

## Gradient similarity in antigemination: evidence from allomorph selection

Aljoša Milenković, Harvard University, aljosamilenkovic@g.harvard.edu

**Overview.** Cross-linguistically, adjacent identical consonants tend to be avoided (McCarthy 1986; Yip 1988). In many languages, antigemination has been found to target not only fully identical but also *sufficiently similar* adjacent consonants (Borowsky 1987; Pierrehumbert 1993; Rose 2000; Frisch et al. 2004; Baković 2005, 2006; Pajak and Baković 2010). This raises questions about how identity is assessed for geminate avoidance purposes. This paper investigates antigemination effects in phonologically-conditioned allomorph selection. I present corpus data from the mobile *a*-alternation in Bosnian (B), Croatian (C) and Serbian (S) (cf. Linzen et al. 2013 for a similar study in Russian). Several morphemes in these languages exhibit stochastic variation between C# and Ca# allomorphs, which is distinct from the well-known vowel-zero alternation. The mobile *a*-morphemes examined in this survey include prepositions [s]~[sa] ‘with’ and [k]~[ka] ‘towards’, and the GEN.SG.M/N adjectival ending [-ɔ:g]~[-ɔ:ga].

**Assimilation and antigemination.** B, C and S display regressive voicing and anteriority assimilation in sandhi, illustrated in (1) by the preposition /iz/ ‘from, out of’.

- |     |    |           |                 |    |            |                     |
|-----|----|-----------|-----------------|----|------------|---------------------|
| (1) | a. | iz avlije | ‘from the yard’ | c. | iʒ ɕɛpa    | ‘out of the pocket’ |
|     | b. | is parka  | ‘from the park’ | d. | if tʃinije | ‘out of the bowl’   |

This is important because, as Baković (2005, 2017) argues, a feature must independently trigger assimilation in a language to be ignored for similarity assessment in antigemination-induced phenomena. In B/C/S, voicing and anteriority trigger assimilation and are precisely the features that segments are allowed to differ in at the input level in antigemination contexts: /s/ is avoided before /s/, but also before /z/, /ʒ/ and /ʒ/, which deviate from /s/ in voicing and/or anteriority. Ultimately, /s/ undergoes total assimilation to all of these fricatives: /s#/ → [ʃ#]. Thus, antigemination effects are sensitive to output representations: the anti-geminate constraint has access to the *potential* output of phonological processes.

**Data.** I examined the realization of three mobile *a*-morphemes in large Internet corpora of B/C/S available on Sketch Engine (Kilgarriff et al. 2004). The corpus search was limited to bigrams that display free variation; lexicalized expressions were filtered out. Given space limitations, I herein report the results of only one of the three case studies conducted: the preposition [s]~[sa]. The results of other case studies are largely consistent with those for [s]~[sa].

**Results.** Allomorph distribution by phonological environment is plotted in Figure 1. The corpus data were statistically analyzed in R (R Core Team 2023) using logistic regression. Separate models were fitted for each language. In each model, the dependent variable was the choice of allomorph (coded as “0” for [s] and “1” for [sa]). The models included three fixed effects: Voicing mismatch (coded “1” if the following word begins with a voiced obstruent, “0” otherwise), Anteriority mismatch (coded “1” if the following word begins with a posterior coronal, “0” otherwise) and Identity (coded “1” if the following word begins with {s, z, ʃ, ʒ}, “0” otherwise). Regression analysis showed that, although B, C and S exhibit different rates of mobile *a* in environments where no phonological effects are at play, the phonological effects on [s]~[sa] realization were consistent across the languages. The models found a significant positive effect of Voicing mismatch on the CV realization in all three languages (B:  $\beta = .12$ ,  $p=.000$ , C:  $\beta = .11$ ,  $p=.000$ , S:  $\beta = .18$ ,  $p=.000$ ). A significant positive effect on the CV realization was likewise found of Anteriority mismatch (B:  $\beta = .3$ ,  $p=.000$ , C:  $\beta = .1$ ,  $p=.000$ , S:  $\beta = .32$ ,  $p=.000$ ). Crucially, there was a strong significant positive association between Identity and the CV realization (B:  $\beta = 4.22$ ,  $p=.000$ , C:  $\beta = 5.19$ ,  $p=.000$ , S:  $\beta = 3.68$ ,  $p=.000$ ). These findings suggest that constraints requiring adjacent obstruents to agree in voicing and anteriority, as well as the anti-geminate constraint, are all independently active in the phonological grammars of B/C/S and favor the the realization of the final vowel in [sa].

