

## **Syllable Structure Spatially Distributed: Patterns of Monosyllables in German Dialects**

Alfred Lameli

*Research Center Deutscher Sprachatlas*

This study presents a micro-typological description of German dialects, focusing on the structure of 13,492 tokens of monosyllables, across 182 locations within Germany. Based on data from the *Phonetischer Atlas der Bundesrepublik Deutschland*, systematic geographical differences in both the segmental and prosodic organization of syllables are explored. The analysis reveals a North–South contrast in the organization of syllable structure. While the North tends toward more simple CVC syllables, the South tends toward the clustering of obstruents. An analysis of sonority dispersion reveals that in southern German, final demisyllables tend to follow more closely the sonority scale. Based on Markov chain models, the study reveals geographical differences in transition probabilities between the segments within monosyllables in German dialects.\*

---

\* This research is funded by the *Deutsche Forschungsgemeinschaft (DFG; German Research Foundation)*, Grant 432304149. I am grateful to a number of people for helpful comments on previous versions of this article: Alexander Werth (Passau), Samantha Link (Marburg), Beat Siebenhaar (Leipzig), and the participants of vital seminar discussions at the universities of Leipzig, Freiburg i. Br., and Marburg. I would also like to thank Tracy A. Hall and two anonymous reviewers, especially reviewer A, for very constructive discussions. A big thank you goes to Ilana Mezhevich for the copy editing. The usual disclaimers apply.

© Society for Germanic Linguistics 2022. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

**Keywords:** German dialects, syllable structure, monosyllables, phonotactics, Markov chains, sonority dispersion

## 1. Introduction.

This article addresses regional differences between syllable structures of different German dialects. While syllable structure of Standard German has been subject of extensive research (see Menzerath 1954, Seiler 1962, Wiese 1988, Ortman 1991, Hall 1992, Kohler 1995, Wiese 2000, Best 2013, Levickij 2014 among many others), very few studies are dedicated to—usually very specific aspects of—regional phonotactics (for example, Karch 1981, Liberman 1997, Spiekermann 2000, Kraehenmann 2003, Hall 2009, Seiler & Würth 2014, Kleber 2017, Klingler, Moosmüller & Scheutz 2017, Caro Reina 2019). A comparative examination of the phonotactic particularities of individual dialect areas is sadly lacking. It is, for example, unclear whether there are regional preferences for monosyllabic or polysyllabic words, preferences for specific consonant clusters or particular sequences within the sonority scale, however defined, or violations of the Sonority Sequencing Generalization (Sievers 1881, Selkirk 1984). More generally, with regard to the major linguistic areas the question arises as to whether Low German dialects prefer different syllable structures from High German dialects, and if so, what the preferred structures are in each case. By answering such questions, a conclusive, typological differentiation of German as a variative language becomes possible.

This is the starting point for the present study, which aims at a micro-typological description of German dialects, focusing on the structure of ca. 13,500 monosyllables across 182 locations, within the Federal Republic of Germany. Monosyllables were chosen because they usually contain the most extensive consonant clusters and thus provide information not only on the maximum number of consonants in the syllables of a language, but also on the prosodic properties of the syllable structure (see, for instance, the contributions in Stolz et al. 2012). Based on a corpus of phonetically-transcribed, phonological words (*Phonetischer Atlas der Bundesrepublik Deutschland*, PAD; see Göschel 1992, 2000), systematic geographical differences in both the segmental and prosodic organization of segments within the onset and the coda in different German dialects are demonstrated. In so doing, this article sheds new light on the phonotactics of German from the perspective of language geography.

The structure of this article is as follows: Section 2 provides background information on the phonotactics of both Standard German and German dialects. Section 3 describes material and methods, while sections 4 to 6 present the results of the individual analyses. Section 4 outlines the regional distribution of monosyllables, while section 5 analyzes the syllable structure of dialects. Section 6 provides a model of monosyllables for selected regions. Section 7 includes a discussion and section 8 concludes.<sup>1</sup>

## 2. Background.

According to Clements & Keyser (1983:29), German belongs to Type IV languages, which are languages with a maximum number of core syllables (CV, V, CVC, VC, with V indicating both long and short vowels), as is typical for most Germanic languages (Oostendorp 2020). Reflecting on the ways in which segments are combined within a syllable, Maddieson (2013) classifies German, on the basis of an evaluation of empirical studies, as a language with a “complex syllable structure,” of the type (C)(C)(C)V(C)(C)(C)(C), subsequently referred to as (C<sup>3</sup>)V(C<sup>4</sup>). By “complexity” Maddieson means clustering of consonants within onset and/or coda.<sup>2</sup> For the sake of exemplification, I present two monosyllables of the type C<sup>3</sup>VC<sup>3</sup> and CVC<sup>4</sup> (see also Seiler 1962:382f.), which fit in with Maddieson’s onset/coda model. Note that V and C refer to individual segments, which is why sequences such as /pf/ are defined as CC (see section 3.2).

(1) a. CCCVCCC            [ʃtʁʊmpf] ‘stocking’

---

<sup>1</sup> In the examples throughout the paper the following abbreviations are used: ALEM=Alemannic, BAV=Bavarian, FRC=East Franconian, GERM=Germanic, LG=Low German, MFRC=Moselle Franconian, MHG=Middle High German, NHG=New High German, OHG=Old High German, POM=Pomeranian, STD=Standard German, SWAB=Swabian, TYR=Tyrolean.

<sup>2</sup> “Languages which permit freer combinations of two consonants in the position before a vowel, or which allow three or more consonants in this onset position, and/or two or more consonants in the position after the vowel, are classified as having complex syllable structure” (Maddieson 2013). For alternative considerations of syllable complexity see, among others, Auer 1993, Maddieson 2009, and Easterday 2019.

b. CVCCCC [hɛkʊbst] ‘autumn’

A more detailed view of syllable structure in Standard German is provided by Kohler (1995:175f.) with reference to monosyllables. Kohler counts the number of consonants in a cluster across morpheme boundaries, and so under his approach not only the ...VC<sup>4</sup> but also the ...VC<sup>5</sup> pattern is possible (see 2; + indicates the morphological boundary).<sup>3</sup> The special status of these patterns becomes clear because, in contrast to less complex monosyllables with, for example, CV structure, these words typically do not form minimal pairs, with the exception of a few examples, as in 2d. However, monosyllables with more than five segments are rather rare in Standard German, according to Menzerath (1954:96).

- |        |              |           |                 |
|--------|--------------|-----------|-----------------|
| (2) a. | CVCCCC+C     | [hɛkʊbst] | ‘autumn-GEN.SG’ |
|        | b. CVCCC+CC  | [ʃimpfst] | ‘scold-2SG’     |
|        | c. CCCVCC+CC | [pʁɔpfst] | ‘graft-2SG’     |
|        | d. CCVC+CC   | [glaʊbst] | ‘belief-2SG’    |
|        |              | [klaʊbst] | ‘cull-2SG’      |

From such configurations Kohler (1995) derives a linear syllable scheme with four structural positions (“Strukturpositionen”; SP) -1 — 0 — +1 — +2 with up to three Cs in the onset (=SP -1) followed by the nucleus (=SP 0), which is typically a vowel, another set of maximally three consonants (=SP +1) and, finally, another set with up to two consonants (=SP +2). While positions SP -1 to SP +1 are found in noninflected forms, SP +2 is found in morphologically complex forms, as in 2. Kohler’s model implies that, as far as the onset is concerned, a counterpart of the morphologically conditioned SP +2—a possible SP -2—does not exist in Standard German. Modifying Maddieson’s template to incorporate Kohler’s proposal leads to a possible syllable structure of the type (C<sup>3</sup>)V(C<sup>3</sup>)+(C<sup>2</sup>), where V and + indicate the boundaries of the structural positions.

The examples provided by 2 demonstrate that the preferred domain of more complex C clusters is the coda rather than the onset, which might conflict with the Coda Law, formulated by Vennemann (1988:21), according to which (additional prosodic characteristics aside) a “syllable

---

<sup>3</sup> See Bergmann 2018 for more detailed information on complex words in German.

coda is the more preferred [...] the smaller the number of speech sounds in the coda.” However, considering, for example, the highly complex syllables in Pacific languages analyzed by Easterday (2019), it is also clear that complex codas are by no means uncommon; there are language systems with much more complex C clusters in both onset and coda, which at the same time follow different prosodic organization.

Considering that Germanic languages “are [...] rather similar in the kinds of syllables they allow” (Oostendorp 2020:33), one might assume that Standard German syllable structure as described above also holds for syllables in German dialects. This may even be the reason why dialect phonotactics has not been thoroughly investigated: To date, there is no comparative study focusing on the phonotactic characteristics of regional varieties within German-speaking areas. This is unfortunate, as there is evidence of certain remarkable differences between Standard German and German dialects. For example, Alber & Meneguzzo (2016:34) point to morphologically complex onset clusters such as in 3a, suggesting that the SP -2—which, according to Kohler (1995), does not exist in Standard German—can be documented rather easily in dialects (see also Bachmann 2000:54 and Lameli & Werth 2017:75). Such morphologically complex clusters are found in family names such as *Gsell*, where the historical *gi-* or *ge-* prefix, as in 3b, is reduced.

- |        |         |             |                   |                   |     |
|--------|---------|-------------|-------------------|-------------------|-----|
| (3) a. | C+CVC+C | [kfo:ɐn]    | ‘drive-PAST.PART’ | TYR               |     |
|        | b.      | CV+.CV.C+VC | [gə.fa:.kən]      | ‘drive-PAST.PART’ | STD |

At the same time, example 3a may indicate a possible preference for monosyllabic structures in dialects, for example, due to schwa loss. However, to date there has been no study dedicated to the phonotactic relationship between monosyllables and polysyllables across German language regions. Provided that there are regions with preferences for monosyllables—as suggested by 3a—it is highly likely that these are regions with more complex C clusters. This hypothesis is partially confirmed, as discussed in section 5.2.

Another well-known difference between the standard language and dialects is described by Lass (1984:206f.; see also Hall 1992:201, 2011:147). In some Standard German words, the velar nasal is to be found in a final syllable/word position, as in 4a, while the same words in Low

German dialects come with an additional stop (=oral release) following the nasal, as in 4b.

- |            |        |         |     |
|------------|--------|---------|-----|
| (4) a. CVC | [dɪŋ]  | ‘thing’ | STD |
| b. CVCC    | [dɪŋk] | ‘thing’ | LG  |

This difference has been used as an argument for different rule orderings in Standard German versus dialects. Starting from the representation /dɪŋg/, Standard German follows the path in 5a, whereas Low German follows the path in 5b. In 5a, final obstruent devoicing fails because /g/ is already deleted, whereas in 5b g-deletion fails because /g/ has been devoiced before g-deletion could occur.

- |             |   |
|-------------|---|
| (5) a. STD: | <i>n</i> -assimilation > <i>g</i> -deletion > final obstruent devoicing |
|             | /dɪŋg/ → [dɪŋ] ‘thing’  |
| b. LG:      | <i>n</i> -assimilation > final obstruent devoicing > <i>g</i> -deletion |
|             | /dɪŋg/ → [dɪŋk] ‘thing’   |

Note that in the same dialects, in the plural form /dɪŋgə/, /g/ is deleted, just as in Standard German: [dɪŋə]. This is because in Low German dialects, *g*-deletion also occurs in the coda before stops, as in [an.fəŋkt] ‘begin-3SG.PRS’, which suggests that the rule order in 5b applies only syllable/word finally (see, for example, Map 39 and the appropriate sound recordings in Schmidt et al. 2008–2021ff. from the South of Hamburg). However, in the region in the North of Hamburg as well as in the East-Frisian dialect area [an.fəŋkt] also occurs, which suggests that even between Low German dialects there might be different rule orders. Consequently, 5b possibly holds for most, but not necessarily for all Low German dialects. How these differences are distributed geographically is another unanswered question.

