BA Thesis

Faculty of the Humanities

BA Linguistics

Testing the OT-writing grammar:

Swiss German geminates and their orthography

UNIVERSITEIT VAN AMSTERDAM

Leda Bonzanini

13375563

26/06/2023

Supervisor: Dr. Silke Hamann

Abstract

This thesis illustrates the application of the *Reading Grammar* (Hamann & Colombo 2017), in the writing direction, to the orthography of geminates in the Swiss dialect of Bern-German and proposes an extension of this model. First, it was empirically tested whether double graphemes correspond to phonological geminates in this dialect, by analyzing the words containing consecutive identical graphemes in Reddit posts flagged with 'Bern' and then the generalizations made on the empirical results were employed to formalize the phoneme-grapheme conversion of geminates using BiPhon-Optimality Theory (Boersma 2011). In general, most double graphemes corresponded to phonological geminates, however lexical exceptions and exceptions due to the optional phonological process of /l/ vocalization were found. The *Reading Grammar* was then applied to these patterns, and it was found insufficient to deal with the exceptions, as it maps only the surface form onto the orthographic form. Therefore, an extra mapping of the underlying form onto the orthographic form had to be introduced to properly formalize these patterns. This approach works with this data set and reconciles polar views on the topic while also encompassing some psycholinguistic aspects of visual word recognition.

Table of Contents

1 Introduction and Literature Review 3
1.1 Language Situation in Switzerland 3
1.2 Phonology 3
1.3 German geminates 7
1.4 Orthographic representation of geminates 7
1.5 Grapheme-Phoneme Conversion in Optimality Theory 8
1.6. Phonological Considerations for the Analysis 10
1.7 Research Question and Hypotheses 11
2 Methodology 11
3 Results 12
3.1 Results per Category 13
3.1.1 Interim Discussion Categories 14
3.2 /l/ vocalization 15
3.2.1 Interim Discussion /lː/ vocalization 16
3.3 Overall discussion 16
4 Constraints and Analysis 16
4.1 Formalization of /lː/ vocalization in orthography 19
4.2 Lexical exceptions 25
5 Discussion 29
5.1 Psycholinguistic Considerations 31
5.2 Surface form vs underlying form mapping 31
6 Conclusion 34
Bibliography 35
Appendix 38

1 Introduction and Literature Review

1.1 Language Situation in Switzerland

Switzerland is a multilingual country, where German, French, Italian, and Romansh are spoken officially. However, the term 'German' is inadequate to describe the linguistic situation of the areas where it is spoken. Standard German is the official language, used in all formal written communication and educational environments. The spoken 'German' is known as 'Swiss-German,' which is an umbrella term for several different Alemannic dialects. Although these dialects are mutually intelligible, they differ from each other. In monolingual households, children primarily speak Swiss-German, and only begin acquiring Standard German in school [1]. It follows that Swiss-German native speakers' German is significantly influenced by the dialect, giving rise to the definition of 'Standard Swiss German,' the local version of German.

1.2 Phonology

One intriguing feature of most Swiss-German dialects is that, unlike German, they have geminates (Marti 1985), creating an asymmetry between the phonology of the official language and the spoken one. Some studies suggest that standard Swiss-German also has geminates, resulting from the influence of dialects on the standard language (Christen 2001; Kolly 2013). Since Standard Swiss-German is the official language of education and teachers themselves are not native speakers of Standard German, they tend to maintain the gemination of intervocalic consonants (Marti 1985). It follows the hypothesis that children acquiring Standard German are exposed to geminates in both the standard and non-standard language.

Swiss-German consonants are primarily distinguished by length, with a singleton vs geminate contrast (Marti 1985; Kraehenmann & Lahiri 2008). The dialect of Bern appears to be more conservative, e.g. in comparison with Basel-German concerning geminate consonants (Naiditsch 1997: 263). Indeed, this dialect did not undergo the degemination process that took place in the shift from Middle High German to New High German (Auer 1999). The following analysis will take into consideration only this phonological feature of the Swiss-German dialect of Bern, a South Alemannic dialect.

[1] https://www.eda.admin.ch/aboutswitzerland/en/home/gesellschaft/sprachen/sprachen-und-dialekte.html (last accessed on Monday, June 26th 2023).

This dialect has a lax/tense and short/long vowel distinction, the table below (1) describes the vowel system of Bern-German. The table divides the vowels into short and long and orders them from the most lax to the most tense, the table is taken and adapted from Schlote's dissertation (2008: 16).

Table 1. *Vowel system of Bern-German (Schlote 2008:16).*

Short vowels	Long vowels
(5) i ü u	(5) i: ü: u:
(4)	(4)
(3) e ö o	(3) e: ö: o:
(2) ɛ œ ɔ	(2) ɛ: œ: ɔ:
(1) æ a	(1) æ: a:

Geminate fricatives are found after both long and short vowels and are therefore independent of vowel length. This is exemplified in the minimal pairs in examples (1a) and (1b).

Throughout the paper, the [ ] will be used to indicate the phonetic realizations, the // to indicate the surface phonological categories, and the <> to indicate the orthographic forms, while the translations will be given in '', the underlying phonological forms will be given in ||. Geminates that cross word boundaries will be represented as two separate singleton sounds divided by the syllable boundary symbol ' .'. Concerning examples, the orthographic form of words will be given in italics.

(1a) Minimal pair: short vowel with long consonant

pfiffe /pfi:f:ɛ/ 'Whistle'

(1b) Minimal pair: long vowel with a long consonant

pfiffe /pfɪf:ɛ/ 'whistled'

(Schlote 2008: 58)

The length of the fricative singleton segments after short vowels can be observed in the first oscillograph in Figure (1) below. The second and third oscillographs show the length of a geminate fricative segment first after a long vowel and then after a short vowel.

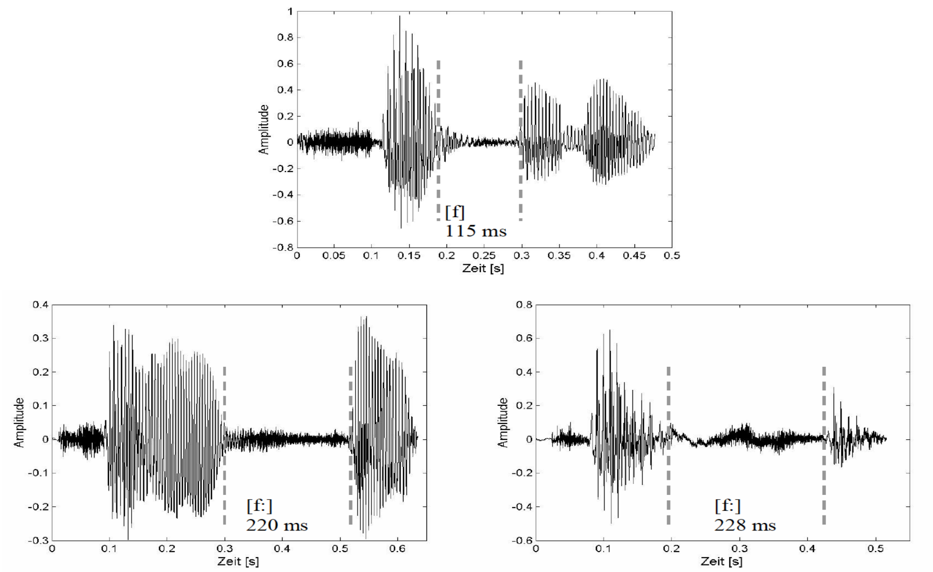


Figure 1. *Oscillographs of singleton fricative after short vowel (top), and geminate fricatives after long (bottom left) and short vowels (bottom right). (Schlote 2008: 105)*

Geminate plosives, on the other hand, are found only after lax short vowels, with middle-length plosives (which are still significantly longer than short plosives) appearing after diphthongs and long vowels as well. Ham (2001: 43) explains this three-way contrast as a consequence of the development of South Alemanic consonants from West Germanic, as shown in Table (2) below.

Table 2. *Historical development of three-way plosive contrast. (Schlote 2008: 32)*

West Germanic	Old Alemanic	South Alemanic
β, ð, ɣ	b, d, g	b̥, d̥, g̊
b, d, g	p, t, k	p, t, k
bː, dː, gː	bː, dː, gː	pː, tː, kː

As shown in this table (2), it is possible to trace back the origin of the plosive three-way contrast (shown in figure 2): the West Germanic fricatives were strengthened in Old Alemannic and then devoiced in South Alemannic, which now corresponds to the singleton plosives in Bern-German. The West Germanic b,d,g were devoiced in Old Alemannic and kept in South Alemannic and correspond to the Bern-German middle-length plosives. The West Germanic geminate voiced plosives were devoiced when going from Old Alemannic to South Alemannic and correspond to the current voiceless geminate plosives. This three-way contrast is shown in the oscillographs below, in Figure (2). The upper oscillograph shows the short, voiceless plosive, that corresponds to the Old Alemanic /d/, the lower one on the left shows the mid-long voiceless plosive, historically a /t/, and the geminate voiceless plosive /t:/ is represented in the lower oscillograph on the right.

