Chapter 18

Animacy influences segmental phonology: The velar–sibilant alternation in BCMS

Marko Simonović
University of Graz

Bosnian/Croatian/Montenegrin/Serbian velar–sibilant alternation is a morphologised process that varies in application rates depending on the context. This article focuses on assibilation in DAT/LOC.SG of the nouns which end in *-a* in the citation form, where varying assibilation ratios are encountered. Two corpus studies targeting nouns with velar-final stems were conducted to establish the influence of phonological factors, animacy, and the presence of a non-alternating /i/ elsewhere in the paradigm on the alternation ratios. The results show that animacy comes out as a significant predictor of the alternation ratios in DAT/LOC.SG in both data sets.

1 Introduction

Bosnian/Croatian/Montenegrin/Serbian (BCMS) assibilation, whereby velars /k, g, x/ alternate with sibilants /ts, z, s/ in front of an /i/-initial affix, is a highly morphologised process whose application rates vary from context to context.[1] Table 1 shows four morphemes which all have the segmental content /i/. The imperative morpheme unexceptionally triggers the alternation, the NOM/VOC.PL morpheme triggers the alternation productively, but exceptions are attested. On

[1]To my knowledge, there are no major differences between the four varieties when it comes to assibilation. As clarified in Section 3, the empirical basis for this study originates from a Croatian and a Serbian corpus.

Marko Simonović. 2025. Animacy influences segmental phonology: The velar–sibilant alternation in BCMS. in Berit Gehrke, Denisa Lenertová, Roland Meyer, Daria Seres, Luka Szucsich & Joanna Zaleska (eds.), *Advances in formal Slavic linguistics 2022*, 531–553. Berlin: Language Science Press. DOI: 10.5281/zenodo.15394199

the other hand, the DAT/LOC.SG morpheme is one of the examples where it is hard to determine whether application or non-application is more common: the morpheme triggers the alternation in some words, fails to do so in others, and triggers it optionally in yet other words. Finally, the GEN.PL morpheme never triggers assibilation.

Table 1: Four assibilation contexts in BCMS

Application ratio	Morphological context	Examples
Categorical	IMP	/leg-i/ → [lezi] 'lie down'
High	NOM/VOC.PL (nouns)	/kirurg-i/ → [kirurzi] 'surgeons' (except in very few cases such as /deʧk-i/ → [deʧki] 'guys')
Medium	DAT/LOC.SG (nouns)	/bajk-i/ → [bajʦi] 'fairy tale (DAT/LOC)' /alg-i/ → [algi] 'alga (DAT/LOC)' /fresk-i/ → [freski] / [fresʦi] 'fresco (DAT/LOC)'
Zero	GEN.PL (nouns)	/bajk-i/ → [bajki] 'fairy tales (GEN)' /alg-i/ → [algi] 'algae (GEN)'

The primary focus of this contribution is on the DAT/LOC.SG ending /i/. I adhere to the assumption of traditional approaches to BCMS, asserting that the underlying form of the exponent of the DAT/LOC.SG morpheme remains consistent. Various factors, then, determine whether this exponent triggers assibilation. In other words, I do not adopt an overabundance analysis (see Thornton 2019 for the general approach and Lečić 2015 for an analysis in terms of overabundance in Croatian). An overabundance analysis would posit two DAT/LOC.SG endings differing solely in their assibilation behavior. The rationale behind this choice not to employ an overabundance analysis lies in the observation that variable assibilation is characteristic of a plethora of unrelated morphological contexts in BCMS. Assuming two different endings competing in all these contexts would face the problem of accounting for the fact that precisely these two endings compete in all these unrelated contexts.

The primary objective of the paper is to establish the factors determining the occurrence of assibilation in DAT/LOC.SG. The factors, as outlined in standard descriptions, will serve as predictors in a statistical model to anticipate assibilation encountered in the corpus (see Lečić 2016 for an overview of statistical modeling of variation observed in corpus data). In addition to phonological and morphological factors, special attention will be given to the influence of animacy on the velar–sibilant alternation in BCMS. This attention is warranted due to its theoretical relevance – the impact of a semantic factor on a segmental alternation.

The rest of the paper is organised as follows. Section 2 provides an overview of the predictors of assibilation and their treatment in the descriptive literature on BCMS. In Section 3, I describe the two corpus studies and the rationale behind them. Section 4 presents the results of the corpus studies, with the main finding being that animacy is a strong predictor of assibilation. Section 5 discusses the results and their theoretical implications. Finally, Section 6 concludes the paper.

2 Predictors of assibilation

Virtually all traditional descriptions of BCMS have a dedicated section on the application of assibilation in each of the morphological contexts. There are three important methodological obstacles in using these descriptions when modelling modern BCMS. Firstly, they often mix prescriptive objectives with descriptive ones, failing to distinguish between the two domains. Secondly, even when fully descriptive, they frequently lack a description of the empirical basis for the descriptions. Finally, they contain long lists of classes or intersections of classes in which assibilation is either favoured or blocked, without any indication of the strength of the generalisations or the size of classes in question. In what follows, I will provide a brief overview of the main phonological, morphological, and semantic factors that influence assibilation in BCMS. I will focus on generalisations that are applicable to a significant number of cases and can be meaningfully tested in a quantitative analysis. The ultimate goal is to define a list of properties which can be used in the quantitative analysis based on corpus data.

2.1 Phonological factors

The most general phonological factor, described in quite some detail in Težak (1986), is the FINAL VELAR, i.e. the difference between the three velars. /k/ is most prone to assibilation, /x/ assibilates in a minority of cases, whereas /g/ takes an intermediate position. This mirrors the size of the relevant classes within the declension class: *k*-final stems are much more common than *g*-final stems, which are in turn much more common than *x*-final stems (e.g., Petrović & Gudurić 2010: 475–478).