The examples in 3 and 4 alone suggest that i) German dialects might follow—at least in part—a different organization of syllables from Standard German; ii) there are regional differences in the structuring of monosyllables in German dialects; iii) dialects differ in the complexity of C clusters in both onset and coda, and iv) from the perspective of generative phonology, dialects might follow different rule orderings. Above all, however, the examples make it clear that a comprehensive

description and analysis of the regional variation of German phonotactics is highly desirable. This article focuses on the regional distribution of monosyllabic structuring in German dialects. In particular, items i), ii), and iii) form the three basic hypotheses addressed in the present study.

### 3. Material and Methods.

#### 3.1. Corpus.

The PAD corpus is a collection of narrow phonetic transcriptions, originally assembled as part of the PAD project during the 1980s and 1990s (see Göschel 1992, 2000).<sup>4</sup> The PAD project documented the speech of native dialect speakers (NORMs according to Chambers & Trudgill 1998:29), who were asked to translate the so-called Wenker sentences into their local dialects. The Wenker sentences are a standard instrument used to analyze phonological and morphological particularities of both High German and Low German dialects. They were developed from 1876 onward, as the basis for the linguistic atlas of the German Empire (*Sprachatlas des Deutschen Reichs*, see Lameli 2014), which was compiled between 1889 and 1923. Originally, the atlas was based on 40 sentences (see Wenker 2013:23). Later, however, depending on the objectives of specific research, more sentences and individual words were added (Bellmann 1970:24–27, Fleischer 2017). Between 1956 and 1990, sound recordings were also made of both the original 40 sentences (ca. 450 words) and additional items.

From these data, Göschel selected recordings from 182 locations throughout the Federal Republic of Germany and constructed a subsample of 201 monosyllabic words (simplex words and compounds), which form the basis of the PAD corpus (see Göschel 1992:64; Nerbonne & Siedle 2005 provide a list of the words). Between 1980 and 1995, a team of professional phonetic transcribers at the Research Center Deutscher Sprachatlas (Marburg) generated narrow IPA transcriptions from this material. The transcriptions were recorded during a supervised process. First, the independent transcriptions of two transcribers were evaluated. Then these two transcriptions were merged into a single transcription, which was understood to be the best possible nonacoustic approximation to the sound signal (Nerbonne 2010:478). To ensure an optimal reliability

---

<sup>4</sup> The PAD corpus is not available online, and it is, at present, not part of REDE (Schmidt et al. 2008–2021), but it will be included in the future.

of the transcriptions, the transcribers' work was evaluated on a regular basis using quantitative measures (see Almeida & Braun 1986). As a result, the PAD corpus mainly focuses on segmental issues but also reflects prosodic particularities (for example, accents). In order to express the fine-grained phonetic detail in the PAD corpus, the phonetic transcriptions make extensive use of diacritics. These diacritics pose a challenge for a phonetic comparison of the dialects but are of minor relevance for the phonotactic analysis, as is demonstrated in the present study.

Even though the transcriptions were completed, the intended atlas was not. From 2003 onward, the handwritten transcriptions were recoded as machine-readable X-SAMPA code at Groningen University. Based on this code, a reversion into machine-readable IPA transcription was compiled at the Research Center in Marburg. The PAD material, as it is used in the present study, contains 30,422 tokens across a total of 182 locations, documenting ca. 3,800 phonetically-specified allophones, which are in contrast with the ca. 40 vowel and consonant phonemes of Standard German (Wiese 2000:20–23). Among these tokens are a total of 13,492 occurrences of monosyllables (44%), which are analyzed in the subsequent sections.

In recent years, the PAD corpus has been used in different studies with different aims, such as the classification of the German-speaking region (Nerbonne & Siedle 2005), the comparison of dialect regions within different languages (Nerbonne 2010), the development of quantitative methodology (Prokić et al. 2012, Prokić 2017), and the analysis of phonological complexity (Lameli & Werth 2017). In all of these instances, linguistic analysis mainly focuses on phonological or morphophonological issues.

### *3.2. Data Preparation and Representation of Segments.*

Considering Nerbonne's (2010:480) results, according to which the six words he selected from the PAD corpus alone provide 322 vowel variants, an extraordinary narrowness of the phonetic transcriptions becomes evident. However, as this narrowness is impossible to handle both in a phonetic and in a phonological analysis, the transcriptions require certain modifications.

For a more detailed description of syllable structure, segments were represented as vowels, glides, liquids, nasals, and obstruents (VGLNO-

tier), which, for example, allows one to see the differences in sonority between the segments (see Clements 1990).<sup>5</sup> In the analysis below Hall's (1992:64f.) proposal was adopted not to differentiate between stops and fricatives. Furthermore, as this study does not consider possible differences between strong and weak elements of the nucleus, in accordance with Maddieson 2013 and Kohler 1995, V represents both long and short vowels, as well as diphthongs. Under this approach, a separate nucleus tier, as set by Clements & Keyser (1983:16), can be dispensed with.

Based on this approach, a representation of demisyllables is derived (Clements 1990), which differentiates initial demisyllables (onset + nucleus) from final demisyllables (nucleus + coda). In so doing, the segments are integrated into a base model of phonotactic structure, consisting of the three subsyllabic constituents: onset, nucleus, and coda. A rhyme tier is not necessary because it would be coextensive with the final demisyllable. For convenience, the term *rhyme* is used continuously even though it is to be understood as a cover term for a final demisyllable. Finally, from the VGLNO-tier, a simpler CV-tier is derived, aimed at the description of more common similarities between the dialects regarding syllable peaks (V, here vowels) and nonsyllabic segments (C, here consonants and glides).

The status of [ts] and [pf] needs to be established. Within a syllable or morpheme, [ts] and [pf] are usually considered to be affricates, although their status as monosegmental or bisegmental units has been discussed at length in the literature since Trubetzkoy 1939. On the one hand, it is not hard to find evidence that they can be replaced by monosegmental sounds to form new words, as in 6, which is one reason in support of defining them as single segments (affricates). This is also reasonable from the perspective of phonotactics: For example, they form the only clusters that occur in both onset and coda (further arguments are provided by Wiese 2000).

- |        |         |          |        |           |     |
|--------|---------|----------|--------|-----------|-----|
| (6) a. | [tsa:n] | 'tooth'  | [va:n] | 'madness' | STD |
| b.     | [pfa:l] | 'picket' | [va:l] | 'whale'   | STD |

---

<sup>5</sup> All data will be made publicly available as part of the *Phonotactics of Dialects in Germany* project (DFG Grant 432304149) in an open-access database permanently hosted by the *Research Center Deutscher Sprachatlas*, in accordance with the FAIR standards.

On the other hand, there are instances, where in the same environment only one of the segments is replaced by another, as shown in 7a,b (Standard German). In addition, looking at the dialects in the PAD corpus, there are instances where the second component of [ts] can be deleted, as in 7c (Gondorf, Upper Saxon region), possibly due to the quality of the vowel.<sup>6</sup> Finally, as example 7d (Linz, Upper Saxon region) shows, the stop alone can also be deleted. It is not clear how to handle instances such as 7c,d, which have not been considered in the literature to date.

(7) a.	[tsa:n]	‘tooth’	[tʁa:n]	‘blubber’	STD
b.	[tɔpf]	‘pot’	[tɔʁf]	‘peat’	STD
c.	[tentsʃ]	‘Tuesday’	[slɔntʃ]	‘Sunday’	SAX
d.	[tentsʃ]	‘Tuesday’	[tense]	‘Tuesday’	SAX

Furthermore, from an empirical point of view, it should be taken into account that all major corpus-based studies on the German language since the 1950s have been using a biphonemic definition, which raises a problem of comparability. To make sure that the data in this study can be analyzed using the models of Kohler (1995) and Maddieson (2013), I followed the tradition of treating these sounds as individual segments within a cluster (see Prinz & Wiese 1991 for a different point of view). The same holds for less homorganic but frequently-realized consonant combinations, such as [tʃ] or [dʒ], which are also defined as bisegmental sounds in the present article.

Finally, the glottal stop must be addressed, the dialectal status of which is somewhat unclear at present. Based on Pröll & Kleiner’s (2016) reading test that involved individual standard German words, the glottal stop is dependent on particular lexemes (see Alber 2001 for different results). This particular speech style effect, however, cannot be identified in the PAD corpus. As regards to southern German dialects, Caro Reina (2019:327) finds that glottal stop insertion is “highly dependent on speech tempo since it only applies in *lento* style.” This observation is in line with an observation made much earlier by Beck (1926) to which Caro Reina (2019:328) additionally refers. This dependency on the speech style explains why in the PAD data the glottal stop occurs only very rarely in

---

<sup>6</sup> This cluster ([tʃ]) is even represented graphically in family names such as *Dietzsch*, which is exclusive to the Upper Saxon region.

prevocalic word initial position, and when it does, it is very unsystematic and without any regional pattern. This is because the PAD data are comprised of longer sentences, not individual words and are thus closer to freely spoken language and allegro style. Consequently, in order not to cause any random effects, the glottal stop is not considered in the present study.

#### 4. Regional Distribution of Monosyllables.

Kleiner et al.'s (2015) examination of the Duden pronunciation dictionary revealed that 43% of Standard German equivalents of the PAD concepts are represented as monosyllables. In the PAD data, the number of monosyllables ranges between 31% and 61% per location (44% on average). This is a considerable variation, with significant spatial clustering, which was confirmed by a test of spatial auto-correlation using the global Moran's  $I$  measure ( $I = .592$ ,  $I_E = -.005$ ,  $p < .001$ ).<sup>7</sup> There are, in other words, geographical preferences for the use of monosyllables versus polysyllables.

Figure 1 shows the areal distribution of 13,492 monosyllabic tokens in the PAD corpus. Diagram A shows the number of monosyllables relative to the total number of words at every PAD location: The larger the extent of the circles, the more frequent are the monosyllables. Diagram B shows the spatial interpolation of the data in diagram A. The preferences for monosyllables are highlighted by an ordinary kriging model: Red indicates a high frequency of monosyllables, blue indicates low frequency.<sup>8</sup>

<sup>7</sup> A spatial autocorrelation test helps establish whether particular values of a spatial data distribution exist independently or in mutual dependency. The global Moran's  $I$  measure is a technique for answering this question (see Moran 1950). Here, and in the following, I report the  $I_E$  value, which is the expected Moran's  $I$  value under the null hypothesis that there is no spatial autocorrelation. The value is due to the weighted mean of four nearest neighbors. For hypothesis testing, Monte Carlo simulation with 1,000 observations was performed.  $I$ , in contrast, is the observed Moran's  $I$  value (with  $-1 \leq I \leq 1$ ), indicating, in this context, that the null hypothesis had to be rejected; spatial autocorrelation was found to be present. Analysis was performed using the  $R$  *spdep* package (Bivand & Wong 2018).

<sup>8</sup> *Kriging* is a geostatistical technique for estimating values for points where there is no direct observation. What makes the method special is that it does not

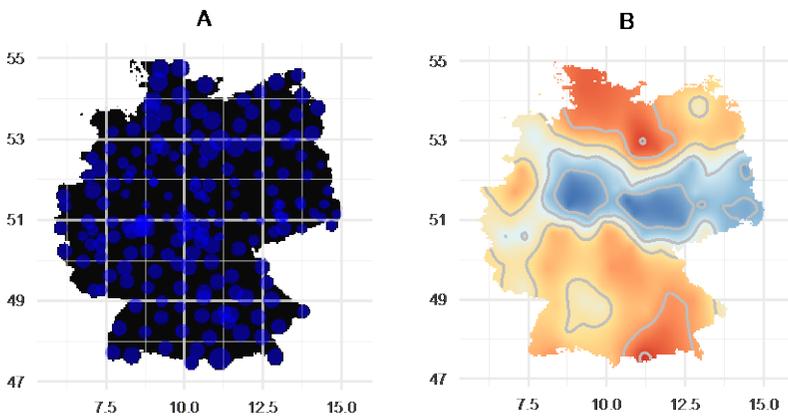


Figure 1. Areal distribution of monosyllables in the PAD corpus.