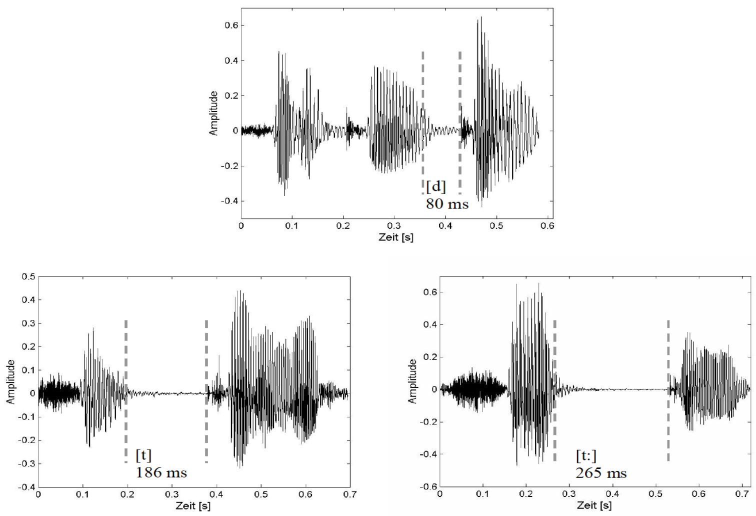


Figure 2. *Oscillographs of singleton vs geminate vs middle-length plosives (Schlote 2008: 98).*

Geminates for nasals and liquids are only found after short lax vowels (Schlote 2008: 106). All geminates can be found in both word- and phrase-medial positions, as well as in word- and phrase-final positions. Only non-lexical geminates created by Sandhi processes can be found in Bern-German (Schlote 2008: 109), therefore, they will be avoided in this analysis. An example of this would be /s:i:b/ as a short form of *'das Sieb' 'the strainer'*. There is however no clear prediction on whether people would merge the contracted form of an article (*'das'*-> *'s'*) and a noun (*'Sieb'*) in writing to match the pronunciation, getting e.g. the orthographic form *'ssieb'*.

1.3 German geminates

Contrarily to Swiss-German, in German there is no singleton/geminate contrast in the phonology, however, double graphemes can be found within phonological words and within phonological words at word boundaries due to compounds or morphological inflection. The phonological realization of these geminates is then avoided by a degemination process (Wiese 2000). Degemination is optional at word boundaries, but mostly present in fast speech and thus considered to be phonetic, not phonological (Hamann 2020). Orthographic consecutive identical consonants are also employed in German to signalize the shortness and laxness of the preceding vowel (a.o. Neef 1992). Example (2) below illustrates these phenomena.

(2a) Fake geminates

Brottaig	bʁoːt + taɪg	[bʁoːtːaɪk]~[bʁoːtˑaɪk]~[bʁoːtaɪk]	'bread dough'

(Hamann 2020: 1)

(2b) Inflectional double grapheme

hatte	hat + tə have 3SG.PST	[hatə]	'he/she/they had'

(Hamann 2020: 1)

In examples (2a) and (2b) the transcription of the precise pronunciation is given in [], to show that in standard German, the gemination is purely phonetic in compound words, but not in contexts of graphic double graphemes caused by inflection.

1.4 Orthographic representation of geminates

The official spoken language in the germanic-speaking Switzerland is often called 'Schweizerhochdeutsch' (standard Swiss German) and the official written language called 'Schriftdeutsch' (the same as standard German, with only a few lexical exceptions). There are also studies indicating that there are geminates in standard Swiss German (the language spoken in educational environments), as a result of the influence of dialects on the standard language (teachers themselves are not native speakers of standard German, and therefore tend to keep the gemination of intervocalic consonants, (e.g. Christen 2001; Kolly 2013)This means that Swiss-German-speaking children acquiring German are exposed to an orthographic way of signaling the length of a character by doubling it, creating two graphemes to represent one phoneme since they see a double consonant and hear a geminate.

Throughout this paper, I will refer to orthographic geminates as 'identical consecutive consonants' or 'double consonants', while for phonological geminates, the term 'geminate' will be employed. Recently, Swiss-German speakers, primarily due to social media's development and spread, have begun to write mostly in Swiss-German for non-official communications (Anschwanden 2001; Siebenhaar 2003). Because of their uncodified status, these dialects are mostly spelled phonetically by relying on the literacy system in place (Aschwanden 2001). Therefore, speakers use the phoneme-grapheme conversion known to them and adapt it to their dialects' phonological needs. This could thus lead to the logical conclusion that if a speaker of Bern-German hears a geminate, they will write it as a double consonant when writing their dialect.

Throughout this paper, it will be assumed that, in general, geminates are phonologically consisting only of a single unit that is mapped in writing onto two graphemes.

1.5 Grapheme-Phoneme Conversion in Optimality Theory

Hamann and Colombo (2017) came up with a model that could account for the grapheme-phoneme conversion of English borrowings containing geminates into Italian. This model, the 'Reading Grammar' can be considered as an extension of the bidirectional phonetics and phonology model (henceforth: BiPhon) (Boersma 2011). The latter is an Optimality Theory based model that accounts for both perception and production, in a single, elegant solution, by simply changing the direction of its steps (see Figure 3). Going top-down represents comprehension, and bottom-up production, for this reason, this will later be referred to as the 'bidirectionality principle' of this model. All constraints are evaluated in parallel, and their rankings are the same independently from the direction.

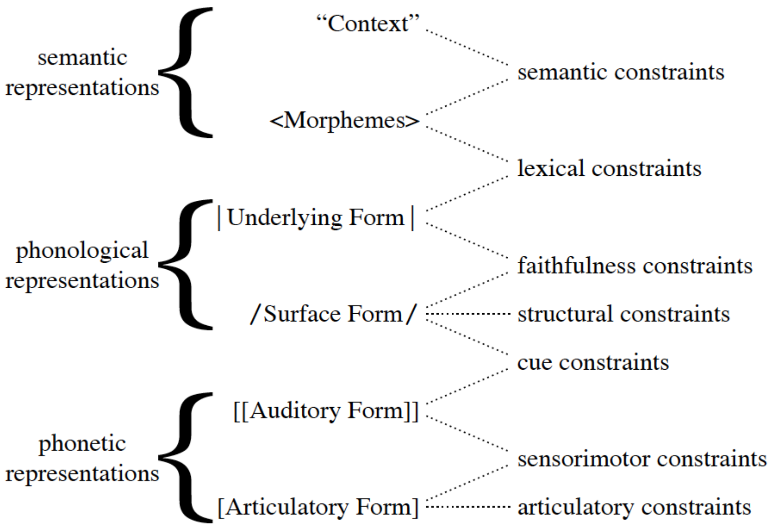


Figure 3. *BiPhon model (Boersma 2011).*

The ‘Reading Grammar’ can be ‘attached’ to the surface form in the BiPhon model and describe the phoneme-grapheme conversion (see Figure 4).

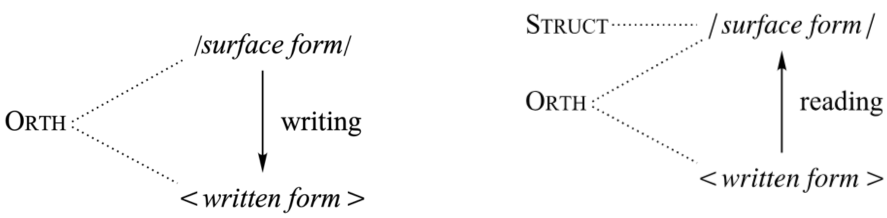


Figure 4. *Reading Grammar in the reading direction (right) and writing direction (left).* (Hamann & Colombo 2017)

Since this is an extension of the BiPhon model, it also rests on the bidirectionality principle, however, it has so far only been employed to model reading, not writing.

1.6. Phonological considerations for the analysis

Bern-German has, as explained above, a three-way phonetic contrast in plosives, however phonologically, I follow the two-way division by Marti (1985) into fortis consonants (middle length and geminate plosives) and lenis devoiced, short plosives.

Furthermore, Bern-German has assimilation processes that take place across word boundaries (Schlote 2008:187) due to the Maximal Onset rule, which causes the CV syllable structure to be carried out also across word boundaries. This causes so-called ‘fake geminates’, that in German occur in writing and in fast speech (Wiese 2000: 231), but always in production in Bern-German (Schlote 2008: 187). Concerning words with clitics, the orthographic form coincides with the phonological word.