Several descriptions account for the non-application of assibilation in some classes by the fact that "the alternation would be experienced as moving away from the citation form of the word" (translation mine) (Pešikan et al. 2010: 47). While this criterion is extremely vague and therefore difficult to implement, Barić et al. (1997: 154) argue that the danger of excessively altering the stem is especially relevant for words with monosyllabic stems (simply because they have less stem material). Following this reasoning, we can implement the factor MONOSYLLABIC STEM as one of the predictors of non-assibilation.

All other phonological factors refer to stem-final consonant clusters, many of which block assibilation. While most such generalisations are tendencies, there is an unexceptional generalisation (first described as "self-evident" in Maretić 1963: 169): Assibilation never applies if the result would lead to total identity with the preceding consonant. This means that stems ending in *-tsk-*, *-zg-* and *-sx-* never alternate. Unfortunately, this generalisation only applies to a handful of items.

The long lists of rules concerning more frequent clusters (all of which are *k*-final) can be summarised as follows: all blockers have an obstruent stop or an affricate as the first member of the cluster, clusters which have a fricative as the first member allow both assibilation and non-assibilation, whereas clusters with more sonorous consonants tend to favour assibilation. We can conclude that the sonority of the consonant preceding the stem-final velar (C_1-SONORITY) is a predictor of assibilation. The more sonorous the first consonant of the cluster is, the more it is probable that assibilation will occur.

2.2 Morphological factors

The most important morphological predictor of assibilation is the specific morphological context. Since the present study only focuses on one specific context, this factor is controlled for. Still, I will take a brief look at two other morphemes which trigger assibilation in the nominal paradigm: NOM/VOC.PL *-i* and DAT/LOC/INS.PL *-ima*. The discussion of these two morphemes will be helpful in formulating a hypothesis concerning the morpheme in focus here.

As mentioned in Section 1, the NOM/VOC.PL morpheme *-i* triggers assibilation, with very few exceptions. This morpheme only shows up in the paradigms of masculine nouns, illustrated by the paradigm of [kirurg] 'surgeon' in Table 2. Such paradigms always contain another form with an assibilation-triggering ending: the DAT/LOC/INS.PL ending *-ima*. Interestingly, the DAT/LOC/INS.PL *-ima* also shows up in the paradigm of neuter nouns, where it is the only assibilation-triggering ending, as illustrated by the paradigm of [blago] 'treasure' in Table 2.

The actual acceptability of the DAT/LOC/INS.PL forms of the handful of neuter nouns with velar-final stems still needs to be established. The parallel forms with and without assibilation cited in the table are based on the description in Marković (2018: 136), who also points out that the forms with assibilation are in this case somewhat more marked. My personal judgement is ineffability in these case forms for all three nouns mentioned by Marković (2018). In other words, for all three nouns ([klupko] 'ball (of yarn)', [blago] 'treasure' and [ruxo] 'attire'), I cannot derive an acceptable form with the ending *-ima*. Either way, it is clear that DAT/LOC/INS.PL *-ima* triggers assibilation much more successfully in masculine paradigms than it does in neuter ones.

Table 2: Assibilation in masculine and neuter paradigms as illustrated by [kirurg] 'surgeon' and [blago] 'treasure' based on Marković (2018)

	Masculine		Neuter	
	SG	PL	SG	PL
NOM	kirurg	**kirurz-i**	blag-o	blag-a
GEN	kirurg-a	kirurg-a	blag-a	blag-a
DAT/LOC	kirurg-u	**kirurz-ima**	blag-u	**blag-ima/blaz-ima**
ACC	kirurg-a	kirurg-e	blag-o	blag-a
VOC	kirurg-u	**kirurz-i**	blag-o	blag-a
INS	kirurg-om	kirurz-ima	blag-om	**blag-ima/blaz-ima**

One possible way of understanding the empirical picture described above is that the tendency towards assibilation is stronger in cases where multiple assibilation-triggering endings occur in a paradigm. This could be due to the cumulative effect of these endings, which may enable the licensing of allomorphy. If this reasoning is correct, we would expect assibilation to be even more limited in paradigms containing another *-i* ending that does not trigger assibilation. The feminine nouns in focus here are actually the ideal testing ground for this hypothesis, because some of them have the GEN.PL ending *-i*, which never triggers assibilation, as mentioned in Section 1.

The distribution of the GEN.PL endings in the feminine declension in focus here can be summarised as follows. Nouns with a single stem-final consonant have the ending *-a* (e.g., [sʊraka] 'magpie.GEN.PL').[2] On the other hand, in nouns

[2]In all traditional descriptions, this GEN.PL ending contains a long vowel and causes a lengthening of the preceding vowel, e.g. [sʊraka] 'magpie.NOM.SG' vs. [sʊraːkaː] 'magpie.GEN.PL'. I am

with a stem-final consonant cluster three endings are attested: *-i* (e.g. in [kriŋki] 'disguise.GEN.PL'), *-aa*, whose first vowel breaks up the consonant cluster (e.g. in [banaka] 'bank.GEN.PL'), and, somewhat marginally, *-a* (e.g., in ?[kriŋka] 'disguise.GEN.PL' and ?[baŋka] 'bank.GEN.PL'). If endings can play a role in favouring/blocking allomorphy in other members of the paradigm, nouns which have GEN.PL in *-i* should be less prone to assibilation in DAT/LOC.SG than nouns which have other endings in GEN.PL. In other words, if -I[GEN.PL] is a strong factor that blocks assibilation, most nouns should behave either as [kriŋka] 'disguise' (i.e., have a GEN.PL in *-i* and no assibilation in DAT/LOC.SG) or as [baŋka] 'bank.GEN.PL' (i.e., not have a GEN.PL in *-i* and exhibit assibilation in DAT/LOC.SG).