Interestingly, figure 1 reveals no explicit coincidence in relation to the larger areas of Low German or High German (see figure A in Appendix for an overview of German dialects). Nevertheless, there are core areas of both larger numbers and smaller numbers of monosyllables (red and blue color in figure 1B, respectively) in both the Low German area and the High German area. Larger numbers are concentrated in the northern region of the Low German area (northern Low German) and in the southeastern region of the High German area (Bavaria), while smaller numbers of monosyllables are concentrated in an East-West direction in the central region.

A number of phenomena contribute to the pattern in figure 1. The first group of phenomena encompasses more recent phonological processes, such as schwa apocopation, as in 8a, or *t*-deletion, as in 8b. It also includes more historical processes, such as *g*-deletion, as in 8c, or morphophonological processes, such as the deletion of the prefix *ge*-before stop, as in 8d. There are also cases in which phonological processes typical of Standard German as well as many dialects did not occur for various historical reasons, such as anaptyxis of *i*, as in 8e, and schwa anaptyxis, as in 8f. Finally, a third group involves phonological processes that are related to schwa but are only sparsely documented, such as schwa

---

interpolate linearly between individual measurement points; rather, estimates are made on the basis of spatial trends within the entire data distribution.

anaptyxis, as in 8g (note that the schwa in [dʊəp] is due to /r/ vocalization), schwa syncope, as in 8h, and lexical reanalysis, as in 8i.

- (8) a. Schwa apocopation (figure 2A)  
 [glø:f] Mirow/Pomeranian  
 [glɛ:.və] Lohra/Central Hessian  
 STD *glaub-e* ‘believe-1SG’
- b. *t*-deletion (figure 2B)  
 [bla:] Herdecke/Westphalian  
 [blɛ.tə] Bockelwitz/Upper Saxon  
*Blätter* ‘leave-PL’
- c. *g*-deletion (figure 2C)  
 [vlɛn] Volkershausen/East Franconian  
 [flɪ:.gn] Großwechsungen/Thuringian  
*fliegen* ‘fly-3PL’
- d. Deletion of prefix *ge-* before stop (figure 2D)  
 [kent] Grünow/Brandenburgian  
 [jə.kant] Kruft/Moselle Franconian  
*gekant* ‘PART.PERF-know-PART’
- e. Anaptyxis of *i* (figure 2E)  
 [mɛ.lɪç] Ballhausen/Thuringian  
 [mɛlk] Ahrbergen/Eastphalian  
*Milch* ‘milk’<sup>9</sup>
- f. Anaptyxis of schwa (figure 2F)  
 [ɛ.ləf] Aachen/Ripuarian  
 [ɛlf] Vielbrunn/Rhine Franconian  
*elf* ‘eleven’

---

<sup>9</sup> In this particular case, the change should be due to the development GERM \**meluk* > OHG *miluh* > MHG *milch*. It might be that /i/ is, in certain cases, a relic from Germanic or Old High German, while in other cases, it might be an epenthetic vowel. Due to the well-documented vowel loss in Middle High German times, /i/ is treated here as an instance of anaptyxis.

- g. Anaptyxis of schwa (figure 2G)  
 [dɔ.rəf] Pfaffenrot/Low Alemannic  
 [dɔəp] Engter/Westphalian  
*dorf* ‘village’
- h. Schwa syncopation (figure 2H)  
 [jə.ʃtɔ:.və] Bonn/Ripuarian  
 [kʃtɔəm] Ramsau/Central Bavarian  
*gestorben* ‘PART.PERF-die-PART’
- i. Lexical reanalysis (figure 2J)  
 [tense] Linz/Upper Saxon  
 [dɪ:ns.tax] Wemb/Low Franconian  
*Dienstag* ‘Tuesday’

This list is by no means complete, but it gives an idea of the diversity of the most typical phenomena reflected in the PAD corpus. Also, most of these processes are among the typical phonological phenomena in the Germanic languages discussed by Hall (2020).

Figure 2 shows the distribution of the phenomena exemplified in 8 that shape up the pattern in figure 1.

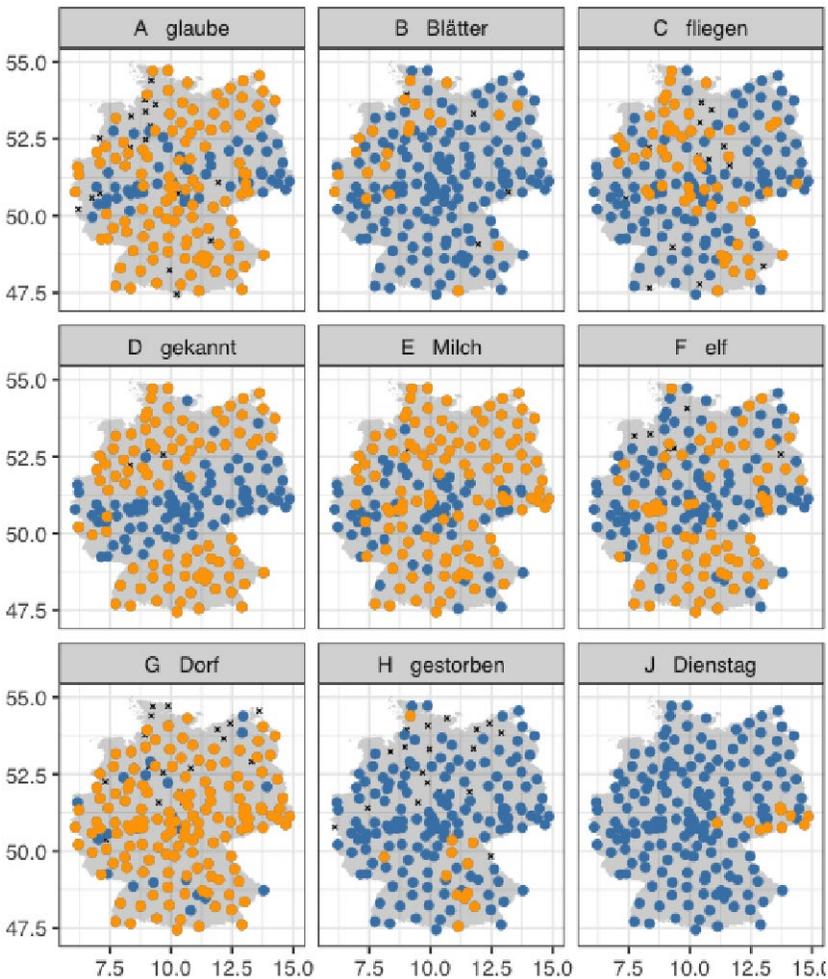


Figure 2. Number of syllables as realized in different dialects; orange=monosyllables, blue=more than one syllable, crosses=no data available.

Figure 2 illustrates that the processes in (8) affect different parts of the language region. Furthermore, they obviously do not necessarily hinder one another. For example, schwa apocopation in 8a and schwa anaptyxis in 8f occur in the same region—in western and eastern Germany. In general, processes that involve schwa seem to be especially relevant among these phenomena. This becomes evident if one combines the maps

showing schwa apocopation, as in 8a, and the deletion of the prefix *ge-*, as in 8d, which, in fact, leads to the loss of a schwa syllable. Together the two processes shape the pattern of the bluish band in figure 1B. This is not really surprising given that schwa deletion has been one of the most important processes in the development of German dialects from the 15th century onward (Lindgren 1953). The loss of schwa not only causes the loss of morphemes (see Lameli 2021) but also leads to a reduction in the number of syllables, as schwa serves as a syllable's nucleus. This, in turn, leads to an increase in the number of monosyllables in German, as stated by Werner (1978:483). At the same time, especially in the case of schwa epenthesis, context seems to be important. As figure 2 demonstrates, /l/ in 8e,f seems to provide a more productive context than /r/ in 8g. Many of the instances of schwa anaptyxis, such as in 8g—for example, in the transition zone between Alemannic/Swabian and Bavarian (central South)—could be related to the apical /r/ allophone, so that epenthesis might be due to a more general context [+sonorant, +anterior]. It is highly likely that similar conditions also hold for other phenomena. This is, however, beyond the scope of the present article.

## 5. The Syllable Structure of German Dialects.

### 5.1. The CV Pattern.

According to Maddieson (2009, 2013), Standard German is a language with complex syllables, by which he means configurations of the (C<sup>3</sup>)V(C<sup>4</sup>) type. Diagram A in figure 3 shows frequency of the CV type among monosyllables in the PAD corpus, whereas B shows the regional distribution of the CVC type, which is the most frequent syllable type. The size of the dots indicates the percentages of CVC-type syllables at the respective location; *AW* refers to the number of CVC-type syllables in the speech of the news anchor, Anne Will.

On the one hand, figure 3A confirms Maddieson's (2009, 2013) assessment of Standard German in terms of configurations, with up to three or four Cs in the onset or coda. On the other hand, 7 of the 18 monosyllabic types in the corpus have only one C in the onset and/or coda (CVC, CV, VC), and the top eight among the most frequent syllables are those with up to two Cs in the same positions (CVC, CV, CVCC, VC, CCVC, CCV, VCC, CCVCC). In contrast, syllables with CCC<sup>+</sup> clusters are rare. The most frequent type is CVC (*S*=4,447), the nucleus of which is usually either a diphthong or a long vowel. This finding is consistent with Kohler's

(1995:226) statistics for monosyllables in the standard German language, where CVC is also the most common type. Following Maddieson, this type is characterized by rather moderate complexity. In regard to the ...C<sup>5</sup> types (as in 2a,b), the PAD data provide no documentation, which is obviously due to the lack of particular lexemes in the corpus.

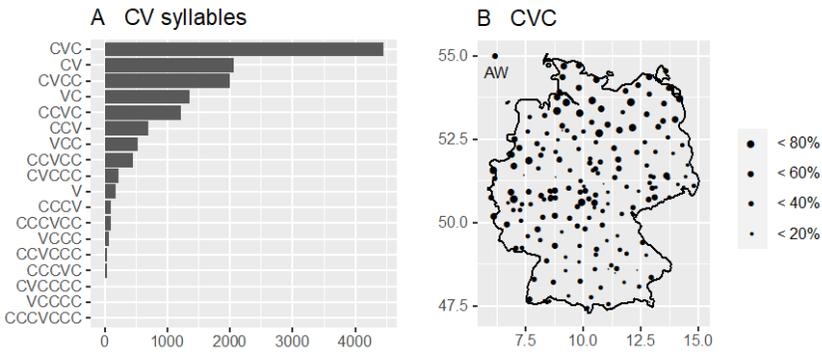


Figure 3. Frequency of CV type among monosyllables in PAD and regional distribution of CVC type.

In addition, figure 3B demonstrates that even if the CVC type can be found at every location, it is clearly dominant in northern Germany, in a smaller region in western Germany, and in the central German area. Comparing this distribution with the distribution in figure 1, a correlation between this pattern and the number of monosyllables and polysyllables is evident for the Low German area. Furthermore, in the upper left corner of figure 3B, a dot marks the spoken language of a professional news anchor on German television (Anne Will), who read the PAD items out loud.<sup>10</sup> Obviously, the news anchor demonstrates a moderate number of CVC-type syllables, thus challenging the assumption that spoken Standard German is more oriented toward northern German phonology (König 1997:250).

<sup>10</sup> This recording was made in the context of the REDE project (Schmidt et al. 2008ff.) and kindly provided by Roland Kehrein (Marburg).