(3) Clitics and intervocalic geminates

nenni /nɛn.ni/ ‘I call’
call-I.NOM

(4) Fake geminates in Bern-German

strommagie /stroːm.magiː/ [stroːmːagiː] ‘energy magic’

Example (3) illustrates how clitics are part of the orthographic and phonological word, and example (4) represents fake geminates in Bern-German. The example is taken from this paper’s data set and shows that, while in German these orthographic geminates do not correspond to phonological ones, they do in Bern-German (cf. example 2a). For this reason, compound words with a double grapheme at the word boundary will also be considered in this analysis.

Concerning /l/ vocalization, /l/ is usually realized as [l] in initial position and vocalized in disyllabic words in medial position (i.e. pronounced as [u]), where underlying a /lː/ would be found, as in example (5) (a.o., Marti 1985). However, this process is highly speaker dependent and typical of the more western part of the Bern-German-speaking area.

(5) /l/ vocalization

alles /auɛs/ ‘everything’

1.7 Research Question and Hypotheses

The aim of this study is thus twofold. First, it is of interest to analyze whether consecutive identical graphemes can be found in the writing system of Bern-German, in the same place where there would be a geminate in the phonology. Additionally, the aim is to model this behavior using the *Reading Grammar* (Hamann & Colombo 2017) in the opposite direction (the ‘writing’ grammar), following the bidirectionality principle of the ‘BiPhon’ model (Boersma 2011).

It is predicted that there will be double consonants in writing and that they will always correspond to a phonological geminate, i.e. there will not be instances where a singleton will be mapped onto two graphemes. If the speakers have a surface form with a geminate, they will map the long phoneme onto two graphemes.

Since no phrase initial geminates are present in this dialect, no initial double consonants are expected. Due to the above-mentioned reasons, all orthographic words (meaningful groups of characters written without white spaces) containing a final geminate plus a clitic, are also expected to be written with a double consonant. Concerning the vocalization of the lateral geminate /lː/ to /u/, no concrete prediction can be done, since the phenomenon is very much speaker dependent and micro-regionally variable, and I do not have access to the precise area of origin of the posts-writers.

2 Methodology

In order to answer the first question, a Reddit API was employed. Reddit API (Application Programming Interface) is a set of tools and protocols that allows developers to programmatically interact with Reddit's platform and retrieve information such as posts, comments, and user data.

Using a Python script (Hannah Pippin 2023) for Reddit API, 2000 posts from the subreddit ‘BUENZLI’ were retrieved, together with the regional origin of the user, as signaled by a ‘flair’, a sort of hashtag with the name of the Swiss canton they were from. This flair is applied by the Reddit user himself and is publicly accessible, as were all the posts retrieved.

The data was then sorted employing a Python program (Van Rossum & Drake 2009) written using the IDE PyCharm (JetBrains 2017) and applying the packages ‘csv’, ‘re’, ‘pandas’, and ‘praw’, which found all instances of words containing consecutive identical consonants. The program also assigned the double consonants to their belonging category: plosive, fricative, trill, lateral, or nasal. Then, from this list, it individuated all instances of English words using the United Kingdom English 30.000 words corpus from the Leipzig Corpora Collection (Leipzig Corpora Collection 2018). The English words were taken out, as not subjects of the analysis.

Due to the lack of reliable, accessible sources concerning the pronunciation of Bern-German words, where in doubt, a native speaker of Bern-German was consulted and asked to pronounce words whose pronunciation I was not sure of. In ambiguous cases, where, even with native knowledge of the singleton/geminate contrast, I was not able to distinguish the contrast with certainty, the words were then analyzed in Praat (Boersma & Weenick 2023).

The orthographic words containing double consonants were then compared to their pronunciation, either as a singleton or as a geminate.

Only unique words (i.e. only one instance per word was considered in cases where multiple were given in the dataset) that had a double consonant in writing were considered for this analysis. This allowed to exclude any effect of frequency in the statistical analysis. The final word list was then manually modified to delete any nonsensical words, non-word sounds such as ‘pssst’, and words where ‘fake double consonants’ were found: e.g. where <ss> corresponds to[sʃ] due to compounding (e.g. in *‘Amtsschimu’* ~ [amtsʃimu] ~*‘bureaucracy’*). Furthermore, if instances of a word were found in compound words with the same syllable structure, only one instance was kept (e.g. ‘millione’ and ‘million each’). Words containing more than one double consonant grapheme were counted twice, one time per segment: e.g. the word ‘aggressiv’ was analyzed both for <gg> and <ss>.

3 Results

Overall, 529 words containing double consonants were analyzed. The table below (3) shows an overview of how many instances of double consonant segments were found per consonantal category:

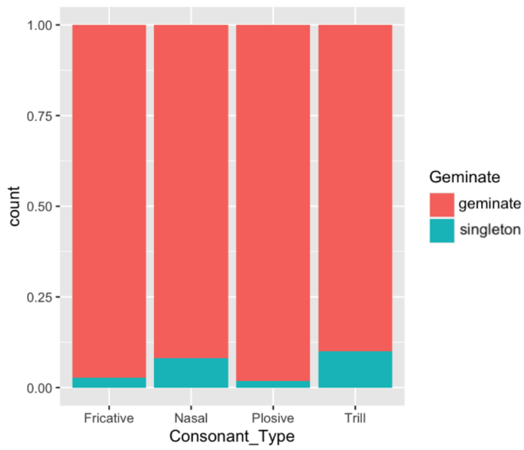
Table 3. *Overall number of words analyzed.*

Plosive	Fricative	Trill	Lateral	Nasals
173	142	30	73	111

Two words were then deleted, as they were diminutives of personal nouns which are not of interest in this analysis.

3.1 Results per category

The graph below shows the proportions of geminates vs singletons corresponding to double consonants in all categories except for laterals since a separate analysis thereof is given in section 3.3. Overall, only 20 instances of singletons out of 529 (~3.8%) corresponding to double graphemes were found.



Graph 1. *Proportions of geminates vs singleton corresponding to double consonants per consonant category.*

The analysis was carried out with Rstudio (Rstudio Team 2020), and the same program was used to code the plotted graphs in the language R, using the packages 'tidyverse' and 'ggplot'. It was found that 97.7% of double plosive consonants correspond to geminate plosives, 97.2% of fricatives, 91.9% of nasals, and 90% of trills (see Table 4).

Table 4. *Proportions of geminates vs singleton corresponding to double graphemes per consonant category.*

	Plosive	Fricative	Nasals	Trill
Geminate	169	138	102	27
Singleton	4 (~2.3%)	4 (~2.8%)	9 (~8.1%)	3 (10%)

An overview of words containing singletons written as a double grapheme is given in the appendix.

3.1.1 Interim Discussion Categories

Across all categories, the correspondence of a double grapheme to a phonological geminate was found to be very systematic, which supports the hypothesis that the length of the grapheme also corresponds to the length of the consonant, when writing Bern-German. The main mismatches between orthography and pronunciation were found in nasals and trills. The latter do not represent a large enough pool of data to draw any conclusions and concerning the nasals, it may be that the original spelling of the French or German words was kept. Concerning the two words containing clitics: *'ufggä'* and *'mussme'* it can be hypothesized that the spelling is normalized to keep it the same as the rest of the paradigm (the stem *'muess'* or *'muss'* is the same throughout the paradigm and 'ggä' is the most common way of spelling the verb 'to go'), creating double consonants that correspond to surface singletons.

Based on these results, it seems likely that the Latin-based loanwords' pronunciation (probably loaned from French) is loaned together with the spelling. Another option would be that if the speaker is conservative enough in their speech and pronounce the word with the original tense instead of the lax vowel (Schlote 2008: 123), it follows naturally that no geminate can be produced. This is however a purely speculative statement and further research is needed to prove this. Furthermore, being a non-native speaker of the language and having consulted only one native speaker, the results may vary in a production study aimed to investigate this. Only one case of a word beginning with two identical consonant graphemes was found: *'ddrunge'* ('drunk'), however, that did not correspond to a geminate in speech.

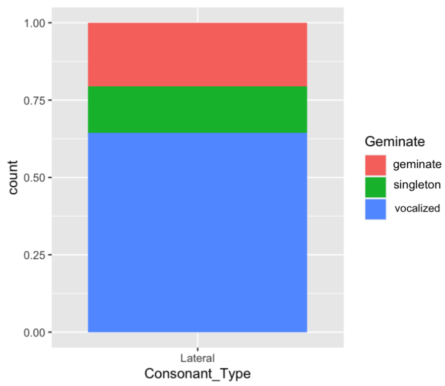
3.2 /l/ vocalization

Concerning the /l/ vocalization process, a lot of instances were found of <ll> being realized as /u/. An overview of the numbers is given below (see Table 5).