Table 3: Assibilation in [baŋka] and lack of assibilation in [kriŋka]

	SG	PL	SG	PL
NOM	baŋk-a	baŋk-e	kriŋk-a	kriŋk-i
GEN	baŋk-e	banak-a	kriŋk-e	kriŋk-i
DAT/LOC	**bants-i**	baŋk-ama	**kriŋk-i**	kriŋk-ama
ACC	baŋk-u	baŋk-e	kriŋk-u	kriŋk-e
VOC	baŋk-o	baŋk-e	kriŋk-o	kriŋk-e
INS	baŋk-om	baŋk-ama	kriŋk-om	kriŋk-ama

2.3 Semantic factors

All traditional descriptions contain lists of classes in which assibilation is blocked, often defined by one semantic and one formal criterion. Almost all such classes are restricted to animates. For instance, Težak & Babić (1992: 92–93) include classes such as: personal male and female names, surnames regardless of their origin, names of pets and domestic animals, terms of endearment, ethnonyms derived using the suffixes *-ka*, *-nka* and *-čanka*, nouns in *-jka* which mean a female person with a certain characteristic or are derived from loanwords, and many others. While none of the descriptive works that I am aware of suggest that there is a direct link between animacy and lack of assibilation, overviews as the one sketched above justify the implementation of ANIMACY as one of the factors that blocks assibilation.

ignoring both vowel length and suprasegmental information here, because there is considerable variation in this respect, including large numbers of speakers who do not have distinctive vowel length.

Indirect evidence that animacy plays a role in blocking assibilation is contained in the discussions of minimal pairs which emerge when words that typically refer to animates (often female inhabitants) acquire an additional inanimate referent, e.g., a factory, a restaurant, etc. In such cases, all sources report that assibilation becomes possible with the inanimate referent. Examples are *Podravka* (the inhabitant of the Podravina region or a factory in Koprivnica, Croatia), *Beograđanka* (the female inhabitant of Belgrade or a building in Belgrade), and *Japanka* (a Japanese woman) vs. *japanka* (flip flop). Most normative sources argue against assibilation in such cases (e.g., Hudeček 2022 and Pešikan et al. 2010: 47), while others just describe it (e.g., Barić et al. 1997: 154). A potential example of an extension in the other direction would be the word *stranka*, which most commonly means 'party' (e.g., political party) but in some contexts can mean 'client'. My intuition is that assibilation is only possible in the former meaning.

3 Methodology

Given the shortcomings of the existing descriptions, in order to get a realistic picture of the data, I obtained data from two web corpora of BCMS: hrWaC and srWaC (Ljubešić & Klubička 2014). Corpus data are especially valuable when studying phenomena that exhibit a significant amount of variation, because they allow for the calculation of the relative frequencies of the specific options. In the case of assibilation in DAT/LOC.SG, which allows for a considerable amount of variation, the relevant construct to be employed here is the ASSIBILATION RATIO. The assibilation ratio of a word is the proportion of the DAT/LOC.SG forms with assibilation, calculated as the number of DAT/LOC.SG forms with assibilation divided by the total number of DAT/LOC.SG forms. For instance, if three DAT/LOC.SG tokens of the word [loziŋka] 'password' were extracted and two of them are [lozint͡si], whereas one is [loziŋki], then the ASSIBILATION RATIO for this noun is 0.67.

The outcome variable, ASSIBILATION RATIO, can be computed for any noun. The same is true for ANIMACY, FINAL VELAR and MONOSYLLABIC STEM. However, some of the predictors can only be meaningfully applied to a subset of nouns. Specifically, C_1-SONORITY and -I[GEN.PL] (implemented here as -I[GEN.PL] RATIO) can only be applied to nouns with stems that end in a consonant cluster. I therefore address the nouns with stem-final consonant clusters in a separate study.

3.1 Study 1: Assibilation with CC-final stems in hrWaC

This study was based on data from the Croatian web corpus hrWaC. In order to obtain nouns with CC-final stems, I first conducted a CQL search for lemmas ending in -CGa, where C is any consonant and G is any velar. The results were ranked by frequency and the 204 most frequent nouns were copied to a separate table.[3] Each noun was then annotated for FINAL VELAR, MONOSYLLABIC STEM, C_1-SONORITY and ANIMACY.

The annotation for ANIMACY was implemented analogously to the category of animacy in the masculine declension, where animacy influences the exponence of the ACC.SG (ACC.SG = GEN.SG for animates, ACC.SG = NOM.SG for inanimates). For instance, the noun [lutka] 'puppet' was annotated as animate because its masculine counterpart [lutak] 'male puppet' declines as animate.

For simplicity, the predictor C_1-SONORITY was implemented as a binary variable. A value of `0` was assigned to cases where the first consonant of the stem-final cluster is an obstruent stop or an affricate, while a value of `1` was assigned to all other cases.

An initial overview of the data showed that FINAL VELAR could not be meaningfully included as a factor, because, among 204 most frequent nouns, there were no *x*-final stems and only eight *g*-final stems. I therefore decided to only include *k*-final items in this study and replaced the eight *g*-final items with the next eight *k*-final items from the frequency list.

For each of the targeted nouns the counts of all the possible DAT/LOC.SG and GEN.PL forms were obtained by processing the results of CQL queries. Based on these counts the values for ASSIBILATION RATIO and -I[GEN.PL] RATIO were calculated. Specifically, since the morphological tags were found to be unreliable, CQLs were used to find strings in which the word in question is preceded by two congruent adjectival words. This method proved to yield a sufficiently precise sample, which could be manually cleaned within the constraints of the available time and manpower. The CQL used for the DAT/LOC.SG form of the word [freska] 'fresco' is shown in (1a), while (1b) shows the CQL used for the GEN.PL forms of the same noun.[4]

(1) a. `[word = ".*oj"][word = ".*oj"][word = "fres(c|k)i"]`

[3] The numbers of items eventually included in the study depended on the available time and manpower. However, it should be pointed out that the sample did include low frequency nouns, whose meaning needed to be looked up.

[4] The employed endings from the adjectival declension uniquely identify the relevant paradigm cells: *-oj* only appears in DAT/LOC.SG, whereas *-ih* [ix] only appears in GEN.PL.

b. `[word = ".*ih"][word = ".*ih"][word = "fres(ki|aka|ka)"]`

The nouns for which one of the queries yielded an empty result were removed and supplanted by the following word from the frequency ranking.