### 5.2. Filling the C Position.

The areal distribution of the CVC type shown in figure 3B raises the question of whether regional preferences should be expected for other syllable types as well. As it turns out, this is indeed the case: Figure 4 shows the regional distribution of syllable types differing in the number of Cs in the onset and/or coda. Diagrams in row A show the distribution of monosyllables with only one C in onset and/or coda; diagrams in row B show monosyllables with two Cs in onset and/or coda, and diagrams in row C show monosyllables with three or more Cs in onset and/or coda. The colored lines in maps in diagram 1 in each row represent decision boundaries, based on a trained  $k$ -nearest neighbor classifier. Percentages refer to the relative number of the particular types in the total of all monosyllabic realizations at the particular location. Diagrams in A/1, B/1, and C/1 at the beginning of each row provide an overview by combining C positions to the left and right of the nucleus. The maps differ only in the number of Cs in onset/coda positions.<sup>11</sup> As a reference point, the number of clusters corresponding to the spoken language of a professional news anchor is highlighted (AW; see previous section). A more precise differentiation between onsets and codas is provided by the maps on the right (A/2, B/2, and C/2 versus A/3, B/3, and C/3).

As figure 3B reveals, monosyllables with a single consonant in either onset or coda, that is, CV... or ...VC (figure 4, row A;  $S=7,855$ ) are most frequent in the PAD corpus; they represent almost half of all instances. All the more remarkable is the fact that there is a clear preference for the CV and VC types in the northern region and parts of central Germany as opposed to southern Germany.

---

<sup>11</sup> Using a machine learning approach, isoglosses have been identified and integrated into these maps. In particular, trained  $k$ -nearest neighbor classification has been used in order to identify so-called decision boundaries. This approach was performed using the *R* FastKNN package (Besanson 2015). The boundaries found in this process must be considered to have a statistical nature.

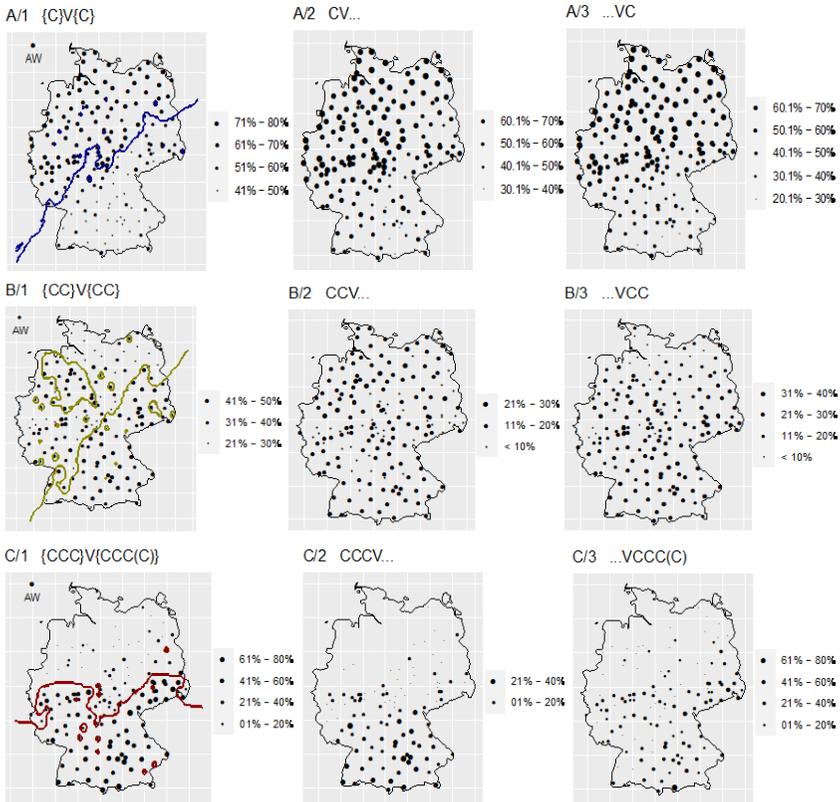


Figure 4. Spatial distribution of different types of monosyllables from the PAD corpus with up to four consonants in either onset or coda.

The distributional pattern of the most complex but also rarest syllables with three or more Cs ( $CCC^+$ ) is almost the opposite. These syllables dominate the southern and central regions of Germany (figure 4, row C;  $S=567$ ). There are two main reasons for their (relatively) high number in southern Germany. First, some of the  $CCC^+$  onset clusters contain an obstruent + /j/, as shown in 9 (see also the examples in 3 and example 12d). Historically, such  $CCC^+$  clusters result from schwa syncope and thus represent a more recent stage of language development (Werner 1978).

- (9) a. [pftelt] SWAB  
 [bə.ftelt] STD  
 ‘ordered’ Bempflingen, Swabian
- b. [kftɔln] FRC  
 [gə.fto:.lən] STD  
 ‘stolen’ Weidenbach, East Franconian
- c. [xftɔm] BAV  
 [gə.ftɔx.bən] STD  
 ‘died’ Langenbruck, Central Bavarian

Second, some of the onset clusters are due to the High German consonant shift, which led to the emergence of /pf/ and /ts/ clusters in most parts of the southern (as well as central German) regions and, additionally, the cluster /kx/ in the Southwest, near the border with Switzerland. Naturally, the question arises as to what extent the patterns in row C of figure 4 depend on /pf/ and /ts/ being defined as bisegmental clusters. As it turns out, however, defining /pf/ and /ts/ as single segments would not alter the picture significantly: The southwestern region (Alemannic) would, indeed, thin out slightly, but the North–South contrast would remain. The eastern (Bavarian) region would remain unaffected as well, due to the fact that more consonant clusters in this area come from the loss of schwa, as in 9.

In addition, there are some other, more rare phenomena, which are due not to language history but to particular phonological processes. For example, CCC coda clusters as in POM [ontʃ] ‘our.DAT-SG.N’, SWAB [ɛlpf] ‘eleven’ or MERC [alts] ‘as’ are due to a process of stop insertion between a sonorant and a fricative in the coda position of the type  $\emptyset \rightarrow P / \{N,L\} \_ F ]_{\sigma}$  under the condition that P is homorganic with either {N,L} or F.<sup>12</sup> Although this particular rule does not systematically apply in

<sup>12</sup> A rule can be stated featurally as follows:

$$\emptyset \rightarrow \left[ \begin{array}{c} -son \\ -cont \\ \alpha place \end{array} \right] / \left[ \begin{array}{c} +son \\ +cons \\ \alpha place \end{array} \right] \text{---} \left[ \begin{array}{c} -son \\ +cont \\ \alpha place \end{array} \right] K_0^n ]_{\sigma}$$

Standard German (STD [ons], [ɛlf], [als]), the phenomenon itself is quite common and is an instance of coarticulation. Although more detailed research is needed on this topic, it is possible that due to regional restrictions there could be a phonological difference between Standard German and dialects.

Finally, the most diffuse clusters are CC (figure 4, row B;  $S=4,862$ ), which show no clear regional pattern. The only syllable type that shows a slight regional tendency is VCC, as its frequency increases from North to South.

The consolidated data on all C positions in figure 4 (maps A/1–C/1), show that particular syllable types become more common as one moves from North to South. In particular, the number of C positions in both the onset and the coda tends to increase:  $C > CC > CCC^+$ . It was hypothesized in section 2 that should there be regions with preferences for monosyllables, these regions could reveal larger C clusters. Contrasting figure 4 with figure 1 shows that this is the case in the southern region but not in the northern region, where CVC is the most prominent type.

Finally, syllables without Cs also occur. In this respect, figure 5 indicates first that syllables consisting of only a V are rare ( $S=172$ ), and second that these syllables mostly occur in the Upper German area, with most of them in the Alemannic language group (including Swabian).

Based on figures 4 and 5, the southwestern region (primarily the Alemannic and Swabian dialect region) is of particular interest, as it exhibits both extremes—the V-only syllables and the  $CCC^+$  clusters (even though the latter are not as frequent as they are in the southeastern region—Bavarian). The V-only syllables occur in weak forms, as shown in 10.<sup>13</sup>

---

This rule feeds final obstruent devoicing, which is typical in all German dialects, as well as in Standard German.

<sup>13</sup> V-only realizations are also found in Standard German ([e:] ‘anyway’; [aɪ] ‘egg’). These lexemes are, however, not documented by the PAD corpus. Other V-only syllables, such as [i:], [a:], [o:], and [ʊɪ] are typical Standard German interjections.

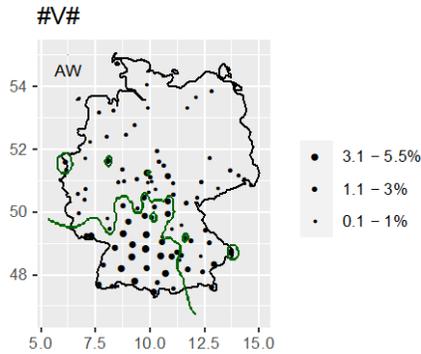


Figure 5. Spatial distribution of V-only syllables in the PAD corpus.

- |         |        |                            |
|---------|--------|----------------------------|
| (10) a. | [ɑ:]   | ALEM                       |
|         | [aʊx]  | STD                        |
|         | 'also' | Endenburg, High Alemannic  |
| b.      | [i:]   | BAV                        |
|         | [ɪç]   | STD                        |
|         | 'ɪ'    | Ulbering, Central Bavarian |

Almost none of these syllables come with a glottal stop, and those that do follow no specific regional pattern, which, to some extent, seems to contradict the results of Alber (2001) and Caro Reina (2019), who demonstrate regional preferences for word initial glottal stop insertion, for example. However, as mentioned above, Caro Reina (2019:327f.) also finds glottal stop insertion to be dependent on *lento* style, which is not documented by the PAD data. Therefore, one can say that the Wenker sentences, as used in the PAD corpus, seem not to trigger glottal stop insertion.

### 5.3. Modeling of CV/VC Sequences in Monosyllables.

While figure 3A shows the frequency of the CV pattern (that is, any syllable consisting of any number of C and/or V), with the CVC type as the most prominent one, figure 3B shows the regional distribution of the latter. In the previous sections, these syllable types were discussed in terms of their position within words, with the most extreme case being a VC<sup>4</sup> sequence. The discussion in the following sections moves down to the

level of segments: I analyze the internal structure of these syllables in order to identify possible combinations of segments in them. Based on the results of this analysis regional models of segment sequences are then developed.

In order to translate the insight in figure 3 into the actual ordering of segments (for instance, CV versus VC), northern Germany, which is dominated by the CVC structure, was contrasted with southern Germany, where a preference for more complex C clusters was found as well as for V-only syllables. For a more detailed description, 33 locations across three different regions—one from northern Germany and two from southern Germany—were randomly chosen, with 11 locations per region ( $L=33\sim 18\%$  of the 182 PAD locations). These locations are highlighted in figure 6.

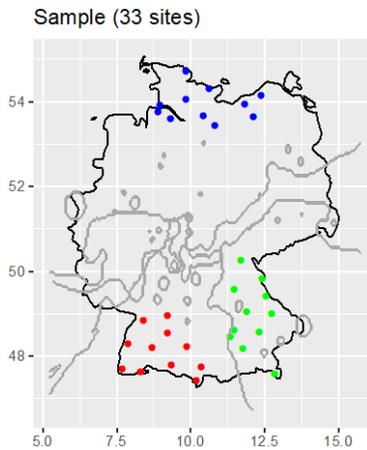


Figure 6. North versus South comparison ( $L=33$ ), together with the decision boundaries from figure 4.

For all three regions, the different CV types were combined across all locations. On this basis, a transition matrix of Cs and Vs was generated in order to assemble individual Markov chains. The transition matrix contained the relative frequency of the joint occurrence of each pair of segments. By way of example, table 1 reports the transition probabilities between segments in CVC syllables, comparing northern Germany with southern Germany. Regional differences in terms of frequency of CVC syllables can be seen in figure 3B. By breaking CVCs down to their

individual C–V–C segments (including the initialization ( $\alpha$ ) and the closure ( $\varepsilon$ ) of the syllable), table 1 illustrates how frequent the individual combinations of these segments are. An overall probability for the entire syllable can then be derived by multiplying the frequency of every segment.