Table 5. */l/ realizations.*

|l:| /l:/	|l:| /l/	|l:| /u/
15 (~21%)	11 (~15%)	47 (~64%)

The table summarizes the number of occurrences of the different realizations of the lateral. The grapheme <ll> is often utilized also to express the vocalized /l/, but being the /l/ vocalization phenomenon very micro-regionally and speaker-dependent, there might be a speaker preference towards using the grapheme <ll> or <u>, as that has also been found in writing in the collected data e.g. *'appenzeu'* ('Appenzell', a region in Switzerland), it was however not subject to any analysis since not of initial interest. The grapheme <ll> thus can correspond to both a /l:/, following short lax vowels or the vocalized segment realized as [u] across syllables and in final position (underlying a lengthened |l|). There were very few instances of a singleton written as <ll>.



Graph 2. /l:/ *realizations.*

3.2.1 Interim Discussion /lː/ vocalization

Overall, the grapheme <ll> mostly corresponds (in 85% of the cases), as predicted to either vocalized ambisyllabic geminate, vocalized lengthened final /l/, or a non-vocalized geminate following a short lax vowel. This seems to be systematic, even though less so than in the other categories.

All the words containing the grapheme <ll> that corresponded to a singleton /l/ are Latin-based loanwords (*'allergisch', 'grilliere', 'homosexuelli', 'illustriert', 'installiert', 'intelligänti', 'kolläg', 'kollaterau', 'kontrollierä', 'million', 'millionäfach'*). As for the other categories, it could be hypothesized that both the spelling and the pronunciation are loaned from the original language, however, this is to be determined by further research.

3.3 Overall discussion

Overall, the conclusion can be drawn that, unlike in other Germanic languages such as English or German, Bern-German does employ double consonants to represent a geminate consonant (at least in the words here analyzed, represented with a maximum of two graphemes). Not all consonants are subject to this orthographic doubling: indeed, no doubling of complex graphemes was found in the data. For instance, the coronal fricative /ʃ/ is typically written as <sch>, and as hypothesized by Wiese (2004), reading a double grapheme consisting of 6 characters such as <schsch>, would be way more effortful, than simply keeping the basic, single grapheme.

4 Constraints and analysis

Hamann & Colombo (2017: 8) developed the *Reading Grammar*, a model to formalize grapheme-phoneme conversion (as explained in the introduction section). Thereby, they introduced the so-called *orthographic constraints*; they formalize the interaction between the phonological Surface Form and the Orthographic Form and vice versa. The following example (6) illustrates a general orthographic constraint.

(6) *<γ>/ /:

Assign a violation mark to every grapheme <γ> that is mapped onto an empty segment in the SF.

(Hamann & Colombo 2017: 8)

Orthographic constraints are based on what seem to be generalizations about preferences in orthography, as in Wiese (2004); this constraint (6) posits that each phoneme should be mapped onto a grapheme and that it should not be represented by an empty grapheme and reflects Wiese's principle (2004) of one sound-one letter for shallow orthographies (Hamann & Colombo 2017:8).

Another relevant, yet absent from the literature constraint, seems to be the following:

(7) <γi>/Pi/:

Assign a violation mark to every grapheme <γ> that is not mapped onto the phonological form /P/ and vice versa as in the conventionally established letter-sound associations (e.g. /a/ <a>).

This constraint (7) is a more specific version of the <γ>/P/: constraint by Hamann & Colombo (2017), which assigns a violation mark to every grapheme <γ> that is not mapped onto the phonological form /P/ and vice versa. This (as all other kinds of constraints, violable) orthographic constraint, together with constraint (6) describes the orthographic principle of necessarily mapping a grapheme onto a phoneme, as claimed by Wiese (2004). In this paper, the assumption is that each writer develops their own set of constraints, based on the association of a phonological unit with the encountered written forms, very similar to what is assumed for the reading process in Hamann & Colombo (2017). It then seems to be appropriate to adjust constraint (5) to be more specific, mapping a surface form onto a specific grapheme, which in that language is commonly employed to represent a certain sound or set of sounds. An example would be mapping the SF /m/ onto the orthographic form <m> in Bern-German. This of course needs to be language specific, since the established (whether by the speakers or by the standard spelling rules) phoneme-grapheme correspondences can vary quite a lot across languages.

In order to represent the mapping of geminates onto a double grapheme constraint (8) is needed. This allows to formalize the grapheme-phoneme conversion of languages with geminates that usually represent them with double identical consonants in writing.

(8) <βiβi>/C:/:

Assign a violation mark if a grapheme of two identical consonantal letters is not mapped onto a surface geminate, and vice versa. (Hamann & Colombo 2017: 9)

This constraint (8) was employed to model the reading direction of English borrowings containing double consonants in Italian. Italian speakers indeed pronounce these words with a geminate, even though it is not present in the English phonology. This means that there must be a constraint that keeps the orthographic geminate also in the phonology. This constraint is fundamental to modeling the writing of geminates in Bern-German.

Relying on BiPhon's bidirectionality principle, these constraints can also be used to model the writing direction. In this case, the constraints apply when going from the phonological surface form (henceforth: SF) to the orthographic form (henceforth: OF). Tableau (1) shows the workings of the three above-mentioned constraints (n. 6, 7, 8).

/pfif:ɛ/	<βiβi> /C:/	<γi>/Pi/	*<γ>/ /
☞ <pfiffe>			
<pfife>	*!		
<pfitte>		*!	
<pfie>			*!

Tableau 1. *Orthographic constraints in Bern-German.*

The chosen candidates represent the two possible ways of graphically representing the word '*pfiffe'* ('to whistle') concerning the geminates, and two candidates that show the workings of the above-presented general orthographic constraints. The first one is the winning candidate, as it does not violate any of the constraints. The second candidate violates the <βiβi>

/C:/, while the third one violates fatally constraint number (7). The rankings of the constraints one over the other does not matter in this case, as the second, third, and fourth candidate are thrown out either way.

4.1 Formalization of /lː/ vocalization in orthography

Concerning the phenomenon of /l/ vocalization, first, the following Tableau (2) illustrates the optional phonological process of /l/ vocalization independently of the length of the segment. For this formalization, faithfulness constraints are needed to keep the underlying form identical to the output, which I here summarize as the constraint 'FAITH- |l| /l/' (see constrain 10). Furthermore, there needs to be a constraint that describes the optional /l/ vocalization and for that, I introduce the constraint '/l/ vocalization'. This constraint generally describes the phonological process, independently of the length of the underlying form (since both the geminate and singleton lateral can be vocalized).

(9) /l/ vocalization

Assign a violation mark to every candidate that does not vocalize /l/.

(10) FAITH- |β| /β/

Assign a violation mark to any surface form that is not faithful to the underlying form.

|l|	FAITH-|l| /l/	/l/ vocalization
☞ /l/		*
☞ /u/	*	

Tableau 2. *Phonological process of /l/ vocalization in Bern-German.*

Tableau (2) shows that the phonologically underlying |l| can have two surface forms: /l/ and /u/. Then, the ranking of the faithfulness constraint over the /l/ vocalization constraint is not relevant, since also in the reversed ranking, one of the two is violated by one candidate. The two candidates are both acceptable outputs in Bern-German.

Following Marti (1985), the /l/ vocalization process is not only very variable micro-regionally and speaker-internally, but it also has a stylistic marking, where the vocalized form is more informal than the non-vocalized one. Based on the above-reported data (section 3), it is clear that at least the consulted native speaker can recognize the grapheme <ll> and read it as <u> while keeping the /l/ realization in other cases. There also seems to be variation in writing (although from the gathered data, it is not possible to establish whether it is inter- or intra-speaker), since sometimes the vocalized /l/ is written as <u>. No case was found in this data set of an <u> read as /lː/, which suggests, that if there is no vocalization in the SF, the phoneme will be written as <ll>.

(11) Writings of vocalized /lː/

aktuell	/aktuɛu/	'current'
appenzeu	/apːɛn͡tsɛu/	'Appenzell'
söll ought.3SG/1SG	/sölː/	'(he/she/it/I) ought'

Tableau (3) below represents how the /l/ vocalization process can be formalized in the writing direction. Here, new constraints are introduced: first, the */l/ <u>, which ensures that the phonological /l/ is not written as <u>. Indeed, the speaker would not have any knowledge of the /u/ if they did not vocalize the /l/ in the first place, and they would consequently be unable to access the grapheme <u>. Furthermore, it is unlikely that a non-vocalizing speaker would employ the grapheme corresponding to the vocalized form. However, since the underlying form of /u/ is |l|, the speaker is able in writing to retrieve the necessary information to write the grapheme <l>. This constraint is there to handle the variation in writing between <l> and <u> for the same SF /u/. [2] The following Tableau (3) shows the working of the /l/ vocalization constraint and the <βiβi> /C:/ constraint.