For the statistical analysis, the data were transformed so that each attestation of the DAT/LOC.SG form in the corpus constituted a separate observation (row in the table). This allowed us to treat the outcome variable ASSIBILATION as a binary variable. All relevant data, including ASSIBILATION, MONOSYLLABIC STEM, C_1-SONORITY, ANIMACY, and -I[GEN.PL] RATIO, are published in Simonović (2024). These data were inputted into a mixed-effects logistic regression model in R, where ASSIBILATION served as the outcome variable, and MONOSYLLABIC STEM, C_1-SONORITY, ANIMACY, and -I[GEN.PL] RATIO were treated as fixed effects. Additionally, the specific noun was included as a random factor to account for random variance between different nouns.

3.2 Study 2: Assibilation with VC-final stems in srWaC

This study was based on data from the Serbian web corpus srWaC.[5] In order to obtain nouns with VG-final stems, we first conducted a CQL search for lemmas ending in -VGa, where V is any vowel and G is any velar. The results were cleaned and ranked by frequency. The 349 most frequent nouns were copied to a separate table and annotated for FINAL VELAR, MONOSYLLABIC STEM and ANIMACY.[6]

The annotation for ANIMACY was implemented as in Study 1. Since the morphological tags were found to be unreliable, the values for ASSIBILATION RATIO were obtained by processing results of two CQL queries. Specifically, CQLs were used to find strings in which the target word is preceded by one of the typical prepositions (2a illustrates this for [baraka] 'barrack') and strings in which the word in question is preceded by a congruent adjectival word (2b).

(2) a. `[lemma = "(o|u|na|prema|k|ka)"][word = "bara(c|k)i"]`

b. `[word = ".*oj"][word = "bara(k|c)i"]`

The search results were manually cleaned and the ASSIBILATION RATIO was calculated for each noun. The nouns for which both queries yielded an empty result

[5]Data collection for this study was conducted in collaboration with participants of the course Collecting and Analyzing Corpus and Experimental Data in Hypothesis-Driven Linguistic Research at the University of Novi Sad.

[6]As with the previous study, the numbers of items eventually included in the study depended on the available time and manpower. However, it should be pointed out that the sample did include low frequency nouns, whose meaning needed to be looked up.

were removed and supplanted by the following word from the frequency ranking.

As in Study 1, the data were transformed so that each attestation of the DAT/LOC.SG form constituted a separate observation (row in the table). This allowed us to treat the outcome variable ASSIBILATION as a binary variable. All relevant data, including values for FINAL VELAR, MONOSYLLABIC STEM, ANIMACY and ASSIBILATION are published in Simonović (2024). These data were inputted into a mixed-effects logistic regression model in R, where ASSIBILATION served as the outcome variable, and FINAL VELAR, MONOSYLLABIC STEM, ANIMACY were treated as fixed effects. Additionally, the specific noun was included as a random factor to account for random variance between different nouns.

4 Results

4.1 Study 1

Before presenting the results of the statistical model, a brief overview of the mean values for the ASSIBILATION RATIO is provided. In this study, the overall mean ASSIBILATION RATIO is 0.33. The means for all groups identified by single values of the binary variables, along with the number of items in these groups, are presented in Table 4.

The mean ASSIBILATION RATIO for animate nouns exhibits a notably low value, also indicating a significant difference of means concerning ANIMACY. Similarly, and as expected, a considerable difference in means is observed for C_1-SONORITY. Specifically, stems in which the first consonant of the stem-final cluster is an obstruent stop or an affricate, display, on average, a lower ASSIBILATION RATIO compared to stems with different consonant configurations.

Interestingly, the difference in means for MONOSYLLABIC STEM is relatively small, but it also deviates from the expected pattern: monosyllabic stems exhibit a higher mean ASSIBILATION RATIO than polysyllabic ones.

The binary predictor variables mentioned earlier, along with the continuous predictor variable -I[GEN.PL]RATIO (with a mean of 0.87 in the dataset), were incorporated as fixed factors in a generalised linear mixed model. The binary variables MONOSYLLABIC STEM, C_1-SONORITY and ANIMACY were stored as factors, while -I[GEN.PL]RATIO was the only numeric factor. Individual lemmas, were included as a random factor, with by-noun varying intercepts. In (3) I provide the formula for the model as implemented in R using the package lme4 (Bates et al.

Table 4: Mean ASSIBILATION RATIO for each value of the binary variables

Variable	Mean AR for 1 (N)	Mean AR for 0 (N)	Difference
ANIMACY	0.04 (77)	0.51 (127)	−0.47
C_1-SONORITY	0.39 (160)	0.13 (44)	0.26
MONOSYLLABIC STEM	0.40 (80)	0.29 (124)	0.11

2015). The complete script is published in Simonović (2024). The summarised results can be found in Table 5.[7]

(3)
```
model1 <- glmer(assib ~ mono + anim + c1son + igenpl + (1 |
noun), family = binomial(link = "logit"), data = analysis1)
```

Table 5: Generalised linear mixed model results

Variable	Coefficient	Std. Error	z value	Pr(> $|z|$)	Odds Ratio
(Intercept)	3.3012	1.0109	3.266	0.00109**	27.1447
MONOSYLLABIC STEM	0.1950	0.6303	0.309	0.75703	1.2153
ANIMACY	−6.7532	0.8073	−8.366	$< 2 \times 10^{-16}$***	0.0012
C_1-SONORITY	−4.9472	0.7643	−6.473	9.62×10^{-11}***	0.0071
-I[GEN.PL] RATIO	−2.3075	0.9823	−2.349	0.01882*	0.0995

The *Coefficient* column in Table 5 provides the estimated coefficients, revealing the log-odds change in the outcome variable for a one-unit change in each predictor. These estimates offer valuable insights into the direction and magnitude of the predictors' impact. Accompanying the estimates, the *Std. Error* column indicates the standard error of each coefficient estimate. This information is crucial for assessing the precision and reliability of the estimated coefficients. The z-statistic is calculated as the coefficient estimate divided by the standard error, where larger values indicate a larger estimated effect size. The p-value represents the probability of observing an effect at least as large as the one found assuming the null hypothesis is true. Using an *alpha*-level of .05, we consider effects where

[7]The significance codes used in this report follow the standard R output format: '***' for p-values $\leq$ 0.001, '**' for p-values $\leq$ 0.01, '*' for p-values $\leq$ 0.05, '.' for p-values $\leq$ 0.1, and no extra symbol for p-values > 0.1. These codes are retained from the R output for consistency in reporting results.