Region	$\alpha \rightsquigarrow C$	$C \rightsquigarrow V$	$V \rightsquigarrow C$	$C \rightsquigarrow \varepsilon$	Total
North	85%	82%	79%	84%	46.26%
Southwest	84%	77%	74%	68%	32.55%
Southeast	85%	74%	75%	67%	31.61%

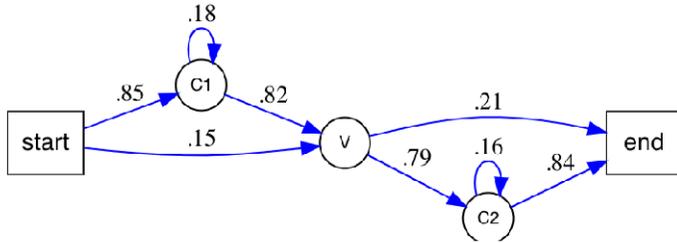
Table 1. Transition probability  $P$  for the prototypical CVC syllables.

As can be seen from table 1, the probability of a CVC syllable is 46.26% in northern Germany (that is,  $.85 \times .82 \times .79 \times .84 = .4626$ ), which is considerably higher than in either southwestern or southeastern Germany. This is to be expected in light of figure 3B. What table 1 does indicate is where transitions between the segments are similar and where they are different. The difference in CVC frequency is thus specified not as the frequency of an individual syllable, but as a difference in possible segment combinations inherent in the dialects. Since these segmental combinations also constitute part of the structure of other syllable types (for example, the VC type), table 1 provides a more refined picture of the possible structure of monosyllables in dialects.

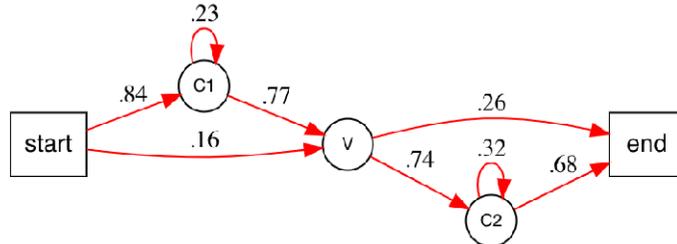
For a better overview of the possible combinations, diagrams of transition probabilities were modeled. This is where Markov chains (which can be understood as finite state automata) become relevant: They provide information on the probability of one particular state becoming another. In the present case, this is the probability  $P$  for the transition from one segment  $i$  of the CV representation to another one  $j$  within the same syllable, taking into account the actual initialization and closure ( $\alpha$  and  $\varepsilon$ ) of the syllable ( $\Rightarrow P_{ij}$ , with  $i, j \in \{\alpha, C, V, \varepsilon\}$ ).<sup>14</sup> The actual models obtained from the transition matrix underlying table 1 are specified in figure 7, where Markov chains represent the transition probability  $P$  of segments.

<sup>14</sup> For logical reasons, zero probability is assumed for the recursion of these edges. The present analyses were performed using the *R* Markov chain package (Spedicato et al. 2020).

North



Southwest



Southeast

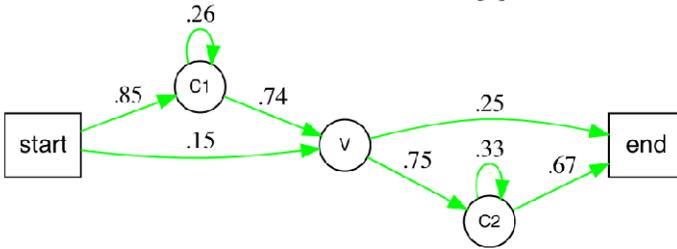


Figure 7. Markov chains for the northern, southwestern, and southeastern locations ( $L=33$ ).

The models in figure 7 show sequences of possible C and V states where the probability of the next state depends on the current state. Note that *C1* and *C2* do not refer to particular sounds but to the positions (or states) where consonants might occur. In the case of C clusters, these positions are filled repeatedly. Accordingly, *V* represents the position of the nucleus. Since the corpus covers only monosyllables, the condition applies that only one *V* per chain is possible. *Start* ( $\alpha$ ) initializes the left syllable edge, and *end* ( $\epsilon$ ) closes the right syllable edge. The models can be read as follows: The probability  $P$  of a monosyllable starting with C is  $P_{\alpha C1}=85\%$  in the northern model,  $P_{\alpha C1}=84\%$  in the southwestern model, and  $P_{\alpha C1}=85\%$  in the southeastern model, while the probability of *V* being the first segment in the word is  $P_{\alpha V}=15\%$  to  $16\%$  to  $15\%$ . According to the highest values in the models

shown in figure 7, the most likely sequence in all three regions is CVC, which concurs with figure 3A and table 1.

Figure 7 and table 1 demonstrate an almost identical probability of  $\alpha$ C transition across all three regions ( $P_{\alpha C1}=84\%–85\%$ ). This is consistent with Kohler's (1995:226) statistics for monosyllables in the German standard language, where almost all monosyllables start with C. At the same time, these results show that dialectal monosyllables tend to comply with the Head Law (Vennemann 1988:13f.), according to which (additional prosodic characteristics aside) a "syllable head is the more preferred [...] the closer the number of speech sounds in the head is to one" (see also figure 3A).

However, most interesting is the replication rate of C states, which is the probability that one C is followed by another. Two findings need to be highlighted. First, replication of C states is much higher in the southern models than in the northern model. This holds for both onsets (Southwest  $P_{C1C1}=23\%$  and Southeast 26% versus North  $P_{C1C1}=18\%$ ) and codas ( $P_{C2C2}=32\%$  and 33% versus  $P_{C2C2}=16\%$ ). Second, the same numbers indicate a clear difference between onsets and codas in the southern models but not in the northern model. This difference corresponds with lower values for CV transitions ( $P_{C1V}=\text{North } 82\%; \text{ Southwest } 77\%; \text{ Southeast } 74\%$ ) and C transitions in the South ( $P_{C2\varepsilon}=\text{North } 84\%; \text{ Southwest } 68\%; \text{ Southeast } 67\%$ ). In other words, as in Standard German, the coda is only the preferred domain of complex C clusters in the southern dialects. The conflict with Vennemann's Coda Law (1988:21) mentioned for Standard German (see example 2) also holds for the South of the language area but not for the North. Regarding the V-only syllables, which were more commonly found in southern Germany, no relevant influence on the southwestern model is obvious.

Taken together, the models thus illustrate important differences in the clustering of consonants, which leads to the question on the prosodic structuring of syllables. For example, are the segments sequenced depending on their degree of sonority and in accordance with the Sonority Sequencing Generalization, which states that in any syllable "there is a segment constituting a sonority peak that is preceded and/or followed by a sequence of segments with progressively decreasing sonority values" (Selkirk 1984:116)? An answer is provided in the following section.

5.4. *Sonority Dispersion.*

In order to identify the possible regional distribution of syllables based on their prosodic properties, an analysis of SONORITY DISPERSION (Clements 1990:302–311) was performed. Sonority dispersion is another way to describe syllable complexity, based on the organization of segments within the syllable according to their degree of sonority. In essence, the aim of a sonority dispersion analysis is to identify sonority differences between adjacent segments in a syllable. To this end, syllables are divided into demisyllables: onset + nucleus versus rhyme (that is, coda + nucleus), with V as the segment shared by the initial and the final demisyllable). Thus, each demisyllable is a sequence of the form  $C_m \dots C_n V$  or  $VC_m \dots C_n$ , with  $n \geq m \geq 0$  (Clements 1990:303). Based on this configuration, sonority dispersion  $D$  is calculated as follows:

$$(11) D = \sum_{i=1}^m \frac{1}{d_i^2}$$

In 11,  $d$  is the difference between the sonority ranks of every adjacent and nonadjacent pair of segments and  $i$ , within each demisyllable, following the sonority scale  $V > G > L > N > O$ . This sonority scale represents the ranks  $4 > 3 > 2 > 1 > 0$ , while  $m$  refers to the number of pairs in the demisyllable.<sup>15</sup> For example, the dialect word [mɛlk] ‘milk’ in 8e has the initial demisyllable (ID) [mɛ] and the final demisyllable (FD) [ɛlk], which form the pairs ID={mɛ} and FD={ɛl,ɛk,lk}. Based on these pairings, sonority dispersion is calculated in 12.

(12) a. Initial demisyllable

$$\begin{array}{l} m\varepsilon \quad NV = 1 / (1-4)^2 = 0.11 \\ \quad \quad D_{ID} = 0.11 \end{array}$$

---

<sup>15</sup> Clements (1990) takes this formula from physics, where it has been used in the computation of forces in potential fields. Clements (1990:304) mentions that the formula has also been used in perceptual linguistics. For a critical assessment, see Parker 2002:22–24.

## b. Final demisyllable

$$\epsilon l \quad VL = 1 / (4-2)^2 = 0.25$$

$$\epsilon k \quad VO = 1 / (4-0)^2 = 0.06$$

$$lk \quad LO = 1 / (2-0)^2 = 0.25$$

$$D_{FD} = 0.25 + 0.06 + 0.25 = 0.56$$

Further examples provided by Clements (1990:304) are GV, VG=1.0 or OV, VO=.06 for CV/VC demisyllables and OLV, VLO=.56 or LGV, VGL=2.25 for CCV/VCC demisyllables. From these examples it becomes evident that segment pairs that follow more closely the aforementioned sonority scale have higher *D* scores, while pairs that skip segments on the sonority scale (for example, VO lacks G, L, N) have lower *D* scores. At the same time, the examples illustrate that *D* takes into account not so much the violation of the Sonority Sequencing Generalization but the (non)adjacency of the segments within the defined sonority scale.

In order to evaluate the PAD data, the monosyllables from PAD were represented as VGLNO-sequences and then divided into demisyllables on which the *D* measure was performed using R programming (V-only demisyllables were excluded). For every location *j* the mean sonority dispersion *D<sub>j</sub>* was calculated, which leaves one with a total of 182 average *D<sub>j</sub>* measures per demisyllable.

According to Clements's (1990:304) crosslinguistic observations, the preferred initial demisyllable (onset + nucleus) tends toward lower *D*, while the preferred final demisyllable (rhyme) tends toward higher *D* (the so-called Dispersion Principle):

We observe that initial demisyllables with low values for *D* are those that show an optimal sonority profile, i.e., a sharp and steady rise in sonority, while in the case of final demisyllables, those with high values for *D* show the best profile, i.e., a gradual drop in sonority.

Thus, according to Clements, the preferred profile of initial demi-syllables is a profile consisting of segments that are not next to each other (that is, nonadjacent) on the sonority scale, such as OV in contrast to GV. Such a profile is expressed by lower *D* values. However, in final demisyllables, the preferred profile is a profile consisting of segments that are next to each other on the sonority scale (adjacent=gradual drop), such as VGL in contrast to VLO. This preferred profile is expressed by higher *D* values.

This observation, which corresponds reasonably to both the Head Law and the Coda Law (Vennemann 1988), was confirmed by an evaluation of the PAD data. On average, sonority dispersion was found to be  $\bar{D}_{ID}=.148$  in initial demisyllables and  $\bar{D}_{FD}=.193$  in final demisyllables. This difference was confirmed by a *t*-test in statistical terms ( $t(23,781)=-13.02$ ,  $p<.001$ ). The values obtained were noticeably rather low, indicating a predominance of CV (79% of which=OV, 13%=NV) and VC demisyllables (65%=VO, 25%=VN), ranging between  $D=.06$  and  $D=1$ . It must be assumed that this is not least due to the dominance of CVC syllables. An overview of the most frequent types of both initial and final demisyllables is provided in table 2.