[2] If one assumes Marti's (1985) position about /l/ vocalization as being an optional and rather informal feature, one could hypothesize that a contextual constraint, a higher level, social constraint such as 'INFORMAL' could be responsible for the choice of /u/ over /l/ and consequently also <u> over <l>.

<table>
<tr><td>|lː|</td><td>/l/ vocalization</td><td><γi>/Pi/</td></tr>
<tr><td>☹ /u/ <ll></td><td></td><td>*!</td></tr>
<tr><td>☞ /u/ <u></td><td></td><td></td></tr>
</table>

Tableau 3. *Interaction of /l/ vocalization constraint and graphic constraints in Bern-German.*

In Tableau (3), the interaction between the structural constraint for /l/ vocalization and the orthographic constraint leads to the elimination of the first candidate, which however is an acceptable pair of SF-OF in Bern-German.

To formalize the interaction between faithfulness and orthographic constraints, a general faithfulness constraint was introduced (10), that keeps the surface form close to the underlying form.

<table>
<tr><td>|lː|</td><td>FAITH- |lː| /lː/</td><td><βiβi> /Cː/</td></tr>
<tr><td>☞ /lː/ <ll></td><td></td><td></td></tr>
<tr><td>/lː/ <u></td><td></td><td>*!</td></tr>
</table>

Tableau 4. *Interaction between faithfulness constraints and orthographic constraints.*

As it can be seen in Tableau (4), the interaction between the phonological constraints and the graphic constraints leads to the winning of only the first output candidate. Indeed, the FAITH- /l/-|l| outranks the orthographic constraint, which is then fatally violated by the second candidate. The following Tableau (5) reconciles the phonological constraint of /l/ vocalization and the graphic constraints needed to represent this phenomenon in a situation, where the speaker vocalizes the /l/, i.e. with the SF being /u/.

|lː|	FAITH- |lː| /lː/	/l/ vocalization	<βiβi> /Cː/	<γi>/Pi/
☞ /lː/ <ll>		*		
/lː/ <u>		*	*!	*
☹ /u/ <ll>	*			*!
☞ /u/ <u>	*			

Tableau 5. *Interaction between orthographic, faithfulness, and structural constraints for the orthography /l/ (non-)vocalization.*

As shown in Tableau (2), the two phonological constraints cannot be ranked one above the other, since either way, at least one of the candidates is going to be fatally violated. The second candidate fatally violates the <βiβi> /Cː/ constraint, which is ranked higher than the <γi>/Pi/ constraint to allow for the second candidate to be thrown out first. However, the <γi>/Pi/ constraint is fatally violated by the third candidate, which is, in reality, an acceptable output. Changing the ranking of these two constraints always leads to the third candidate being thrown out, even though acceptable. So, a writer could choose to write the underlying |l|, surface form /l/, with the OF <l>, because they are rather non-vocalizers or <u> if they are vocalizers in an informal context. Tableau (5) thus shows that mapping the SF onto an OF does lead to the winning of all the acceptable outputs. At this point, it appears necessary to make use of the underlying form in the mapping, in order to account for this issue. The following Tableau (6) shows how it is possible to formalize the /l/ vocalization mapping onto orthography. The person writing then maps the UF onto one of the possible surface forms, and will finally represent the SF with one of the acceptable OFs for that SF, depending on the contextual needs. At the same time, the speaker can employ the 'direct' way that directly maps the UF onto the OF, as shown in Figure (4) below For this reason, the <βiβi> |Cː| constraint needs to be introduced. It works exactly like the <βiβi>/Cː/ constraint however, instead of the SF, it employs the underlying form.

(12) <βiβi> |C:|

Assign a violation mark if a grapheme of two identical consonantal letters is not mapped onto an underlying geminate, and vice versa.

Tableau (6) below shows the working of this constraint and its interaction with the orthographic constraints.

|al:ɛs|	FAITH- |l:| /l:/	/l/ vocalization	<βiβi> |C:|	<γi>/Pi/	<βiβi> /C:/
☞/auɛs/ <alles>	*			*	
☞/auɛs/ <aues>	*		*		
/al:ɛs/ <aues>		*	* (!)	*(!)	*
☞/al:ɛs/ <alles>		*		*	

Tableau 6. *Interaction of the graphic constraints for /l/ vocalization and the <βiβi> |C:| constraint.*

The candidates are all potential SF and OF pairs that can be generated from the UF of the word *'alles'* ('everything'). First off, the point needs to be made, that the winning candidates in this Tableau are based on the generalizations that could be made from the dataset across words. The word *'alles'* was not found to be written in the three acceptable ways, however other words with an underlying geminate /l/ were, as shown in example (11). The involvement of the UF is necessary for this process to be represented correctly, and for this reason, the <βiβi> /C:/ constraint is outranked by the <βiβi> |C:|, i.e. the same constraint that maps an underlyingly, not necessarily superficially long consonant onto a double grapheme. The <βiβi> |C:| and <γi>/Pi/ constraints are not ranked with respect to each other, so no candidate is thrown out by either of them, however, the third candidate violates both, which consists then of a fatal violation. The <βiβi> |C:| constraint represents the direct route, from the UF to the SF, while the <γi>/Pi/ the slower route, mapping the SF onto the OF. Since both of them are viable options, they cannot be ranked with respect to each other. The decision of whether to take the direct or the slower route is dependent on several factors, such as the necessity of orthographically representing a SF that differs a lot from the UF, e.g., in this case, speakers seem to feel like the vocalized /l/ differs too much from the underlying |lː| for them to keep the <ll> representation in an informal register context.

This modification to the model entails that if the UF is not modified by any structural constraint, it is going to be directly connected to the OF and represented through the interaction of ORTH and faithfulness constraints. If the SF is different from the UF, an interaction between the structural and ORTH constraints is needed. For the sake of clarity, ORTH- |UF| and ORTH- /SF/ constraints are distinguished. The first constraints interact with the faithfulness constraints, and the second with the structural constraints. Figure 4 illustrates how the OF is connected both to the UF and to the SF. The person writing either goes from a UF to a SF and then will produce a OF, or they can choose the direct route, making a direct UF-OF connection. The reasons as to why a speaker would choose the direct route (UF -> OF) or the slower one (SF-> OF) hypothesis will be further discussed in section 5.

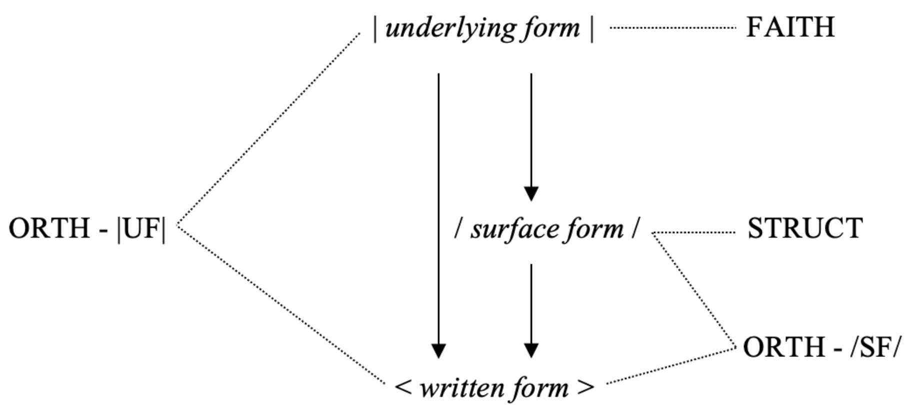


Figure 5. *Reading grammar, including the underlying form.*

Using the above-presented model (see figure (5)), a speaker can produce /u/ and write <ll>, by taking the direct way: from the UF to the OF and doing the same for the surface form /l/. However, if the speaker produces /u/ and writes <u>, they are taking the slower route, through the SF /u/ since they need the knowledge of the vocalized form in order to produce the <u> grapheme. The speaker can also go from the UF |lː| to the SF /lː/ and then to the OF <ll>, however, this entails an extra step through the SF, which is not necessary since no knowledge about phonological processes is needed here.

The following Tableau shows how the direct route from the UF to the OF works, when it comes to the general pattern that geminates are written with a double grapheme. This Tableau, together with Tableau (1) reinforces the theory that it is indeed possible to take either the direct route or the slower one and get to the same orthographic form. Here, since no phonological process modifies the surface form in a way that speakers feel the need to show it in writing, it is possible to suppose that they will prefer taking the direct route.

