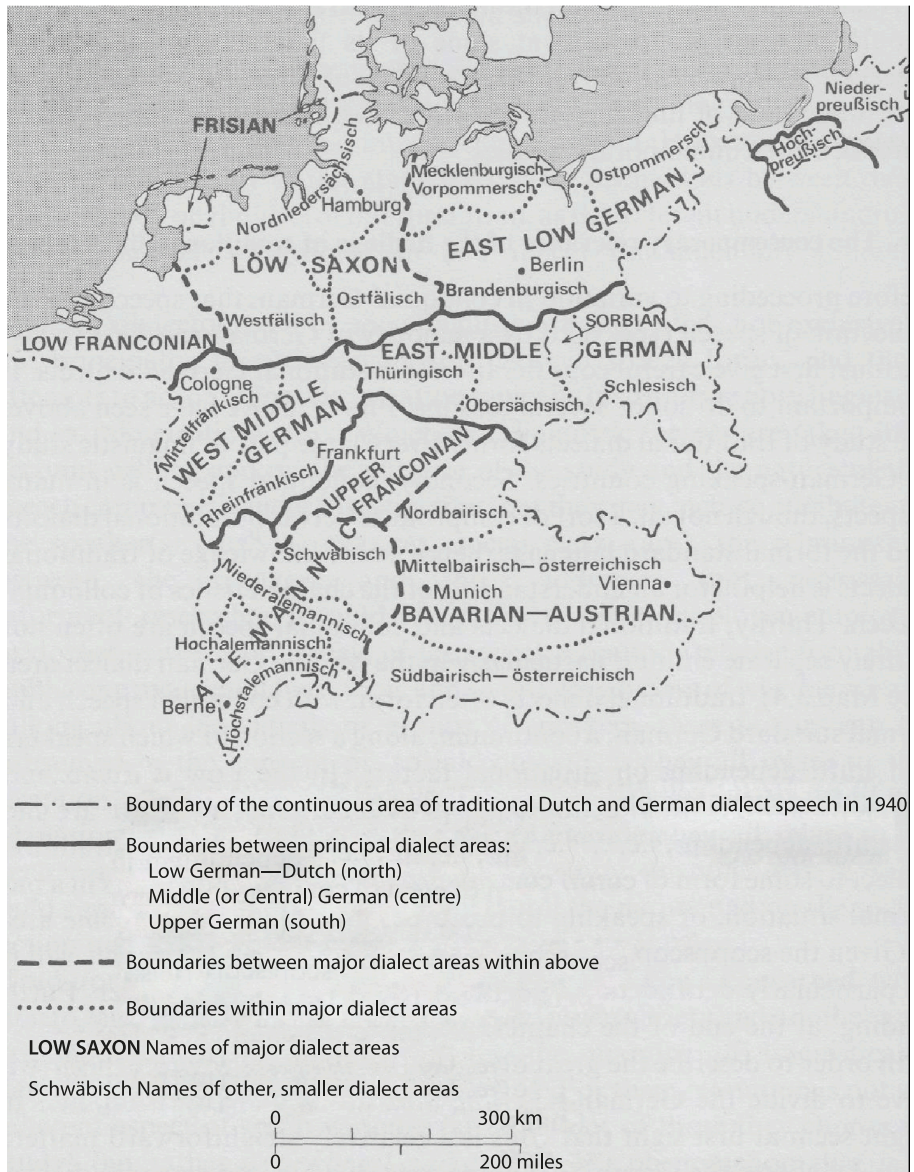


Figure 12. The German language area around 1940 (Barbour & Stevenson 1990:76)

Address for correspondence

Alfred Lameli
Deutsches Seminar – Germanistische Linguistik
Albert-Ludwigs-Universität Freiburg
Platz der Universität 3
79098 Freiburg
Germany
germanistik.uni-freiburg.de