Rank	Initial Demisyllable			Final Demisyllable		
	Type	Count	%	Type	Count	%
1	OV	6905	51.18	VO	4666	34.58
2	V	2125	15.75	V	3022	22.40
3	OLV	1468	10.88	VN	1677	12.43
4	NV	1118	8.29	VOO	1293	9.58
5	OOV	709	5.25	VLO	878	6.51
6	LV	609	4.51	VNO	738	5.47
7	OOOV	214	1.59	VL	707	5.24
8	ONV	207	1.53	VLOO	179	1.33
9	GV	96	0.71	VLN	111	0.82
10	OOLV	18	0.13	VNOO	104	0.77

Table 2. Most frequent demisyllable types in the PAD corpus.

In order to test for regional differences, average *D* was calculated for every location documented in the corpus and plotted in figure 8. The upper panel (figure 8A,B) shows the mean scores for sonority dispersion per site. For both conditions, initial and final demisyllables, a Moran's I test revealed significant spatial autocorrelation ( $I_{ID}=0.172$ ,  $I_E=-0.005$ ,  $p<.001$ ;  $I_{FD}=0.294$ ,  $I_E=-0.005$ ,  $p<.001$ ).<sup>16</sup> That is, significant regional differences

<sup>16</sup> As an alternative measure, *D* was also calculated ignoring nonadjacent pairs (such as [ek] in example 7), as proposed by Clements (1990). This, however, revealed no major difference. Even though the *D* values were slightly higher, the relationship between locations in respect of their mean *D* value was equivalent.

in  $D$  values were found to be present. The bottom panel (figure 8C,D) reports the results of a test for spatial association following Getis & Ord 1992, 1996.<sup>17</sup> The locations identified here are those that statistically i) have particularly high (brown color) or low (green color)  $D$  scores, and ii) are adjacent to locations with similar scores. However, because the values are so small, they should only be considered indicative of a tendency.

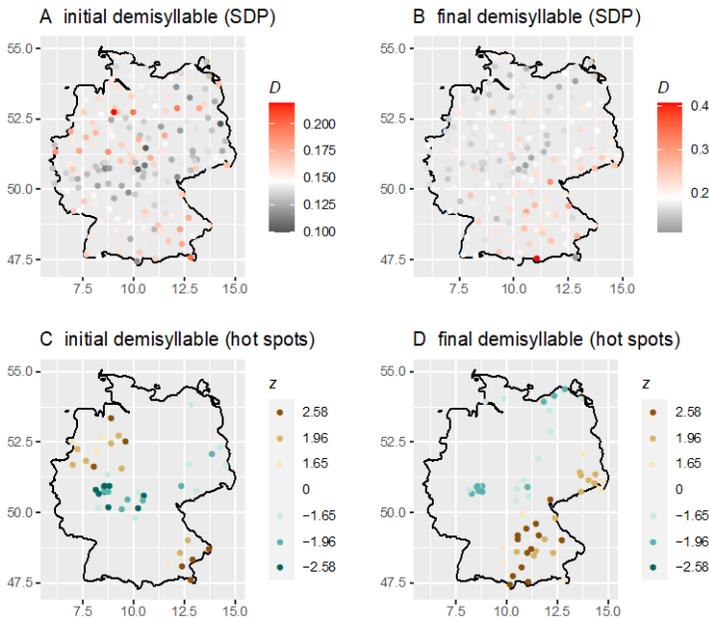


Figure 8. Regional distribution of sonority dispersion measure ( $D$ ).

In diagrams A and B,  $D$  is shown for initial and final demissyllables (onset + nucleus versus rhyme) and white color indicates the average of the distribution. Diagrams C and D show  $G_i^*$  scores of the  $D$  distribution for both initial and final demissyllables. Note that  $D$  scores in C and D have been  $z$ -transformed, which means that the mean of the distribution is zero and the standard deviation is  $z=1$ .

<sup>17</sup> The so-called  $G_i^*$  statistics proposed by Getis & Ord (1992) aim to detect clustering of particularly high or low values within a spatial data distribution. To this end, each data point is interrelated with its neighboring points. The values of such a neighborhood relation are then validated against the overall distribution.



## 6. Modeling Monosyllables in German Dialects.

Finally, in order to obtain a comprehensive answer to the question of the phonotactic structuring of German dialects, the approach described in section 5.3 was used: A total of 55 locations were identified across five different regions, with 11 locations per region ( $L=55 \sim 30\%$  of PAD locations). In each region, Markov chains for both initial and final demisyllables were calculated, as shown in figure 9. It represents the transition probabilities  $P$  for each pair of adjacent segments in both the initial and final demisyllables (probabilities  $P < 5\%$  are not reported). Red arrows indicate high transition probability and blue arrows indicate low transition probability. Location subsets within regions A, D, and E are the same as in the previous analysis (figure 6). B represents the more western region, with a lower occurrence of monosyllables (figure 1), and C is a location subset within the central region, with lower  $D$  values in the onset (figure 8A,C).

With respect to the initial demisyllable, in the region A model, a syllable most likely starts with an obstruent ( $P_{ao}=72\%$ ), followed by a vowel ( $P_{ov}=76\%$ ) and another obstruent ( $P_{vo}=45\%$ ). This shows that the coda is the more diverse subsyllabic unit (for example, onset:  $P_{ol}=14\%$ , coda:  $P_{lo}=44\%$ ; onset:  $P_{og} < 5\%$ , coda:  $P_{go}=22\%$ ). In regard to the filling of onset and coda, one is more likely to find an open coda ( $P_{ve}=21\%$ ) than an open onset ( $P_{av}=15\%$ ), which is in line with more general expectations. The same tendency holds for all other regions. However, there is a clear decrease in  $P_{ov}$  transition from North to South (A:  $P_{ov}=76\%$ ; B:  $75\%$ ; C:  $75\%$ ; D:  $69\%$ ; E:  $66\%$ ), indicating more complex O clusters, which is confirmed by the increase of  $P_{oo}$  (A:  $P_{oo}=07\%$ ; B:  $07\%$ ; C:  $12\%$ ; D:  $15\%$ ; E:  $17\%$ ).

In relation to the final demisyllable, the situation is rather different. In regard to O clustering, first, there is an increase from North to South (A:  $P_{oo} = 07\%$ ; B:  $16\%$ ; C:  $19\%$ ; D:  $20\%$ ; E:  $24\%$ ). Second, while  $P_{oo}$  is identical in the initial and the final demisyllable, it is much higher in the models of regions B–E. Since the northern area was selected as being representative of the CVC region, this is the expected result.



E: 40%), which contributes to higher *D* values in the South. Interestingly, this transition is observed in the Alemannic/ Swabian region (*D* in figure 9), and the values are higher here than in the Bavarian region (*E* in figure 9). In contrast, there are higher values for  $P_{Le}$  and  $P_{OO}$  in southeastern Germany, which leads to slightly sharper contrasts in Bavaria. This difference is also evident in figure 8D, where in the eastern region of southeastern Germany *D* values occur around the mean or even in the first quartile.

More generally, figure 9 demonstrates that on this level of phonological representation, almost no violation of the Sonority Sequencing Generalization is in evidence above the 5% cut. Certainly, a more fine-graded scale differentiating, for instance, stops and fricatives, would reveal a more detailed picture. However, the OLGV representation adopted in this study obviously enables very clear modeling of the given data. The only exceptions are GL transitions in the initial demisyllables in the region A model ( $P_{GL}=20\%$ ), which are due to lenition from [g] to [j], as in 14, which is well known in Low German dialects (Simmler 1983).

(14) [jɾəʊ̯] ‘great’

Diekhusen, North Low German

From a more theoretical perspective, however, this exception can be eliminated if one interprets these /j/ sounds not as glides, but as fricatives (see Hall 2014:334 for similar considerations with respect to Westphalian glides). Since fricatives are obstruents, the OLGV order is not violated.

## 7. Discussion.

### 7.1. Syllable Structure and Regional Patterns.

Prior work modeled sound sequencing in syllables of Standard German. Based on previous research, Maddieson (2013) defines German as being among the languages with the highest degree of syllable complexity in terms of C clustering in onset and coda. Expanding the focus on language variation, the present study contributes to a more differentiated view. Hypotheses i)–iii) outlined in section 2 have been confirmed: There have been found significant differences in the architecture of syllables between German dialects and Standard German; furthermore, regional differences in the structuring of monosyllables in German dialects have been identified, and these differences mainly concern the complexity of C clusters in both onset and coda.



coincides almost perfectly with the region where monosyllables are less frequent in figure 1.<sup>18</sup>

These general findings lead to a more specific result, namely, the North–South divide in terms of syllable complexity. This result has been substantiated not in the least by the probabilities of segment transition. Focusing on individual phenomena, such as schwa apocopation, the analysis has revealed a clear (categorical) separation between North and South. In addition, the models that were developed based on the stochastic application of Markov chains reveal a continuous increase in obstruent clusters in both onset and coda from North to South. The most prominent areas are the northern coastal region (that is, northern Low German and Pomeranian, see Appendix) and the southwestern region (that is, Alemannic, including Swabian) together with the southeastern region (that is, parts of Bavarian).

It must be noted that these regions represent larger areas with similar phonotactic characteristics. Specifically, the pattern of the Bavarian region seems to reach Upper Saxon, as figure 5B indicates. However, altogether, these regions define the hot spots of the North–South divide, where the northern dialects tend toward more simple syllable structures and the southern dialects toward more complex clusters. Consequently, with regard to Maddieson’s definition of syllable structure, it is assumed that northern and southern dialects could represent two different types, among which the syllable structure of Low German dialects is “moderately complex” and the syllable structure of High German dialects is “complex” (Maddieson 2013).

If the zone between (C)V(C) and (C<sup>3</sup>)V(C<sup>3+</sup>) (see figure 4A/C) is assumed to be the border zone between the North and the South region, it must be stated that this geographical division does not correspond to traditional dialect classification. However, in part this finding is in line with more recent studies on the connection between speech rate and regional varieties (Hahn & Siebenhaar 2016, Siebenhaar & Hahn 2019). These studies also find a limited correspondence between regional borders

---

<sup>18</sup> See <https://www.regionalsprache.de/Map/5uEUewyu>, last accessed September 16, 2020.

and traditional dialect classification.<sup>19</sup> As has been shown, it is the deletion of schwa and other schwa-related phenomena (such as *ge-* deletion) that have a huge impact on the architecture of phonological systems and are thus a major issue in the regional differentiation of both lexical phonology and syllable phonology (Werner 1978, Birkenes 2014).

As a rough synopsis of the analyses, three syllable types can be highlighted: CVC, V, and (C<sup>3</sup>)V(C<sup>3+</sup>). With respect to the latter, its connection with the deletion of schwa and other schwa-related phenomena has already been established. It has also been demonstrated that with regard to coda clusters, the Bavarian language in the western region—more precisely, the contact area between Bavarian and Swabian Bavarian—primarily exhibits the most frequent occurrences of the (C<sup>3</sup>)V(C<sup>3+</sup>) type. Equally, it is in this region where the ordering of segments in the coda follows the sonority scale most consistently (expressed as a measure of sonority dispersion, figure 8B,D). This leads to the conclusion that with increasing C clustering, mechanisms that order segments according to their sonority are at work. At the same time, regional characteristics can be observed, such as the more frequent NO transitions in southwestern Germany as opposed to a sharper sonority contrast between the segments in southeastern Germany (see the Markov chains for demisyllables in figure 8D,E). In all cases, however, the Sonority Sequencing Generalization seems to apply. Exceptions, such as a putative GLV sequence in 14, resolve when the glides in question are considered as fricatives.

V-only syllables, by contrast, occur almost exclusively in the southwestern region. As they occur mainly in weak forms, such as pronouns or prepositions, further studies could investigate whether these forms have a specific function in higher level prosodic categories, such as intonational or phonological phrases. In any case, the V-only syllables contribute to the fact that the southwestern region seems to be a region of phonotactic particularities, as has also been demonstrated in other studies (see among others Nübling & Schrambke 2004, Caro Reina 2019).