<table>
<tr><th>|pfif:ɛ|</th><th><βiβi> |C:|</th><th>FAITH- |f:| /f:/</th><th><γi>/Pi/</th></tr>
<tr><td>☞ /pfif:ɛ/ <pfiffe></td><td></td><td></td><td></td></tr>
<tr><td>/pfif:ɛ/ <pfife></td><td>*!</td><td>*</td><td></td></tr>
<tr><td>/pfifɛ/ <pfiffe></td><td></td><td>*!</td><td></td></tr>
<tr><td>/pfifɛ/ <pfie></td><td></td><td></td><td>*!</td></tr>
</table>

Tableau 7. *Formalization of the UF-OF route of the general pattern <βiβi> |C:|.*

This Tableau has the same candidates presented in Tableau (1), however, the constraints are different since the pair being evaluated by the orthographic constraints is the UF-OF pair. First off, the <βiβi> |C:| constraint, as shown in Tableau (6), maps an underlyingly long consonant onto a double grapheme, which allows for the surface form to be different from the UF and the OF to be still represented with the double grapheme. The surface form is then evaluated by the faithfulness constraints in the direct route, in this case by the FAITH- |f:| /f:/ constraint, which throws out the candidates with a singleton in the SF.

4.2 Lexical exceptions

Concerning the words like *'mussme'* and *'ufggä',* no phonological explanation was found, as to why there would be a singleton instead of the expected geminate in speech, and the only possible explanation seems to be the articulatory difficulty of the segments involved that can hardly allow for a double as long closure phase. The orthographic explanation is however straightforward: the PARADIGM constraint (Wiese 2004) postulates that *‘mussme’* is written with <ss>, as the different forms of the paradigm need to maintain a sort of graphic identity (see constraint 13). This allows to account for those words that do not belong to the specific category of lexical exceptions that usually store SF and OF together, but have a phonological singleton and a written double consonant in this dataset. Tableau (3) below illustrates how this is ranked. It follows that the paradigm of *‘muesse’* or *‘musse’* (‘to have to’), that contains a double grapheme to indicate the geminate /s:/ present in the majority of the forms in the paradigm, is also kept in the forms that do not have a phonological geminate due to other phonological processes.

(13) PARADIGM

Assign a violation mark to any component of a paradigm that does not maintain the paradigm’s orthographic identity, independently of how the word is pronounced.

Constraint (13) aims to maintain the graphic identity of a paradigm, since it is easier for people reading to visually recognize the word and consequently its meaning if all other words of the paradigm are written in the same way.

/mʊɛsmə/	PARADIGM	<βiβi> /C:/
☞ <mussme>		*
<musme>	*!	

Tableau 8. *Ranking of the* PARADIGM *constraint in Bern-German.*

Based on the data gathered, it is clear that the graphic representation needs to be preserved throughout the paradigm, which necessarily leads to the PARADIGM constraint being ranked over the <βiβi> /C:/ constraint in Tableau (8). The inverted order would lead to the winning of the second candidate (the written form with a single consonant), which, however, does not reflect the way speakers represent this word graphically.

This constraint, when analyzed from the perspective of BiPhon's model leads to the conclusion that the underlying form always needs to be mapped onto the orthographic form of the words in a paradigm, so that all words contained in it can be written the same way, regardless of any phonological processes that might apply due to the word's context. The following Tableau (9) formalizes this phenomenon by employing the UF-OF mapping, with the interaction of a structural constraint that applies in the specific word's context.

|mʊɛsːmə|	<βiβi> |Cː|	*/sːm/
☞ /mʊɛsmə/ <mussme>		
/mʊɛsːmə/ <musme>	*!	*
/mʊɛsːmə/ <mussme>		*!
/mʊɛsmə/ <musme>	*!	

Tableau 9. *Interaction of structural and underlying-orthographic constraints.*

Tableau (9) shows the interaction between structural constraints and underlying orthographic constraints. This works under the assumption that this word entails an underlying geminate, since the rest of the paradigm has one, and that due to phonological processes, that could not be explained, it has a surface singleton. In order to represent this phonological process, a constraint was introduced that assigns a violation mark to every geminate /sː/ before an /m/, since no example of this happening was found in the data (see constraint 14).

(14) */sːm/

Assign a violation mark to any geminate /sː/ followed by the segment /m/.

Moving onto the words containing singletons but written with a double grapheme, which all appear to be Latin-based loanwords (probably loaned from French), I will follow the hypothesis, based on the data obtained, that both the spelling and the pronunciation are stored together. Indeed, they appear to be lexical exceptions. Therefore, a constraint is needed, specific for these words that maps the specific UF onto the specific OF. For this analysis, the

ENTRY (UF | OF) constraint (n. 15) will be needed, as it assures that the UF and the OF are stored together and that the OF or the SF do not deviate from what is stored in the lexicon.

(15) ENTRY (UF | OF)

Assign a violation mark to every lexical exception that stores the UF and the OF together that is not written as the stored OF.

Also, in this case, it becomes necessary to involve the UF. Indeed, once the loanword has been adopted in another language system, not only the surface form is available to the speaker but also the underlying form. This is because once the loaned word has entered the vocabulary, children will acquire it directly from the parents (who only have a stored surface form, since they borrow the word) and they will not be able to recognize it as a loaned word, it will just be part of their vocabulary, meaning that they will have both an underlying and a surface form. In this case, the UF is then going to be mapped onto a specific OF that is learned in education settings. Since the loanword is now part of the system of the language and the OF is also loaned completely, a mismatch between the spoken and the written form (a singleton mapped onto a double grapheme) comes up. This is illustrated in Tableau (10) employing the word *'ammortisiert' and 'to absorb a shock'*.

|ɑmortisiᵊrt|	FAITH-|m| /m/	ENTRY(|ɑmortisiᵊrt| | <ammortisiert>)
/am:ortisiᵊrt/ <ammortisiert>	*!	
/amortisiᵊrt/ <amortisiert>		*!
☞ /amortisiᵊrt/ <ammortisiert>		
/am:ortisiᵊrt/ <amortisiert>	*!	*

Tableau 10. *Ranking of the ENTRY(UF | OF) constraint in Bern-German.*

The chosen candidates in Tableau (10) represent the two possible ways of representing this word graphically (as far as geminates are concerned): either with a double grapheme (candidate one) or with a single one (candidate two). The first candidate is the winning candidate, as it represents the OF as it is stored. The second candidate violates the ENTRY (UF | OF) constraint fatally, as based on the generalizations made on the collected data, this kind of words have a stored SF with a singleton but an OF with a double consonant. The two constraints are not ranked with respect to each other, since either order leads to the winning of the third candidate.

5 Discussion

In this paper, I investigated the way of graphically representing geminates in Bern-German, by searching for words containing identical double graphemes in Reddit posts and analyzing them in comparison to their phonological form. A critical point to the methodology employed with respect to the empirical question is that it does not allow checking whether singletons are always realized as single graphemes and whether there are phonological geminates that are not represented with a double grapheme. Indeed, only words containing double graphemes were analyzed here, so it was only possible to confirm that most words that contain a double grapheme correspond to phonological geminates, not to prove that all phonological geminates are written with a double grapheme or that singletons are always written with a single grapheme. Further research could focus on empirically testing this phoneme-grapheme correspondence in a more systematic way. Furthermore, it was not possible to formulate predictions about the writing of non-lexical, word-initial geminates in Bern-German, due to the lack of orthographic data concerning this phenomenon. Nevertheless, it would be interesting to understand this phenomenon better, by conducting specific empirical research on the phoneme-grapheme conversion of word-initial geminates. Another point should be made about the methodology: only one native speaker was consulted for this study. Indeed, the main goal of this paper was to analyze the phoneme-grapheme conversion of geminates in Bern-German, however, I was faced with the complete lack of IPA transcriptions in the Bern-German dictionaries[3], which only contained intuitive transcriptions of the words employing the alphabetic system. In the case of a non-standardized spelling, this, of course, can lead to different interpretations of the phoneme-grapheme conversion in different people, and thus to an unreliable transcription.

[3] https://www.berndeutsch.ch/ , https://www.idiotikon.ch/ (both last accessed on Monday, June 26th 2023).

