

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

Marko Simonović, University of Graz
marko.simonovic@uni-graz.at



Velar-sibilant alternation in BCMS

- Highly morphologised process
- Velars /k, g, x/ → sibilants /^hts, z, s/ in front of /i/-initial suffix.

/ruk-i/ → **rut^hsi** ‘hand(dat/loc)’

- Not all /i/-initial suffixes trigger the alternation.
- Even those that trigger the alternation, do so at different rates.

Categorical (non-)sibilarisation

<p>Imperative of the \emptyset/e conjugation</p>	<p>/pek-i/ → petsi ‘bake’ /rek-i/ → retsi ‘say’ /leg-i/ → lezi ‘lie down’</p>
<p>Genitive plural of some nouns with a citation form in <i>-a</i> and a CC-final stem.</p>	<p>/bajk-i/ → bajki ‘fairy tales(gen)’ /majk-i/ → majki ‘mothers(gen)’ /maŋg-i/ → maŋgi ‘mangas(gen)’ /fresk-i/ → freski ‘frescos(gen)’</p>

Almost categorical sibilantisation

Nominative
plural of
nouns with a
citation form
in a consonant

/obelisk-i/ → **obeslists̩** ‘obelisks’

/kirurg-i/ → **kirurzi** ‘surgeons’

/monarx-i/ → **monarsi** ‘surgeons’

but exceptionally:

/detʃk-i/ → **detʃki** ‘guys’

/pelazg-i/ → **pelazgi** ‘Pelasgians’

/bronx-i/ → **bronxi** ‘bronchi’

Medium sibilisation ratio

Dative/locative
singular of
nouns with a
citation form
in *-a*

/bajk-i/ → bajt̃si ‘fairy tale(dat/loc)’

/majk-i/ → majt̃si ‘mother(dat/loc)’

/lozink-i/ → loziŋki/loziŋt̃si ‘password(dat/loc)’;

/fresk-i/ → freski/frest̃si ‘fresco(dat/loc)’

/tsurk-i/ → tsurki ‘girl(dat/loc)’

/alg-i/ → algi ‘alga(dat/loc)’

Research question

- What are the predictors of sibilantisation triggered by DAT/LOC.SG /i/?

Candidate 1: **GEN.PL /i/**.

Never triggers the alternation.

This may block sibilantisation (e.g., by Lexical Conservatism).

baŋk-a vs kriŋk-a

'bank'

	Singular	Plural
Nom	baŋk-a	baŋk-e
Gen	baŋk-e	banak-a ??baŋk-i *baŋk-a
Dat/Loc	banṡ-i *baŋk-i	baŋk-ama
Acc	baŋk-u	baŋk-e
Voc	baŋk-o	baŋk-e
Ins	baŋk-om	baŋk-ama

baṅk-a vs kriṅk-a

'disguise'

	Singular	Plural
Nom	kriṅk-a	kriṅk-e
Gen	kriṅk-e	kriṅk-i *krinak-a *kriṅk-a
Dat/Loc	kriṅk-i *krints̄-i	kriṅk-ama
Acc	kriṅk-u	kriṅk-e
Voc	kriṅk-o	kriṅk-e
Ins	kriṅk-om	kriṅk-ama

Candidate 2: **Animacy**

Hints from BCMS descriptive literature.

Silić & Pranjković (2005: 161):

- names of persons and animals never alternate,
- inhabitants and nationals never alternate.

‘Minimal pairs’ from Težak (1986: 401-402)

Podrav[ts]-i ‘factory name(dat/loc)’

Zor[ts]-i ‘factory name(dat/loc)’

Podravk-i ‘woman for Podravina(dat/loc)’

Zork-i ‘human name(dat/loc)’

Candidate 3: **Target velar (/k/ vs. /g/ vs. /x/)**

Težak (1986:401): /k/ sibilises the most, and /x/ the least.

Candidate 4: **Monosyllabic base**

Monosyllabic bases may resist the alternation in order to maintain recoverability.

/kuk-i/ → **kuki** ‘hook(dat/loc)’

/okuk-i/ → **okutsi** ‘curve(dat/loc)’

Candidate 5: C_1 in stem-final CCs

Obstruent stops and affricates as C_1 don't go well with sibilants.

/repk-i/ → repki 'national team(dat/loc)'

/tʃetk-i/ → tʃetki 'brush(dat/loc)'

/zvetʃk-i/ → zvetʃki 'rattle(dat/loc)'

/mask-i/ → mastsi 'masl(dat/loc)'

/zbirk-i/ → zbirtsi 'collection(dat/loc)'

/stauk-i/ → stautsi 'item(dat/loc)'

Measuring sibilisation (and GEN.PL /i/) based on corpus data

Sibilisation in DAT/LOC.SG (the dependent variable) and **GEN.PL /i/** measured as ratios of tokens of the specific form in the corpus.