$p < .05$ to be significant. Lastly, the *Odds Ratio* column shows the exponentiated coefficients, offering a clear understanding of the multiplicative change in odds for a one-unit change in each predictor. This column provides practical insights into the implications of the predictors on the odds of the outcome.

Summarising the findings, Table 5 indicates, that among the four predictors, only two exhibit a highly statistically significant relationship with ASSIBILATION: ANIMACY and C_1-SONORITY both demonstrate particularly strong negative associations, as evidenced by their low *p*-values. The predictor -I[GEN.PL] RATIO shows a significant negative association with ASSIBILATION, aligning notably in the expected direction. It should, however, be noted that the magnitude of the effect size (OR) is somewhat lower than all the other significant predictors. Finally, the predictor MONOSYLLABIC STEM does not show a statistically significant relationship with the outcome variable. Consequently, the unexpected positive difference of means observed earlier can be attributed to chance rather than a meaningful association.

4.2 Study 2

As with Study 1, I begin with a brief overview of the mean values for the ASSIBILATION RATIO. In this study, the overall mean ASSIBILATION RATIO is 0.75, which is much higher than in the previous study. The means for all groups identified by single values of the binary variables, along with the number of items in these groups, are presented in Table 6.

The mean ASSIBILATION RATIO for animate nouns exhibits a notably low value, indicating a significant disparity in means concerning ANIMACY. Interestingly, the difference in means for MONOSYLLABIC STEM is also relatively high, and it goes in the expected direction: monosyllabic stems exhibit a lower mean ASSIBILATION RATIO than polysyllabic ones.

Table 6: Mean ASSIBILATION RATIO for both binary variables

Variable	Mean SR for 1 (N)	Mean SR for 0 (N)	Difference
ANIMACY	0.13 (40)	0.83 (309)	−0.70
MONOSYLLABIC STEM	0.44 (101)	0.87 (248)	−0.42

Table 7 shows the mean ASSIBILATION RATIO for the 3 values of the variable FINAL VELAR. As expected, *k*-final stems have the highest mean, whereas the *x*-final stems have the lowest mean.

Table 7: Mean ASSIBILATION RATIO for the three values of FINAL VELAR

Mean AR for k (N)	Mean AR for g (N)	Mean AR for x (N)
0.86 (248)	0.54 (82)	0.19 (19)

Both binary predictor variables discussed above, as well as the categorical predictor variable FINAL VELAR were included in a generalised linear mixed model. Additionally, the specific noun was entered as a random variable. In (4) I provide the formula for the model as implemented in R using the package lme4. The complete script is published in Simonović (2024). The results are summarised in Table 8.

(4)
```
model2 <- glmer(assib ~  finalvelar + anim + mono + (1 | noun),
family = binomial(link = "logit"), data = analysis2)
```

Table 8: Generalised linear mixed model results

Variable	Coefficient	Std. Error	z value	$Pr(>\|z\|)$	Odds Ratio
(Intercept)	9.4696	0.4308	21.980	$< 2 \times 10^{-16}$***	1.296×10^{4}
FINAL VELAR:G	−5.2410	0.8887	−5.897	1.23×10^{-9}***	0.00529
FINAL VELAR:X	−10.0896	1.6726	−6.032	1.62×10^{-9}***	4.15×10^{-5}
ANIMACY	−9.9452	1.2072	−8.238	$< 2 \times 10^{-16}$***	4.80×10^{-5}
MONOSYLLABIC STEM	−6.3654	0.7886	−8.072	6.90×10^{-16}***	0.00172

First, it is important to note that the categorial variable FINAL VELAR was dummy coded, resulting in two of its values appearing in the list. The baseline value, FINAL VELAR:K, is used as the reference category for comparison.

The model results indicate that all incorporated predictors exhibit a negative association with ASSIBILATION. Specifically, for the variable FINAL VELAR, which was omitted in Study 1, both g and x demonstrate negative associations with the outcome. Similarly, ANIMACY maintains a negative association, consistent with the findings of Study 1. Notably, unlike in Study 1, MONOSYLLABIC STEM also shows a negative association.

5 Discussion

Having presented the results of the two corpus studies, an evaluation of the three types of factors presented in Section 2 is in order.

Regarding the phonological factors, C_1-SONORITY has clearly come out as an important predictor in Study 1, as did the FINAL VELAR in Study 2. Indirectly, the difference between the mean ASSIBILATION RATIO in the two studies (0.33 vs. 0.75), although not statistically tested, points in the direction of a more general influence of the sonority of the segment preceding the stem-final velar.[8] The issue is somewhat less clear when it comes to the factor MONOSYLLABIC STEM, which only came out as significant in Study 2.

As regards the morphological factor -I[GEN.PL] RATIO, its relation with ASSIBILATION was found to be statistically significant, but of relatively weak magnitude compared to the other factors. Especially in the context of the ongoing debate on the relevance of paradigms for phonological computation (see, e.g., Bobaljik 2008), the presented findings cannot be taken as firm evidence that other paradigm cells influence assibilation.

Finally, ANIMACY unequivocally emerges as an influential factor in determining the application of assibilation. The described pattern then joins other, better described and understood, animacy effects in BCMS morphology. Animacy has been well described to influence the exponence of ACC.SG in the main masculine declension in BCMS, leading to minimal pairs such as, e.g., [tip-a] 'guy.ACC.SG' vs. [tip] 'type.ACC.SG'. The influence of animacy on BCMS tonal patterns has also been discussed in the literature, especially for the DAT/LOC.SG ending [-ú], which seems to realise its underlying High tone only in inanimate monosyllables, leading to minimal pairs such as [tíip-u] 'guy.DAT/LOC.SG' and [tiip-ú] 'type.DAT/LOC.SG' (vs. [tíip-a] 'guy/type.GEN.SG'; see Martinović 2012 for a recent quantitative analysis).