The results showed that, in general, geminates are represented graphically with identical consecutive graphemes with only a few exceptions. Within these cases of inconsistency, two main patterns were found: the first concerns lexical exceptions: Latin-based loanwords, that have a phonological singleton corresponding to two graphemes; it can be hypothesized that the orthographic form was loaned into Bern-German together with the surface form and in time, the underlying form (as described in section 4). However, since no previous literature was found concerning this phenomenon and it was not the scope of the present paper, this was not further investigated. The other pattern of exceptions is the group of words containing a vocalized /l:/ that is underlyingly a geminate lateral, but it surfaces as either /u/ or /l:/ and is graphically represented with either <u> or <ll>. While formalizing this latter case, it was found that the mapping from the surface form onto the orthographic form is not sufficient to account for the re-emergence of <ll> in writing, after surfacing as /u/ in the surface form. Both the phonological underlying form and the surface form then need to be mapped onto the orthographic form, contrary to what was proposed by Hamann & Colombo (2017). Mapping only the underlying form directly onto the orthographic one, nevertheless, means losing any knowledge of the phonological processes happening from the UF to the SF, which, however, also do play a role in the writing process. Indeed, a speaker necessarily needs to apply the /l/ vocalization process to the |l:|, before going to the OF <u>. Otherwise, there would be a great mismatch in the employment of the established sound-grapheme correspondences; in fact, in German, the literacy language in Switzerland, the <l> is only employed to indicate the sound /l/ and never /u/. For this reason, it would not be logical to assume that Bern-German speakers randomly assigned two OF: <u> and <l> to the sound /l/. This is reinforced by the fact that the alternative orthographic form to <l>, <u>, is exactly the grapheme that is generally employed to represent the sound /u/, which is the vocalized form of /l/. It naturally follows that there needs to be a double mapping both from the UF and from the SF onto the OF. The more direct route, the UF -> OF route, as explained in the previous section, can be chosen in speed writing, of common and usual words, and the slower route (SF -> OF) in spelling new, complex, or infrequent words. The slower route is furthermore always taken when there is a phonological process influencing the written form. From the psycholinguistic literature concerning the reading process, indeed one learns that the grapheme-phoneme conversion route is usually taken when a word is infrequent, irregular, or long (Warren 2009: 141).

5.1 Psycholinguistic considerations

Considering the psycholinguistic literature about dual routes models of visual word recognition such as Coltheart et al's. (2001), it is attested that visual word recognition happens via two routes: the grapheme-phoneme conversion route (e.g. when reading non-words) and a whole-word route for frequent, regular words and experienced readers. The whole route entails that the reader recognizes the word by only focusing on a few of its graphemes, which are enough to activate the whole word form, evidence from this can be found in studies about the word superiority effect, as described by Warren (2013: 140), where letters are recognized faster if found in words rather than non-words. This together with the fact that 'typing mistakes are less detected if they preserve the word shape' (Warren 2013:141), can be taken as evidence that letter to letter recognition is needed also in the whole-word route and that they do, in fact, happen at the same time. It can then be assumed that in the phoneme-grapheme conversion route, as presented in this paper, the slower route is taken when reading infrequent, long, or irregular words, or graphemes that represent the surface form after a phonological process that distances the UF so much from the SF that speakers feel the need to represent it graphically. Instead, the direct route would be taken when recognizing the few letters of the word that are sufficient to activate the form of the whole word. If BiPhon's bidirectionality principle is assumed, and apply this concept also to the writing process, the direct route will be employed within the grapheme-phoneme conversion route when writing the more frequent words and straightforward phoneme-grapheme correspondence, while the slower route will be taken when writing less frequent, complex or long words. Nevertheless, while this seems to hold from a theoretical point of view, empirical studies are needed to test this.

5.2 Surface form vs underlying form mapping

The most discussed approaches of phonology-orthography conversion within the branch of phonology are indeed the morphophonemic one (a.o. Chomsky & Halle 1968), another term to describe the 'direct route' that maps the UF onto the OF and the phonemic one (a.o. Gudschinsky 1958, 1970, 1973), i.e., the correlate of the original *Reading Grammar*, where the surface form is taken as the phonological input. Snider (2014: 2) indeed claims that: "Consistently representing either the underlying form or the phonemic form is problematic, and neither approach can be recommended as a good overall strategy.". Both approaches have been shown to not be sufficient to explain different orthographic phenomena on their own. A few examples can be found also in this paper: the phonemic approach, i.e. the slower route of the above-presented model, cannot account for alternating spellings in uncodified languages and cannot deal with lexical exceptions, as it needs to keep the orthography as close as possible to the phonological output form (SF). The morphophonemic approach alone, i.e. the direct route of the above-presented model, however, cannot account for the phonological process that may apply to a certain UF and that can potentially influence the orthographic form if too distant from the UF e.g. the /l/ vocalization in the above-presented data. Yet, allowing for both structural and faithfulness constraints to interact with the orthographic constraints, while mapping both forms onto the orthographic form, allows to account for both the cases where the UF differs from the SF and those where it does not. The speaker, when writing, then goes from the underlying form to the surface form and then to the orthographic form. The passage through the surface form allows to account for non-transparent sound-orthography correlations, such as the previously formalized case of /l/ vocalization, where an underlying |l:| can be written as <u>. The underlying form is then likely to be employed directly in orthography, when there are such particular phonological processes applying to the UF, that force the speaker to represent these changes in writing too. Overall, it then seems that the *Reading Grammar* was not able to account for within-speaker variation due to the interaction and influence of different registers, such as in the case of /l/ vocalization. However, this model did provide the necessary connection to show the interaction between the structural and the orthographic constraints, showing that phonological processes are fundamental to graphical forms in transparent orthographies.

Snider (2014) proposes a theory that considers the lexical level to be the one that speakers are mostly aware of, when writing, the lexical processes are then defined as "processes that produce sounds that the native speaker is fully aware of" (Snider 2014). In his analysis, the author claims that if the processes between the UF and the SF are lexical, one should employ the phonemic mapping, i.e., the SF-OF mapping; if they are post-lexical, speakers prefer representing them using the UF, as they have no awareness about those processes and consequently do not feel like reflecting them in the orthography.

Concerning the previously- presented data, it is clear that speakers of Bern-German do have awareness concerning the underlying form of vocalized /u/, be it because of the influence of standard Swiss-German or the bimodal distribution of this phenomenon (due to its optionality). This seems to be indeed a lexical process, where the speakers are aware of all forms: the UF, the SF, and the OF, of course. It is possible to establish that the speakers are aware of this because the consulted native speaker was able to vocalize the |l:| even if the OF was a <ll>. The fact that if the |l| is not vocalized, it cannot be written with a <u> signals that once the speaker has awareness about a certain phonological process, they will not go back to the UF to represent the sound graphically, but they will stick to the SF. However, speakers do not seem to be aware of the fact that they represent the loanword lexical exceptions with a double grapheme, where it has a singleton in the phonology, which suggests that they use the UF, since there does not seem to be any exception to this rule (in the analyzed data set), and since the OF is loaned together with the UF, the speakers do not need to have any awareness about this process. This goes against Sinder's (2014) theory that either one form or the other is needed to represent either lexical or post-lexical processes, since it appears that both forms need to be involved for either form if one considers e.g. complex word writing. However, his theory could be giving an explanation as to why both forms are needed in terms of the general processes that apply.

An important point to be made is that this model was proven to be working for this data set in the writing direction, however, to get a full picture of whether the model is actually bidirectionally implementable or not. Another future question is whether this model is language and/or phenomenon-specific or can be applied to other phenomena in another language.

Concerning the paradigm of the verb *'muessen'*, it was impossible to establish what reason led to the realization of a singleton instead of a geminate. The current literature about syllable weight is insufficient to explain this phenomenon, and neither a moraic nor a segmental approach were useful in explaining this. For this reason, further research could focus on the formalization of syllable weight in this dialect by employing different approaches or describing this more in detail. Schlote (2008) does give an overview of syllable weight in Bern-German, however, the author does not seem to choose any particular approach over the other. Furthermore, there is a lack of empirical studies testing the phonology of this dialect. The traditional literature such as Marti (1985) does describe the phonology of the language, but no trace is found of any empirical data supporting this in later studies, except for Schlote (2008). The main findings of this study, as far as the phoneme-grapheme correspondences in Bern-German are concerned, definitely show that there seem to be at least two groups of words that do not represent a surface geminate with a double grapheme, i.e. the loanwords lexical exceptions and the paradigms exception such as '*mussme'*.

6 Conclusion

In this paper, it is proposed that the orthographic form of Bern-German geminates, two identical consecutive graphemes, is needed to be mapped onto both the underlying and the surface form. In the first part of the paper, the way of graphically representing geminates in Bern-German was empirically researched, by looking at words containing double graphemes in Reddit posts flagged as belonging to the Bern area. Three main groups were found: a general pattern that maps geminate onto double graphemes, a group of loanwords that represent lexical exceptions, and the vocalized representation of /l/, written as <u>. The general pattern of orthographic representing geminates, and the exceptions that were found while answering the first research question were then formalized by employing the *Reading Grammar* by Hamann & Colombo (2017) and relying on the principles of BiPhon-OT (Boersma 2011). While the *Reading Grammar* by Hamann & Colombo (2017) could account for the general pattern, it was found not sufficient to formalize the exceptions. It was then necessary to develop a new model able to account for these, adding the mapping of the underlying form to the orthographic form. This model seems to work when formalizing all patterns. Then, the theoretical, practical, and psycholinguistic implications of this model were discussed. In general, a few topics were touched upon, but not further discussed due to the time limitations, such as the formalization of this phenomenon in the reading direction applying the new model, its potential language-specificity, and applicability to other processes. Further research is furthermore needed concerning the generalizations about phoneme-grapheme correspondences in this language since most previous literature is either outdated or lacks empirical support.