- *lozinka* ‘password’ has 3 DAT/LOC.SG tokens: 1 *loziŋki* and 2 *loziŋsi*.
 - The sibilisation ratio for *lozinka* is 0.67.
- *lozinka* has 21 GEN.PL tokens and all of them are *loziŋki* (0 *loziŋka*, 0 *lozinaka*).
 - The GEN.PL /i/ratio for *lozinka* is 1.

Study 1: CC-final stems in hrWaC

GEN.PL /i/ and C₁ only apply to CC-final stems.

- Independent variables
 - GEN.PL /i/ (ratio),
 - **Animacy** (1 or 0),
 - **Monosyllabicity** (1 or 0),
 - C₁=**stop/affr** (1=obstruent stop/affricate, 0=all others).
 - ~~**Final velar** (/k/ or /g/ or /x/)~~
 - Needed to be excluded due to extremely few items with /g, x/.

Study 1: Data set and first annotation

- In hrWaC (Ljubešić & Klubička 2014), a CQL search was conducted for lemmas ending in -CGa, where C is any consonant and G is any velar.
[lemma = ".*(a|e|i|o|u)(b|c|č|ć|d|đ|dž|f|g|h|j|k|l|m|n|nj|p|r|s|š|t|v|z|ž)(k|g|h)a"]
- The results were ranked by frequency.
- The 130 most frequent nouns were copied to a separate table and annotated for **Animacy**, **Monosyllabicity**, **C₁=stop/affr** and **Final velar**.
- **Extremely uneven distribution of the final velars:** no final /x/, final /g/ in only 3 items and final /k/ in 127 items.
 - Items that don't have a final /k/ were excluded and replaced with /k/-final items.

Study 1: Getting the ratios

- CQL searches for DAT/LOC.SG and GEN.PL were conducted
 - In each case two congruent adjectival words + target from
 - [word = ".*oj"] [word = ".*oj"] [word = "lozin(c|k)i"]
 - [word = ".*ih"] [word = ".*ih"] [word = "lozin(ki|aka|ka)"]
- Search results manually cleaned and the two ratios were calculated for each noun.
- The nouns for which one of the searches yielded an empty result were removed and supplanted by new words from the lemma frequency ranking.

Study 1: Regression model

Predictor	<i>beta</i>	<i>p</i>	Uniqueness	<i>r</i>	Fit
GEN.PL_i	-0.15	0.02	2%	-.07	
Animacy	-0.56	< .001	30%	-.52**	
Monosyllabicity	-0.07	0.33	0%	-.04	
C1_stopaffr	-0.50	< .001	22%	-.44**	
					$R^2 = .521^{**}$

Study 2: VG-final stems in srWaC*

- Independent variables
 - **Animacy** (1 or 0),
 - **Monosyllabicity** (1 or 0),
 - **Final velar** (/k/ or /g/ or /x/)
 - /g, x/ lumped together because rare.

* Study conducted in collaboration with the participants of the course *Collecting and analyzing corpus and experimental data in hypothesis-driven linguistic research* at the University of Novi Sad.

Study 2: Data set and first annotation

- In srWaC (Ljubešić & Klubička 2014), a CQL search was conducted for lemmas ending in -VGa, where C is any vowel and G is any velar.
 - [lemma = ".*(a|e|i|o|u)(k|g|h)a"]
 - [lemma = ".*(b|c|č|ć|d|đ|f|g|h|j|k|l|m|n|p|r|s|š|t|v|z|ž)r(k|g|h)a"]
- The results were ranked by frequency.
- The 300 most frequent nouns were copied to a separate table and annotated for **Animacy**, **Monosyllabicity** and **Final velar**.
- **Uneven distribution of the final velars:** 19 /x/-final stems, 79 /g/-final stems and 202 /k/-final stems.
 - Turned into a binary variable: /k/-final items vs. /g,x/-final items.

Study 2: Getting the ratios

- 2 CQL searches for DAT/LOC.SG were conducted:
 - Preposition + target form
 - [lemma = "(o|u|na|prema|k|ka)"] [word = "bara(c|k)i"]
 - Congruent adjectival word + target form
 - [word = ".*oj"] [word = "bara(k|c)i"]
 - [word = ".*(o|e)m"] [word = "vladi(k|c)i"]
- Search results manually cleaned and the ratio calculated for each noun.
- The nouns for which the search yielded an empty result removed and supplanted by new words from the lemma frequency ranking.