Prima facie, the assibilation pattern seems much more gradient than the other two animacy-controlled patterns. The closest we get to a categorical effect is the blocking of assibilation in animates. It is therefore worthwhile to take a closer look at the exceptional animates that display assibilation. The main insight is that there are extremely few animates that display assibilation more often than not (i.e. have assibilation ratios above 0.5). In Study 1, these are only 3 (out of 77 animates): [majka] 'mother', [pomajka] 'foster mother' and [djevojka] 'girl(friend)'. In Study 2, out of 40 animate nouns, 5 have assibilation ratios higher than 0.5: [supruga] 'wife', [unuka] 'granddaughter', [sluga] 'servant', [sʋastika] 'sister-in-

[8]The difference would have been even bigger if stems ending in consonant clusters in /g/ and /x/ had been included in Study 1. I am not aware of a single noun from this group that undergoes assibilation in modern BCMS. Barić et al. (1997: 154) mention [kaʋga] 'conflict' as the only Cg-final stem that undergoes assibilation, but most modern speakers seem to either not know this word or use it without assibilation. Including such items in Study 1 would have then additionally lowered the ASSIBILATION RATIO in this study

law' and [vladika] 'bishop'. The fact that all of these nouns refer to roles suggests that roles might belong to a distinct category between animates and inanimates. If this holds true, we can assert that there exists a clear prohibition on assibilation within the category of true animates in the DAT/LOC.SG, and that the possibility of assibilation emerges for entities falling lower on the animacy hierarchy, such as roles and (other) inanimates.

6 Conclusion

The present study aimed to investigate the influence of phonological, morphological and semantic factors on the application of assibilation in DAT/LOC.SG in BCMS nouns. Through a comprehensive analysis of data and statistical modeling, it has become evident that ANIMACY plays a central role in determining the occurrence of assibilation in this context.[9] A detailed analysis of the individual exceptions to the generalisation that animates do not allow assibilation showed that assibilation is restricted to animates that have the meaning of roles.

While providing a complete formal account of the observed pattern is reserved for future research, the results of this study facilitate the formulation of desiderata for such an account. In the spirit of advancing incrementally, the following steps are suggested to establish a connection between the phenomenon described here and its closest related phenomena.

The most closely related phenomenon appears to be the tonal pattern observed in the DAT/LOC.SG forms of the main masculine declension. The two phenomena both exhibit a more intimate phonological interaction with case endings in inanimates compared to animates. This interaction is manifested as a tonal shift in one case and as assibilation in the other. However, a notable distinction lies in the fact that masculine declension roles do not permit the imposition of the DAT/LOC.SG ending's tonal pattern.

The next in line closely related domain is the occurrence or absence of assibilation elsewhere in the nominal and adjectival declensions. The DAT/LOC.SG data presented above suggest the presence of a boundary that hinders phonological interactions between the stem and the case ending in animates. However, this boundary seems to disappear in the plural cases of the masculine declension, where animates undergo assibilation without restriction (e.g., in [ʧex] 'Czech man.NOM.SG', [ʧesi] 'Czech man.NOM.PL', [ʧesima] 'Czech man.DAT/LOC/INS.PL').

[9] As argued by one of the reviewers, the statistical tests used here are telling in terms of the statistical significance of the coefficients rather than on predictive power for novel data. We leave the latter type of analysis to future works.

Conversely, it is noteworthy that the adjectival declension never allows assibilation, despite having numerous *i*-initial case endings.

Future research will also profit from more extensive data collection, not only from corpora, but also from elicited production, wug experiments, etc. It is worth noting, that although the present study encompassed a sizable data sample, certain nouns had to be excluded due to the absence of encountered forms, particularly in Study 1, where the absence of GEN.PL forms led to the exclusion of many nouns with attested DAT/LOC.SG forms. Moreover, it is possible that there are further factors which were not included in the analysis.

Finally, an important aspect that was not addressed here is the precise representation of animacy. The observed consistency in assibilation among animate entities suggests the possibility of formalising animacy as the presence of an additional feature or structure.

Overall, the findings of this study contribute to our understanding of the intricate relationship between animacy and phonological processes in BCMS. While animacy's influence on other aspects of BCMS morphology has been previously described, this study unveils a novel finding by demonstrating its comprehensive impact on the application of segmental phonological alternations.

Abbreviations

ACC	accusative	LOC	locative
DAT	dative	NOM	nominative
GEN	genitive	PL	plural
IMP	imperative	SG	singular
INS	instrumental	VOC	vocative

Acknowledgments

The author thanks the audience of FDSL 15 and the anonymous reviewers, as well as Peđa Kovačević and the participants of the course Collecting and Analyzing Corpus and Experimental Data in Hypothesis-Driven Linguistic Research held at the University of Novi Sad. Special thanks go to Igor Marchetti and Maja Miličević Petrović for all the help with the statistical analysis. This research was funded in whole by the Austrian Science Fund (FWF, Grant-DOI 10.55776/I6258).

References

Barić, Eugenija, Mijo Lončarić, Dragica Malić, Slavko Pavešić, Mirko Peti, Vesna Zečević & Marija Znika. 1997. *Hrvatska gramatika* [Croatian grammar]. Zagreb: Školska knjiga.

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using {lme4}. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Bobaljik, Jonathan. 2008. Paradigms (optimal and otherwise): A case for skepticism. In Asaf Bachrach & Andrew Nevins (eds.), *Inflectional identity*, 29–54. Oxford: Oxford University Press.

Hudeček, Lana. 2022. O podravki koja radi u Podravci, Pepeljugi i Branki. *Hrvatski jezik: znanstveno-popularni Časopis za kulturu hrvatskoga jezika* 9(4). 28–32.