---

<sup>19</sup> At least, the separation found between the (C)V(C) pattern (figure 4B) and sonority dispersion in final demisyllables (figure 8B) resembles the separation between Low German, together with western German, found in a dialectometric analysis by Lameli (2013:194). However, one should not disregard the possibility that this occurs purely by accident.

In regard to the distribution of CVC syllables, the situation is different. Even though this syllable pattern occurs in all regions, it is particularly common in northern Germany. Furthermore, as figure 3B shows, it is less frequent in those regions where schwa deletion did not take place. One possible explanation for this could be that forms with historical schwa apocopation may be realized as monosyllables, whereas forms where schwa was not deleted (or reinserted) are realized as polysyllabic, as shown in 16 (see section 3.2 for the handling of V).

- (16) a. CVC [seɪp] ‘soap.ACC.SG’ Haddorf, Northern Low German  
 b. CV.CV [saɪ.fə] ‘soap.ACC.SG’ Schellroda, Thuringian  
 c. CV.CV [sɔɐ̯.fɪ] ‘soap.ACC.SG’ Oberau, Bavarian

From this perspective, that CVC syllables tend to occur in the absence of schwa apocopation is in line with the hypothesis of different syllable structures inside and outside the region of schwa apocopation (see Lameli 2021). However, the stark difference between North and South in terms of C clustering suggests that there must be other factors influencing the particular architecture of syllables in different regions. In this regard, it should be considered that Low German schwa apocopation is a more recent process (documented around the second half of the 16th century; see Foerste 1966) than High German apocopation (documented around the middle of the 13th century; see Lindgren 1953), so that the much higher frequency of CVC in the coastal region is due to a different, albeit also schwa-related development. Furthermore, as this region forms part of the linguistic continuum between the Low German-speaking area and northern Europe (see Höder 2016), further studies should investigate the extent to which this pattern is related to the regional characteristics of the syllable structures of neighboring languages, such as Danish.

In addition, 16c exemplifies morphological conditions that were not addressed in this study: It represents a relic of weak inflection, typical of Bavarian, where the Early New High German *-en* ending (NOM.SG.F.) underwent schwa syncope in more recent times. As a result, /n/ became the new—now consonantal—nucleus. The phonology–morphology interface has been subject of recent studies on Standard German (Bergmann 2018 among others). Examples such as 16c suggest that it would be worthwhile extending this research to German dialects (see also Birkenes 2014). It thus becomes evident that the phonotactics of dialects

contributes not only to a more comprehensive description of language geography but also to a more comprehensive understanding of German as a variative language.

### *7.2. Methodological Reflection.*

With its quantifying approach to regional phonotactics, the present study has entered a field that has been underexplored so far. Different approaches, such as the measure of sonority dispersion and the modeling of Markov chains, were adopted for the analysis in order to gain different perspectives on the subject. It goes without saying that each of these approaches is able to yield only an approximation to the phonotactic structure of dialects, as the results depend on a particular model, which, in turn, can be questioned. The representativeness of data is a crucial point. The analysis focuses on the phonotactic structure of a sample of 201 monosyllabic words (realized as 13,492 tokens) representing different parts of speech. In total, approximately half of the Wenker sentences were studied. The question is whether the results can be considered to be generally representative. To answer this question the potential phonotactic possibilities in dialects needs to be clarified. However, since no comparative studies are available in this area, such an answer is hardly possible. An alternative would be to examine a selection of different phonological processes known from the standard language; but such an approach would very likely give rise to problems associated with a lack of regional balancing, as Standard German is affected by different dialects, especially by East Middle German dialect variants.

Against this background, an explorative approach based on the PAD data seemed to be a good starting point in order to construct a big picture of the phonotactics across regions in the first place. More importantly, the words analyzed in this study are representative of the Wenker sentences. This is not a small claim, considering that knowledge about the dialects in Germany is substantially shaped by translations of the Wenker sentences. Thus, this material serves as common ground between this study and previous research. For example, the fact that the phonological processes discussed by Hall (2020) are documented in the PAD corpus points to the fundamental suitability of the PAD data for phonotactic analysis. At the same time, it is also clear that some prosodic phenomena (such as tonal accents) are underrepresented in the PAD. It would therefore be important to further expand the corpus in the future.

Furthermore, the use of Markov chains must be addressed. In the present study, First Order Markov Chains were used to develop a memoryless transitional model consisting of known states (for example, C, V, start, and end). There are certainly more powerful techniques, such as Hidden Markov Models, which are also suited for modeling latent effects related to unknown states (see, among others, Tjong-Kim-Sang 1998). These techniques could be used in future work when more information on the underlying phonological systems is available. For the present exploratory purposes, however, the chosen method was not only appropriate but also very powerful.

In regard to the use of sonority dispersion, this approach has provided a new perspective on regional phonotactics. In the present study, Clements's (1990) simplified sonority scale was used. This scale is advantageous in an exploratory study because it allows to effectively capture rough structures. By contrast, a more detailed sonority scale would be advantageous, for example, to obtain a more fine-grained picture of the southern regions, where more C clusters clearly occur. This would be a worthwhile task for future work.

## **8. Conclusion.**

The primary goal of this article was to explore the areal variation of syllable structure in German dialects, in order to take a step toward a more comprehensive view of areal phonotactics. To this end, 13,492 tokens of monosyllables across 182 locations were analyzed. By focusing on the structure of monosyllables, new insights into the phonological micro-variation of German were gained. In particular, new information on dialectal variation came to light, which concerns the structure of syllables as well as specific arrangements of individual segments, both of which vary between northern and southern Germany. The study reveals that monosyllables are quite frequent in both the North and the South. The most typical monosyllable in northern German dialects is CVC, while the dialects in southern Germany tend toward more complex monosyllables, with obstruent clusters. An analysis of sonority dispersion yields two major findings: On the one hand, final demisyllables in southern German tend to be made up of segments more closely arranged on the sonority scale. On the other hand, initial demisyllables in middle German are often comprised of segments that are far apart on the sonority scale. This study

thus constitutes a good starting point for further crossregional, in-depth analysis of micro-variation in syllable structure.

APPENDIX

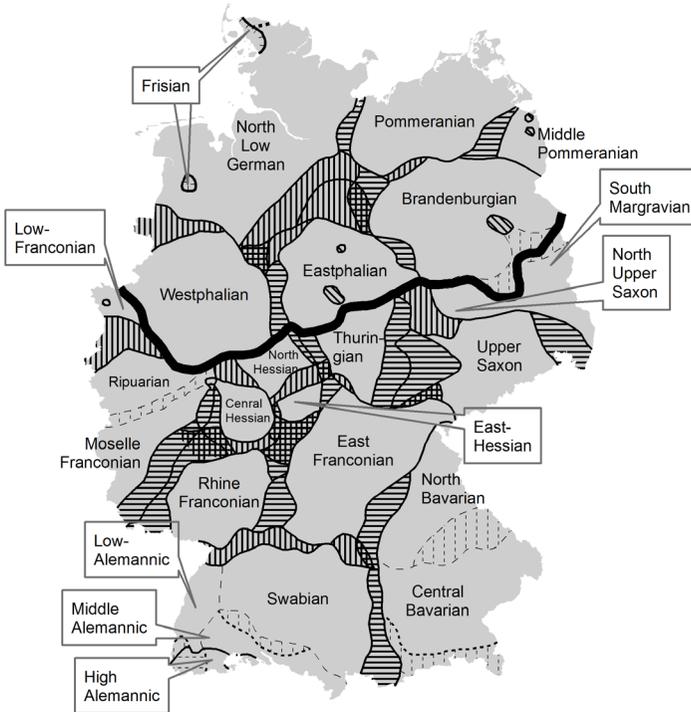


Figure A. Dialects within the Federal Republic of Germany (following Wiesinger 1983).

REFERENCES

Alber, Birgit. 2001. Regional variation and edges: Glottal stop epenthesis and dissimilation in standard and southern varieties of German. *Zeitschrift für Sprachwissenschaft* 20. 3–41.

Alber, Birgit, & Marta Meneguzzo. 2016. Germanic and Romance onset clusters—how to account for microvariation. *Theoretic approaches to linguistic variation (Linguistics Today 234)*, ed. by Ermenegildo Bidese,

- Federica Cognola, & Manuela Caterina Moroni, 25–52. Amsterdam: John Benjamins.
- Almeida, Antonio, & Angelika Braun. 1986. “Richtig” und “Falsch” in phonetischer Transkription. Vorschläge zum Vergleich von Transkriptionen mit Beispielen aus deutschen Dialekten. *Zeitschrift für Dialektologie und Linguistik* 53. 158–172.
- Auer, Peter. 1993. *Is a rhythm-based typology possible? A study of the role of prosody in phonological typology (KontRI Working Paper 21)*. Konstanz: University of Konstanz.
- Bachmann, Armin R. 2000. *Die Mundart von Eslarn in der Oberpfalz: Phonologie, Morphologie, Glossar*. Stuttgart: Steiner.
- Beck, Ernst. 1926. *Lautlehre der oberen Markgräfler Mundart*. Halle (Saale): Buchhandlung des Waisenhauses.
- Bellmann, Günter. 1970. *Tonbandaufnahme ostdeutscher Mundarten 1962–1965. Gesamtkatalog*. Marburg: Elwert.
- Bergmann, Pia. 2018. *Morphologisch komplexe Wörter: Prosodische Struktur und phonetische Realisierung*. Berlin: Language Science Press.
- Besanson, Gaston. 2015. *FastKNN. R package version 0.0.1*.
- Best, Karl-Heinz. 2013. Silbenlängen im Deutschen. *Glottology* 4. 36–44.
- Birkenes, Magnus Breder. 2014. *Subtraktive Nominalmorphologie in den Dialekten des Deutschen: ein Beitrag zur Interaktion von Phonologie und Morphologie*. Stuttgart: Steiner.
- Bivand, Roger, & David W. S. Wong. 2018. Comparing implementations of global and local indicators of spatial association. *TEST* 27. 716–748.
- Caro Reina, Javier. 2019. *Central Catalan and Swabian: A study in the framework of the typology of syllable and word languages*. Berlin: de Gruyter.
- Clements, George N. 1990. The role of the sonority cycle in core syllabification. *Papers in laboratory phonology I: Between the grammar and physics of speech*, ed. by John Kingston & Mary E. Beckman, 283–333. Cambridge: Cambridge University Press.
- Clements, George N., & Samuel Jay Keyser. 1983. *CV phonology. A generative theory of syllables*. Cambridge, MA: MIT Press.
- Chambers, Jack K., & Peter Trudgill. 1998. *Dialectology*. 2nd edn. Cambridge: Cambridge University Press.
- Easterday, Shelece. 2019. *Highly complex syllable structure: A typological and diachronic study*. Berlin: Language Science Press.
- Fleischer, Jürg. 2017. *Geschichte, Anlage und Durchführung der Fragebogen-Erhebungen von Georg Wenkers 40 Sätzen. Dokumentation, Entdeckungen und Neubewertungen*. New York, NY: Olms.
- Foerste, William. 1966. Geschichte der niederdeutschen Mundarten. *Deutsche Philologie im Aufriss*, ed. by Wolfgang Stammer, vol. 1, 1730–1989. 2nd edn. Berlin: Erich Schmidt.