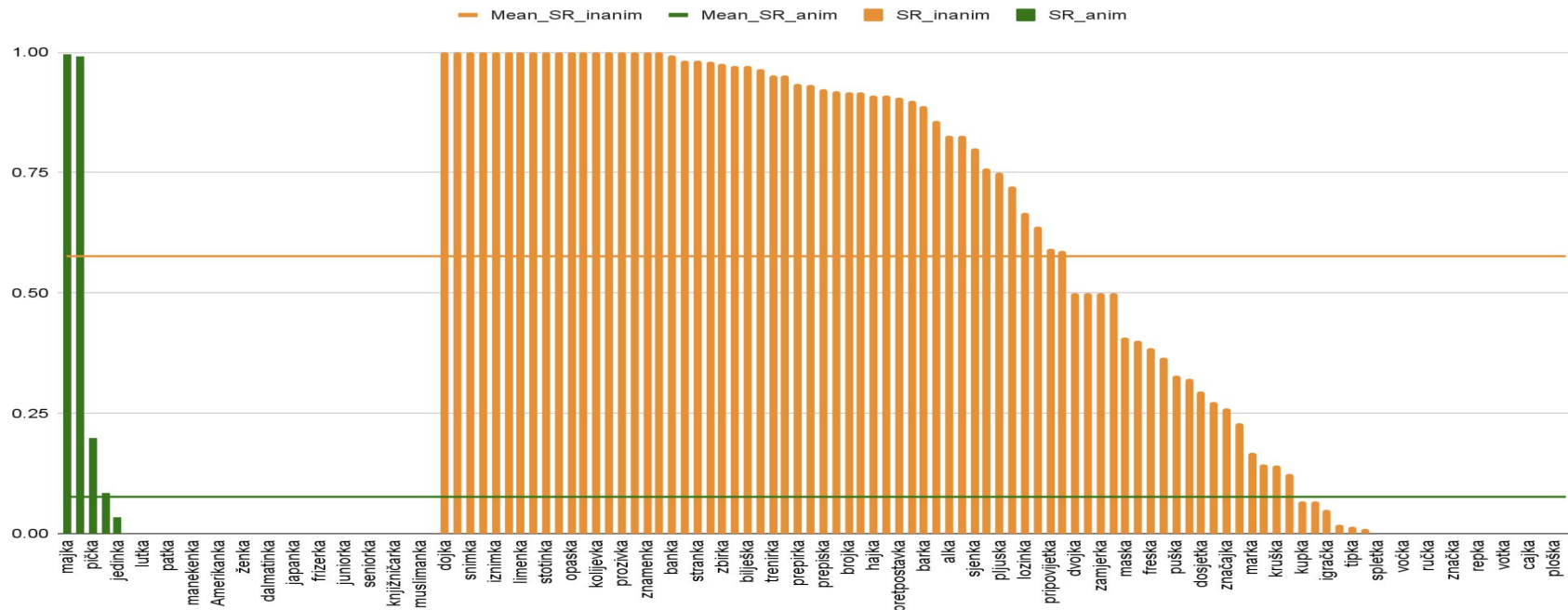
Study 2: Regression model

Predictor	<i>beta</i>	<i>p</i>	Uniqueness	<i>r</i>	Fit
Animacy	-0.43	< .001	17%	-.53**	
Monosyllabicity	-0.29	< .001	7%	-.44**	
Velar_k	0.26	< .001	7%	.37**	
					$R^2 = .454^{**}$