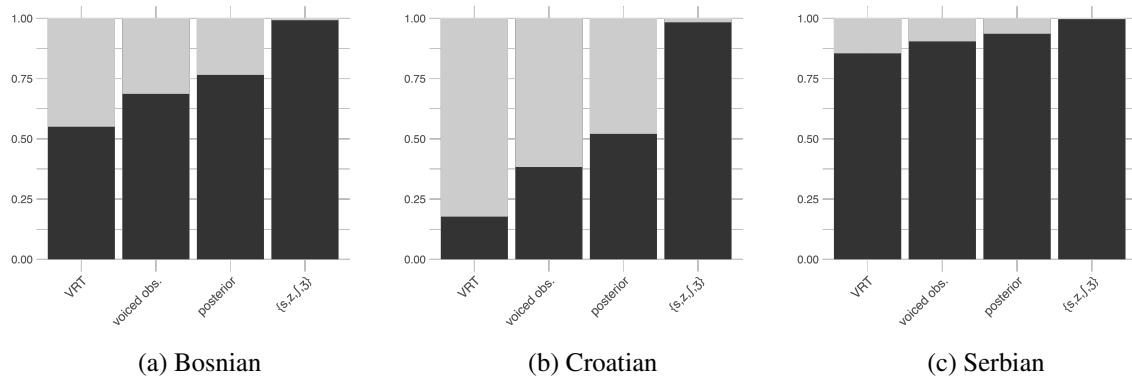


Figure 1: Vowel-less (light grey) and CV realization (dark grey), by phonological environment (x-axis) and language (faceting category). (VRT = vowels, sonorants, voiceless obstruents that do not trigger any assimilation process: {p,t,k,f,x,ts}.)

**Gradient avoidance effects.** The [s] allomorph is avoided before coronal fricatives at varying rates, as shown in Figure 2 for all three languages combined. The strength of avoidance effects depends on the degree of similarity between /s/ and the following consonant in the underlying form. In full identity contexts (/s#s/), geminate avoidance is nearly categorical. In the partial identity pairs /s#z/ and /s#ʃ/, where the following consonant differs from /s/ by just one feature—voicing (z) or anteriority (ʃ)—the avoidance is less strict. The weakest avoidance is observed in the /s#z/ pair, where both voicing and anteriority disagree.

A logistic regression analysis with stepwise difference coded comparisons showed that partial identity pairs (/s#z/ and /s#ʃ/) (coded “1” in the model) significantly reduced the likelihood of the CV allomorph relative to the /s#s/ pair (coded “0” [baseline]):  $\beta = -.92$ ,  $p=.000$ . The /s#z/ pair (coded “2”), with the most feature mismatches at the input level, further reduced the likelihood of the CV allomorph compared to the /s#z/ and /s#ʃ/ pairs:  $\beta = -.71$ ,  $p=.000$ . This indicates that the anti-geminate constraint can prevent the emergence of a sandhi cluster that would result in a geminate (Adler and Zymet 2021), while also being sensitive to the underlying similarity of cluster members.

**Implications.** The novel geminate avoidance data from B/C/S inform phonological theory by showing that the strength of geminate avoidance effects is determined by the degree of feature overlap between adjacent consonants. Antigemination effects were observed in both full and partial identity contexts, but with varying degrees of robustness: sequences of fully identical consonants, such as /s#s/, are avoided more stringently than those of partially identical consonants, such as /s#z/ or /s#ʃ/. Further, in partial identity contexts, the rate of geminate avoidance is gradiently correlated with the degree of similarity between adjacent consonants: the more features the consonants share, the stronger the avoidance effect. I discuss the implications of gradient similarity avoidance for the theory of phonological variation in probabilistic models such as MaxEnt HG (Goldwater and Johnson 2003). Using the `maxent` .ot R package (Mayer et al. 2024), I provide a MaxEnt learner for B/C/S antigemination data. The model based on the constraint set à la Baković 2005 fails to capture the pairwise contrasts between different antigemination environments in Figure 2, flattening the probabilities of CV realization across these environments. I therefore consider potential alternatives to Baković 2005’s approach to antigemination.

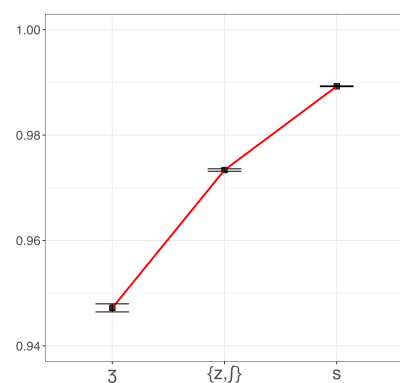


Figure 2: Rate of the CV allomorph across antigemination environments (y-axis), as a function of similarity (x-axis) for all three languages combined. Whiskers indicate Wilson confidence intervals.