- Getis, Arthur, & J. Keith Ord. 1992. The analysis of spatial association by the use of distance statistics. *Geographical Analysis* 24. 189–206.
- Getis, Arthur, & J. Keith Ord. 1996. Local spatial statistics: An overview. *Spatial analysis: Modelling in a GIS environment*, ed. by Paul Longley & Michael Batty, 261–277. Cambridge: Geoinformaiton International.
- Göschel, Joachim. 1992. *Das Forschungsinstitut für Deutsche Sprache "Deutscher Sprachatlas". Wissenschaftlicher Bericht*. Marburg: Forschungsinstitut für deutsche Sprache.
- Göschel, Joachim. 2000. Der Phonetische Atlas von Deutschland. *Лужнословенски Филолог* 56. 283–288.
- Hahn, Matthias, & Beat Siebenhaar. 2016. Sprechtempo und Reduktion im Deutschen (SpURD). *Elektronische Sprachsignalverarbeitung*, ed. by Oliver Jokisch, 198–205. Dresden: TUDpress.
- Hall, Tracy Alan. 1992. *Syllable structure and syllable-related processes in German*. Tübingen: Niemeyer.
- Hall, Tracy Alan. 2009. Liquid dissimilation in Bavarian German. *Journal of Germanic Linguistics* 21. 1–36.
- Hall, Tracy Alan. 2014. The phonology of Westphalian German glides. *Journal of Germanic Linguistics* 26. 323–360.
- Hall, Tracy Alan. 2020. Phonological processes in Germanic languages. Putnam & Page 2020, 11–32.
- Höder, Steffen. 2016. Niederdeutsch und Nordeuropa: Eine Annäherung an grammatische Arealität im Norden Europas. *Niederdeutsches Jahrbuch* 139. 103–129.
- Karch, Dieter. 1981. *Phonemdistribution dargestellt an rheinischen Ortsmundarten. Teil 1: Ergebnisse*. Marburg: Elwert.
- Kleber, Felicitas. 2017. Complementary length in vowel-consonant sequences: Acoustic and perceptual evidence for a sound change in progress in Bavarian German. *Journal of the International Phonetics Association* 50. 1–22.
- Klingler, Nicola, Silvia Moosmüller, & Hannes Scheutz. 2017. Vowel and consonant sequences in three Bavarian dialects of Austria. *Proc. Interspeech 2017*, ed. by Francisco Lacerda, 2983–2987. 10.21437/Interspeech.2017.
- Kleiner, Stefan, Ralf Knöbel, & Max Mangold. 2015. *Duden—Das Aussprachewörterbuch. 7., komplett überarbeitete und aktualisierte Auflage*. Berlin: Dudenverlag.
- König, Werner. 1997. Phonetisch-phonologische Regionalismen in der deutschen Standardsprache. Konsequenzen für den Unterricht "Deutsch als Fremdsprache"? *Varietäten des Deutschen. Regional- und Umgangssprachen*, ed. by Gerhard Stickel, 246–268. Berlin: de Gruyter.
- Kohler, Klaus J. 1995. *Einführung in die Phonetik des Deutschen. 2., neubearbeitete Auflage*. Berlin: Schmidt.

- Kraehenmann, Astrid. 2003. *Quantity and prosodic asymmetries in Alemannic. Synchronic and diachronic perspectives*. Berlin: Mouton de Gruyter.
- Lameli, Alfred. 2013. *Strukturen im Sprachraum. Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland*. Berlin: Walter de Gruyter.
- Lameli, Alfred. 2014. *Erläuterungen und Erschließungsmittel zu Georg Wenkers Schriften. Unter Mitarbeit von Johanna Heil und Constanze Wellendorf*. Hildesheim: Olms.
- Lameli, Alfred. 2021. Remarks on the geography of schwa apocope in German dialects. *Ex Oriente lux. In honorem Nicolae Saramandu*, ed. by Manuela Nevaci, Irina Floarea, & Ioan-Mircea Farcaș, 683–701. Alessandria: Edizioni dell'Orso.
- Lameli, Alfred, & Alexander Werth. 2017. Komplexität und Indexikalität. Zum funktionalen Gehalt phonotaktischer Wortstrukturen im Deutschen. *Linguistische Komplexität—ein Phantom?* ed. by Mathilde Hennig, 73–96. Tübingen: Stauffenburg.
- Lass, Roger. 1984. *Phonology. An introduction to basic concepts*. Cambridge: Cambridge University Press.
- Levickij, Viktor. 2014. Deutsche Phonemverbindungen im Anlaut des Wortstammes. *Glottology* 3. 35–42.
- Liberman, Anatoly. 1997. The reaction of monosyllables to apocope in German dialects. *Insights in German linguistics*, ed. by Werner Winter, 99–133. Berlin: Mouton de Gruyter.
- Lindgren, Kaj B. 1953. *Die Apokope des mhd. -e in seinen verschiedenen Funktionen*. Helsinki: Suomalainen Tiedeakatemia.
- Maddieson, Ian. 2009. Calculating phonological complexity. *Approaches to phonological complexity*, ed. by François Pellegrini, Egidio Marsico, Iona Chitoran, & Christophe Coupé, 85–109. Berlin: Mouton de Gruyter.
- Maddieson, Ian. 2013. Syllable structure. *The world atlas of language structures online*, ed. by Matthew Dryer & Martin Haspelmath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at <http://wals.info/chapter/12/>, accessed on September 16, 2020.
- Menzerath, Paul. 1954. *Die Architektonik des deutschen Wortschatzes*. Bonn: Dümmler.
- Moran, Patrick A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* 37. 17–23.
- Nerbonne, John. 2010. Mapping aggregate variation. *Language and space. An international handbook of linguistic variation, vol. 2: Language mapping, part I*, ed. by Alfred Lameli, Roland Kehrein, & Stefan Rabanus, 476–495. Berlin: De Gruyter Mouton.

- Nerbonne, John, & Christine Siedle. 2005. Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede. *Zeitschrift für Dialektologie und Linguistik* 72. 129–147.
- Nübling, Damaris, & Renate Schrambke. 2004. Silben- versus akzentsprachliche Züge in den germanischen Sprachen und im Alemannischen. *Alemannisch im Sprachvergleich*, ed. by Elvira Glaser, 281–320. Stuttgart: Steiner.
- Oostendorp, Marc van. 2020. Germanic syllable structure. Putnam & Page 2020, 33–48.
- Ortmann, Wolf Dieter. 1991. *Lautnachbarschaften im Deutschen*. München: Goethe-Institut.
- Parker, Stephen G. 2002. *Quantifying the sonority hierarchy*. Amherst, MA: University of Massachusetts Amherst dissertation.
- Prinz, Michael, & Richard Wiese. 1991. Die Affrikaten des Deutschen und ihre Verschriftung. *Linguistische Berichte* 133. 165–189.
- Prokić, Jelena. 2017. Quantitative diachronic dialectology. *From semantics to dialectometry. Festschrift in honor of John Nerbonne*, ed. by Martijn Wieling, Martin Kroon, Gertjan van Noord, & Gosse Bouma, 293–301. London: College Publications.
- Prokić, Jelena, Çağrı Çöltekin, & John Nerbonne. 2012. Detecting shibboleths. *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH: Visualization of linguistic patterns and uncovering language history from multilingual resources*, ed. by Miriam Butt, Sheelagh Carpendale, Gerald Penn, Jelena Prokić, & Michael Cysouw, 72–80. Stroudsburg, PA: The Association for Computational Linguistics.
- Pröll, Simon, & Stefan Kleiner. 2016. Silbengrenzen im Gebrauchsstandard. Empirie, Theorie und Typologie. *Deutsche Sprache* 44. 193–213.
- Putnam, Michael, & B. Richard Page (eds.). 2020. *The Cambridge handbook of Germanic linguistics*. Cambridge: Cambridge University Press.
- Schmidt, Jürgen Erich, Joachim Herrgen, Roland Kehrein, & Alfred Lameli (eds.). 2008–2021. *Regionalsprache.de (REDE): Forschungsplattform zu den modernen Regionalsprachen des Deutschen*. Prepared by Bock Dennis, Robert Engsterhold, Brigitte Ganswindt, Heiko Girnth, Simon Kasper, Roland Kehrein, Alfred Lameli, Slawomir Messner, Christoph Purschke, & Anna Wolańska. Marburg: Forschungszentrum Deutscher Sprachatlas. Available at <https://regionalsprache.de/>, accessed on May 10, 2021.
- Seiler, Guido, & Kathrin Würth. 2014. Monosyllabic lengthening in German and its relation to the syllable vs. word language typology. *Syllable and word languages*, ed. by J. Caro Reina & Renata Szczepaniak, 140–159. Berlin: de Gruyter.
- Seiler, Hansjakob. 1962. Laut und Sinn: Zur Struktur der deutschen Einsilbler. *Lingua* 11. 375–387.

- Selkirk, Elisabeth. 1984. On the major class features and syllable theory. *Language sound structure: Studies in phonology. Presented to Morris Halle by his teacher and students*, ed. by Mark Aronoff, Richard T. Oehrle, Frances Kelley, & Bonnie Wilker Stephens, 107–113. Cambridge, MA: MIT Press.
- Siebenhaar, Beat, & Matthias Hahn. 2019. Vowel space, speech rate and language space. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia*, ed. by Sasha Calhoun, Paola Escudero, Marija Tabain, & Paul Warren, 879–883. Canberra: Australasian Speech Science and Technology Association Inc.
- Sievers, Eduard. 1881. *Grundzüge der Phonetik. Zur Einführung in das Studium der Lautlehre der indogermanischen Sprachen*. Leipzig: Breitkopf & Härtel.
- Simmler, Franz. 1983. Konsonantenschwächung in den deutschen Dialekten. *Dialektologie. Ein Handbuch zur deutschen und allgemeinen Dialektforschung*, vol. 2, ed. by Werner Besch, Ulrich Knoop, Wolfgang Putschke, & Herbert Ernst Wiegand, 1121–1129. Berlin: De Gruyter.
- Spedicato, Giorgio Alfredo, Kang Tae Seung, Yalamanchi Sai Bhargav, Thoralf Mildenerger, Yadav Deepak, Córdón Ignacio, Jain Vandit, & Giorgino Toni. 2020. *markovchain. R package version 0.8.4*.
- Spiekermann, Helmut. 2000. *Silbenschnitt in deutschen Dialekten*. Tübingen: Niemeyer.
- Stolz, Thomas, Nicole Nau, & Cornelia Stroh. 2012. *Monosyllables. From phonology to typology*. Berlin: Akademie Verlag.
- Tjong-Kim-Sang, Erik Fajoen. 1998. *Machine learning of phonotactics*. Groningen: s.n.
- Trubetzkoy, Nikolaj S. 1939. *Grundzüge der Phonologie*. Göttingen: Vandenhoeck & Ruprecht.
- Vennemann, Theo. 1988. *Preference laws for syllable structures and the explanation of sound change. With special reference to German, Germanic, Italian, and Latin*. Berlin: Mouton de Gruyter.
- Wenker, Georg. 2013. *Schriften zum "Sprachatlas des Deutschen Reichs". Gesamtausgabe*, ed. by Alfred Lameli, Mitarbeit von Johanna Heil, & Constanze Wellendorf. Hildesheim: Olms.
- Werner, Otmar. 1978. Schwa-Schwund und Phonotaktik im Deutschen. *Studia linguistica Alexandro Vasilii filio Issatschenko a collegis amicisque oblate*, ed. by Henrik Birnbaum, Lubomir Durovic, Gunnar Jacob, Nils Ake Nilsson, Anders Sjoeborg, Dean S. Worth, 127–147. Lisse: Peter de Ridder.
- Wiese, Richard. 1988. *Silbische und lexikalische Phonologie*. Tübingen: Niemeyer.
- Wiese, Richard. 2000. *The phonology of German*. Oxford: Oxford University Press.
- Wiesinger, Peter. 1983. Die Einteilung der deutschen Dialekte. *Dialektologie. Ein internationales Handbuch*, ed. by Werner Besch, Ulrich Knoop,

Wolfgang Putschke, & Herbert Ernst Wiegand, 807–900. Berlin: Walther de Gruyter.

Research Center Deutscher Sprachatlas  
Pilgrimstein 16  
35032 Marburg  
Germany  
[lameli@uni-marburg.de]