Lečić, Dario. 2015. Morphological doublets in Croatian: The case of the instrumental singular. *Russian Linguistics* 39. 375–393. DOI: 10.1007/s11185-015-9152-7.

Lečić, Dario. 2016. *Morphological doublets in Croatian: A multi-methodological analysis*. Sheffield: University of Sheffield. (Doctoral dissertation).

Ljubešić, Nikola & Filip Klubička. 2014. bs,hr,srWaC - web corpora of Bosnian, Croatian and Serbian. In Felix Bildhauer & Roland Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg: Association for Computational Linguistics. DOI: 10.3115/v1/W14-0405.

Lüdecke, Daniel. 2024. *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.16. https://CRAN.R-project.org/package=sjPlot.

Maretić, Tomislav. 1963. *Gramatika hrvatskoga ili srpskoga književnog jezika*. Zagreb: Matica hrvatska.

Marković, Ivan. 2018. *Hrvatska morfonologija* [Croatian morphonology]. Zagreb: Disput.

Martinović, Martina. 2012. The interaction of animacy with phonetic and phonological factors in Neoštokavian pitch accents. In Jaehoon Choi, E. Alan Hogue, Jeffrey Punske, Deniz Tat, Jessamyn Schertz & Alex Trueman (eds.), *Proceedings of the 29th West Coast Conference on Formal Linguistics*, 161–168. Somerville.

Pešikan, Mitar, Jovan Jerković & Mate Pižurica. 2010. *Pravopis srpskoga jezika, izmenjeno i dopunjeno izdanje* [Orthography of Serbian: Revised and expanded edition]. Novi Sad: Matica srpska.

Petrović, Dragoljub & Snežana Gudurić. 2010. *Fonologija srpskog jezika* [Phonology of Serbian]. Beograd: Institut za srpski jezik SANU / Beogradska knjiga / Matica Srpska.

Simonović, Marko. 2024. *Sibilarisation in the Dative/Locative Singular in BCMS (Data from hrWaC and srWac) [Data set]*. DOI: 10.5281/zenodo.13825333.

Težak, Stjepko. 1986. Sibilarizacija u suvremenomu hrvatskom književnom jeziku. *Filologija* 14. 395–402.

Težak, Stjepko & Stjepan Babić. 1992. *Gramatika hrvatskoga jezika: Priručnik za osnovno jezično obrazovanje*. Školska knjiga.

Thornton, Anna M. 2019. Overabundance: A canonical typology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation* (Studies in Morphology 5), 223–258. Cham: Springer.

Appendix A Plots for predicted probabilities

Below I report the plots for predicted probabilities for each predictor. These were obtained in R using the package sjPlot (Lüdecke 2024) in R. The formula was implemented as `plot_model(model1, type = "pred", terms = "name_of_the_predictor")`.