Bibliography

Aeschimann, K., et al. *Wörterbuch Berndeutsch*. https://www.berndeutsch.ch/ (26/06/2023)

Auer, Peter. 1999. *Variabilität der intervokalischen Position in deutschen Trochäen.* In: Butt, Matthias & Nanna Fuhrhop (eds.). 1999. *Variation und Stabilitat in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietaten des Deutschen und anderer Sprachen* (Germanistische Linguistik 141–142, 1998). Hildesheim ; New York: G. Olms.

Aschwanden, Brigitte. 2001. *Wär wot chätä?' Zum Sprachverhalten deutschschweizerischer Chatter*. Hannover : Gottfried Wilhelm Leibniz Universität, Seminar für deutsche Literatur und Sprache. https://doi.org/10.15488/2915.

Biemann, C., Heyer, G., Quasthoff, U., & Richter, M. 2007. The Leipzig corpora collection-monolingual corpora of standard size. *Proceedings of Corpus Linguistic, 2007.*

Boersma, Paul. 2007. Some listener-oriented accounts of h-aspiré in French. Lingua 117, 1989–2054.

Boersma, Paul. 2011. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In Anton Benz & Jason Mattausch (eds.), *Linguistik Aktuell/Linguistics Today*, vol. 180, 33–72. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/la.180.02boe.

Boersma, Paul & Weenink, David. 2023. Praat: doing phonetics by computer [Computer program]. http://www.praat.org/. (31/05/2023).

Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon & Johannes Ziegler. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review* 108(1). 204–256. https://doi.org/10.1037/0033-295X.108.1.204.

Hamann, Silke. 2020. One phonotactic restriction for speaking, listening, and reading. In Martin Evertz-Rittich & Frank Kirchhoff (eds.), *Geschriebene und gesprochene Sprache als Modalitäten eines Sprachsystems*, 57–78. De Gruyter. https://doi.org/10.1515/9783110710809-004.

Hamann, Silke & Ilaria E. Colombo. 2017. A formal account of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. *Natural Language & Linguistic Theory* 35(3). 683–714. https://doi.org/10.1007/s11049-017-9362-3.

Gudschinsky, Sarah C. 1958. Native reactions to tones and words in Mazatec. Word 14: 338-345.

Gudschinsky, Sarah C. 1970. More on formulating efficient orthographies. The Bible Translator 21.1: 21-25.

Gudschinsky, Sarah C. 1972. Notes on neutralization and orthography. Notes on Literacy 14: 21-22.

Gudschinsky, Sarah C. 1973. A Manual of Literacy for Preliterate Peoples. Ukarumpa, Papua New Guinea: Summer Institute of Linguistics.

JetBrains. 2017. *PyCharm*. JetBrains. Available at: https://www.jetbrains.com/pycharm/ [Accessed September 2022].

Kolly, Marie-José. 2013b. Akzent auf die Standardsprachen: Regionale Spuren in “Français fédéral” und “Schweizerhochdeutsch.” *Linguistik Online* 58(1). https://doi.org/10.13092/lo.58.239.

Kraehenmann, Astrid & Aditi Lahiri. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *The Journal of the Acoustical Society of America* 123(6). 4446–4455. https://doi.org/10.1121/1.2916699.

Leipzig Corpora Collection. 2021.*United Kingdom.* Leipzig Corpora Collection. 30.000. *https://wortschatz.uni-leipzig.de/en/download/English#eng-uk_web_2002* .

Marti, Werner. 1985. *Berndeutsch-Grammatik für die heutige Mundart zwischen Thun und Jura*. Bern: Francke.

Naiditsch, Larissa. 1997. Zur Rekonstruktion der Silbenstruktur-Entwicklung in den Oberdeutschen Mundarten. *NOWELE. North-Western European Language Evolution* 31–32. 253–276. https://doi.org/10.1075/nowele.31-32.21nai.

n.d. *Languages and dialects.* https://www.eda.admin.ch/aboutswitzerland/en/home/gesellschaft/sprachen/sprachen-und-dialekte.html.

Neef, Martin, Anneke Neijt & Richard Sproat. 2012. *The Relation of Writing to Spoken Language*. Tübingen: Max Niemeyer Verlag.

RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Schlote, Elke. 2008. *Vokalqualität im Berndeutschen : Synchrone und diachrone Untersuchungen*. [Dissertation]. urn:nbn:de:bsz:352-opus-78975.

Schmidt, Larissa, Lucy Linder, Sandra Djambazovska, Alexandros Lazaridis, Tanja Samardžić & Claudiu Musat. 2020. A Swiss-German Dictionary: Variation in Speech and Writing. arXiv. http://arxiv.org/abs/2004.00139. (17 March, 2023)

Siebenhaar, Beat & Walter Vögeli. 2010. Mundart und Hochdeutsch im Vergleich. *Preprint für eine Neuauflage; vollständige Überarbeitung von: Walter Voegeli: Mundart und Hochdeutsch im Vergleich, in: Peter Sieber ; Horst Sitta (Hrsg.): Mundart und*

Hochdeutsch im Unterricht : Orientierungshilfen für Lehrer, Aarau ; Frankfurt am Main ; Salzburg : Sauerländer, 1988, Reihe Studienbücher Sprachlandschaft ; Bd. 1, S. 75–86, ISBN: 978-3-7941-3026-9, ISBN: 3-7941-3026-X. http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/14130.

Siebenhaar, Beat. 2003. Sprachgeographische Aspekte der Morphologie und Verschriftung in schweizerdeutschen Chats. *Linguistik Online* 15(3). https://doi.org/10.13092/lo.15.818.

Snider, Keith. 2014. *Orthography and Phonological Depth.* In: Cahill, Michael & Keren Rice (eds.). 2014. *Developing orthographies for unwritten languages* (Publications in Language Use and Education 6). Dallas, Texas: SIL International.

Spaelti, Philip. 2002. *Weak Edges and Final Geminates in Swiss German*. GLSA (Graduate Linguistic Student Association), Dept. of Linguistics, University of Massachusetts. https://doi.org/10.7282/T3T72FGC.

Staub, Friedrich, Tobler Ludwig, Bachmann Albert & Gröger, Otto (Hg.). 1973. *Schweizerisches Idiotikon. Wörterbuch der schweizerdeutschen Sprache*. Frauenfeld: Huber. https://www.idiotikon.ch/ (26/06/2023).

Warren, Paul. 2013. *Introducing psycholinguistics.* Cambridge University Press.

Wiese, Richard. 2005. How to optimize orthography. *Written Language & Literacy* 7(2). 305–331. https://doi.org/10.1075/wll.7.2.08wie.

Wiese, Richard. 2002. *The phonology of German* (The Phonology of the World’s Languages). Reprinted. Oxford: Oxford Univ. Press.

Appendix

Words containing phonological singletons but orthographic double consonants.

Plosive	Fricative	Nasal	Trill
mettwoch [mɛtvɔχ] ‘Wednesday’	assoziiert [ɑsot͡siᵊrt] ‘associate’	ammortisiert [ɑmortisiᵊrt] ‘to absorb a shock’	korruptionsübergang [korupt͡sio:n] ‘corruption’
ra**gg**lettöfeli [rɑglɛtœfeli] ‘Raclette-oven’	interessanti [ɪnteresɑnti] ‘interesting’	asymmetrisch [ɑsimetrɪʃ] ‘asymmetrical’	korräkt [korɛkt] ‘correct’
raggle**tt**öfeli [rɑglɛtœfeli] ‘Raclette-oven’	mussme [mʊɛsme] ‘have to + me’	grammatik [gramatik] ‘grammar’	narrativ [narati:f] ‘narrative’
ufggä [u:fgæ] ‘to go up’	professionelli [profesionɛl:i] ‘professional’	immunsystem [imunsʏste:m] ‘immune system’	
		kommentar [komenta:r] ‘comment’	
		kommentiert [komentiᵊrt] ‘commented’	
		kommunikation [komunikat͡sio:n] ‘communication’	
		kommuniziert [komunɪt͡siᵊrt] ‘communicate’	
		umprogrammiere [umprɔgrɑmi:re] ‘toprogramdifferently’	