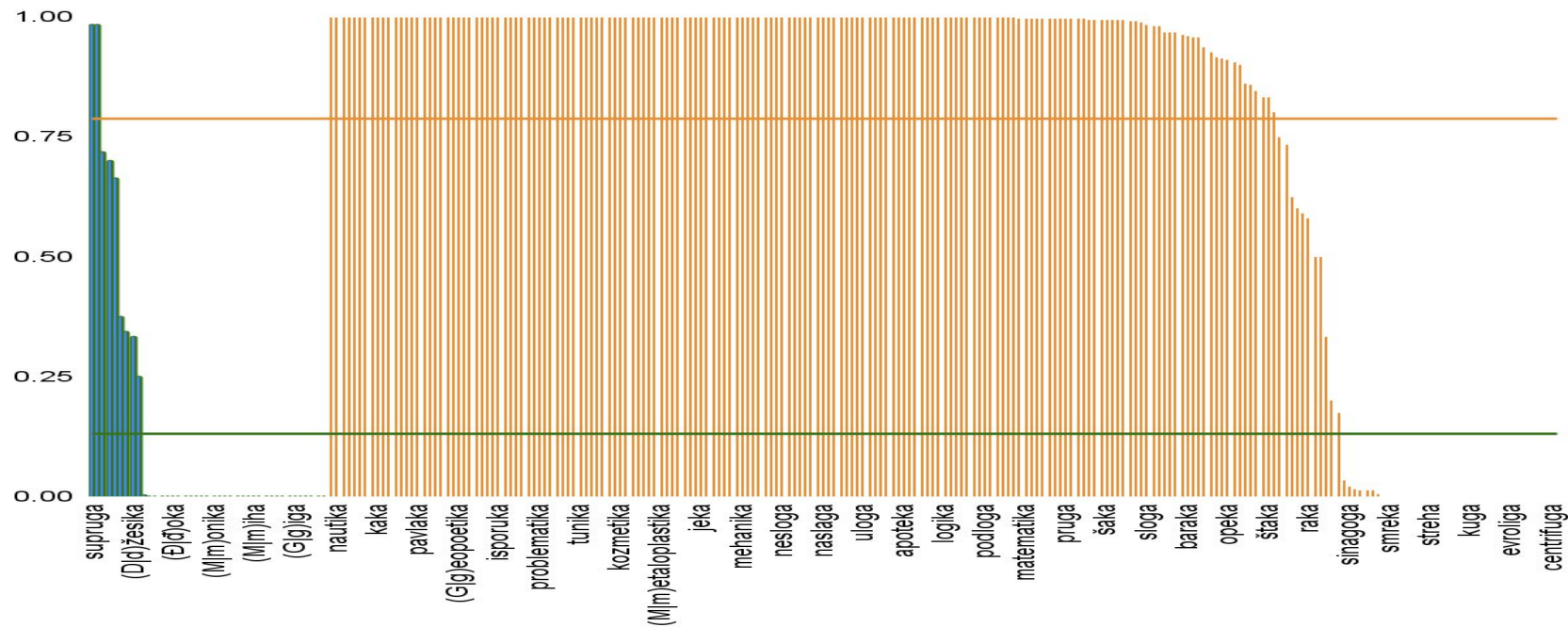
Zooming in on Animacy

- Animates and inanimates in both studies have a different mean sibilisation ratio
 - Study 1: animate: 0.07, inanimate: 0.57;
 - Study 2: animate: 0.15, inanimate: 0.81
- Few animate nouns are attested with sibilisation. In both datasets combined, only 7 animate nouns are found that have sibilisation ratios above 0.4.
 - Study 1 (29 animates): *majka* ‘mother’ and *djevojka* ‘girl(friend)’.
 - Study 2 (41 animates): *supruga* ‘wife’, *unuka* ‘granddaughter’, *sluga* ‘servant’, *svastika* ‘sister-in-law’ and *vladika* ‘bishop’.
 - They all refer to **roles**.

Animates vs. Inanimates in Study 1



Animates vs. Inanimates in Study 2



Where else Animacy matters

- ACC.SG in the main ‘masculine’ declension
 - *t-og tip-a* ‘that-ACC.SG guy-ACC.SG’ vs. *t-aj tip* ‘that-ACC.SG type-ACC.SG’
 - Not amenable to a phonological account.
- DAT/LOC.SG ending *-u*, that realises its underlying High tone only in inanimate monosyllables with stems in a single consonant.
 - *tí:p-u* ‘guy-DAT/LOC.SG’ vs. *ti:p-ú* ‘type-DAT/LOC.SG’ (see Martinović 2012 for a recent quantitative analysis).

Uniting the two DAT/LOC.SG observations?

- Non-modular solutions:
 - Special FAITHFULNESS constraints replicating the animacy scale from names to roles.
- Modular solutions:
 - Animates are spelled out differently due to an additional projection.

Issues to take into account

- The limitation on sibilisation in animates does not generalise to other inflectional endings.
 - /tʃex-i/ → tʃesi ‘Czechs’
 - /slovak-i/ → slovatsi ‘Slovaks’
- Roles behave as regular animates with the DAT/LOC.SG ending -u.
 - /brát-ú/ → brát-u ‘brother.DAT/LOC.SG’
 - /kú:m-ú/ → kú:m-u ‘best man.DAT/LOC.SG’

References

- Ljubešić, N & Klubička, F.** (2014). {bs,hr,sr}WaC - 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.
- Maretić, T.** (1963). Gramatika hrvatskoga ili srpskoga književnog jezika. Zagreb: Matica hrvatska.
- Martinović, M.** (2012). The Interaction of Animacy with Phonetic and Phonological Factors in Neoštokavian Pitch Accents. 29th West Coast Conference on Formal Linguistics (pp. 161-168). Cascadilla Proceedings Project.
- Silić, J., & Pranjković, I.** (2005). Gramatika hrvatskoga jezika: za gimnazije i visoka učilišta. Zagreb: Školska knjiga.
- Težak, S.** (1986). Sibilizacija u suvremenom hrvatskom književnom jeziku. *Filologija*, 14, 395-402.