A.1 Study 1

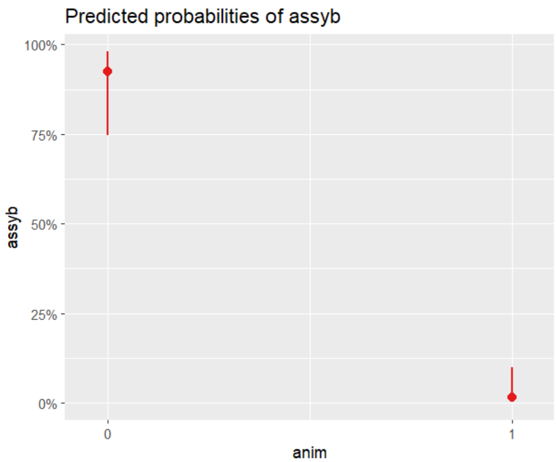

Figure 1: Animacy

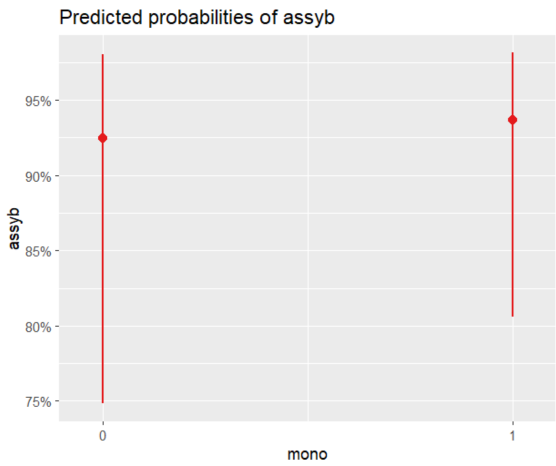

Figure 2: monosyllabic stem

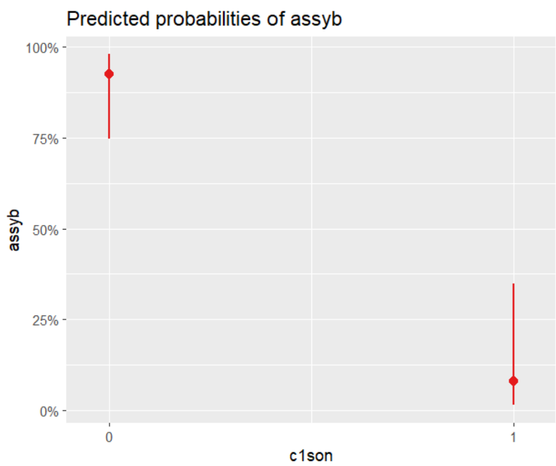

Figure 3: C_1-sonority

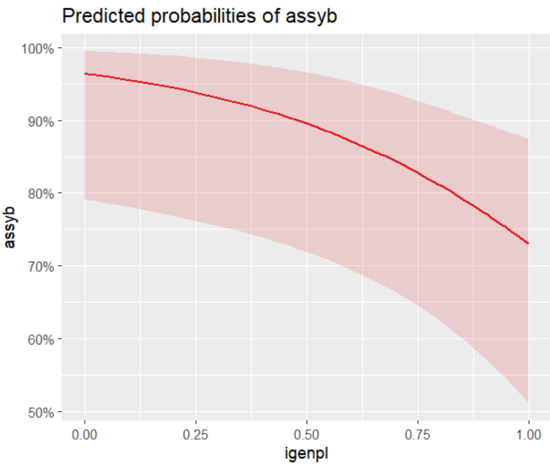

Figure 4: I[GEN.PL] ratio

A.2 Study 2

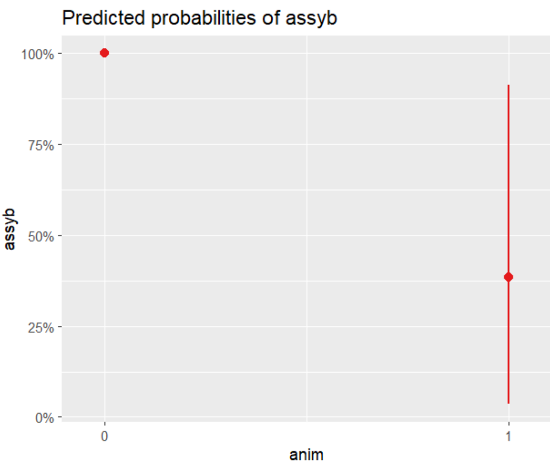

Figure 5: Animacy

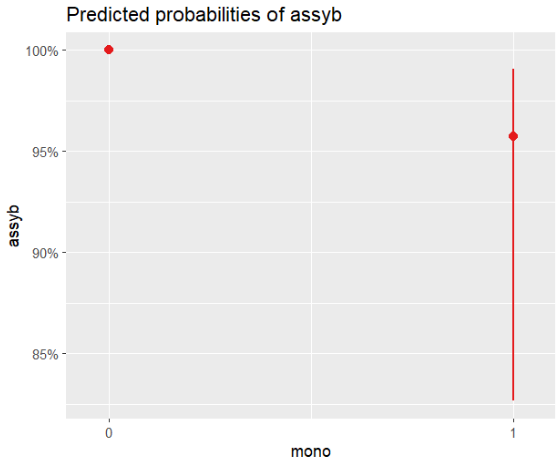

Figure 6: monosyllabic stem

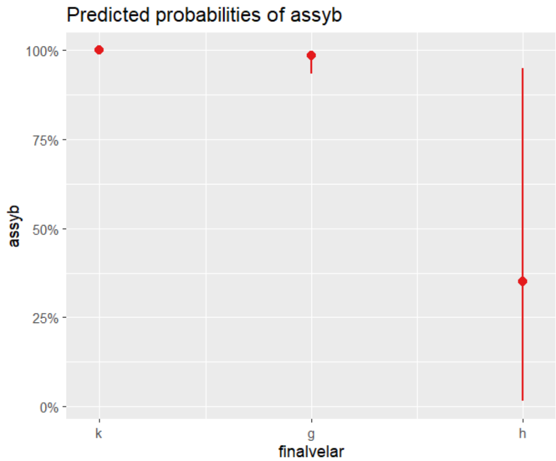

Figure 7: Final Velar

Appendix B R session info

```
R version 4.4.1 (2024-06-14 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 10 x64 (build 19045)
```

Matrix products: default
Locale:

```
[1] LC_COLLATE=Dutch_Netherlands.utf8
    LC_CTYPE=Dutch_Netherlands.utf8
    LC_MONETARY=Dutch_Netherlands.utf8
    LC_NUMERIC=C
    LC_TIME=Dutch_Netherlands.utf8
```

Time zone: Europe/Vienna
tzcode source: internal

Attached base packages:

```
[1] stats     graphics  grDevices utils     datasets  methods   base
```

Other attached packages:

```
[1] sjPlot_2.8.16   lme4_1.1-35.5   Matrix_1.7-0    apaTables_2.0.8
    devtools_2.4.5  usethis_3.0.0   psych_2.4.6.26  readxl_1.4.3
```

Loaded via a namespace (and not attached):

```
[1] gtable_0.3.5       xfun_0.47          ggplot2_3.5.1
    htmlwidgets_1.6.4  remotes_2.5.0      insight_0.20.3     lattice_0.22-6
    sjstats_0.19.0     vctrs_0.6.5        tools_4.4.1        generics_0.1.3
    datawizard_0.12.2  parallel_4.4.1     tibble_3.2.1       fansi_1.0.6
    pkgconfig_2.0.3    RColorBrewer_1.1-3 ggeffects_1.7.0    lifecycle_1.0.4
    farver_2.1.2       compiler_4.4.1     stringr_1.5.1      sjmisc_2.8.10
    munsell_0.5.1      mnormt_2.1.1       httpuv_1.6.15      htmltools_0.5.8.1
    later_1.3.2        pillar_1.9.0       nloptr_2.1.1       urlchecker_1.0.1
    tidyr_1.3.1        MASS_7.3-60.2      ellipsis_0.3.2     cachem_1.1.0
    sessioninfo_1.2.2  boot_1.3-30        nlme_3.1-164       mime_0.12
    sjlabelled_1.2.0   tidyselect_1.2.1   digest_0.6.37      performance_0.12.2
    stringi_1.8.4      dplyr_1.1.4        purrr_1.0.2        labeling_0.4.3
    splines_4.4.1      fastmap_1.2.0      grid_4.4.1         colorspace_2.1-1
    cli_3.6.3          magrittr_2.0.3     pkgbuild_1.4.4     utf8_1.2.4
```

```
broom_1.0.6         withr_3.0.1         scales_1.3.0        promises_1.3.0
backports_1.5.0     cellranger_1.1.0    memoise_2.0.1       shiny_1.9.1
knitr_1.48          miniUI_0.1.1.1      profvis_0.3.8       rlang_1.1.4
Rcpp_1.0.13         xtable_1.8-4        glue_1.7.0          pkgload_1.4.0
rstudioapi_0.16.0   minqa_1.2.8         R6_2.5.1            fs_1.6.4
```