OCP-place beyond phonotactics: sandhi place restrictions in Croatian

Aljoša Milenković

Harvard University

The 31st Manchester Phonology Meeting

May 30, 2025

Antigemination

- Ban on adjacent identical (or sufficiently similar) consonants (McCarthy 1986; Yip 1988; Odden 1988)
  - ❖ **PROCEDURAL**: no phon. rule may result in a geminate (McCarthy 1986; Yip 1988; Côté 1997, 2004)
  - ❖ **PHONOTACTIC**: ban on cooccurrence within morphemes (Pierrehumbert 1993; Frisch et al. 2004)

- Unified phenomenon or two separate phenomena?

Antigemination & OCP

- In OT, procedural and phonotactic antigemination captured using different machinery
- **Status of OCP-place?** (Rose 2000; Baković 2005)
- assimilation-antigemination link
(Baković 2005, 2006, 2017; Pająk & Baković 2010)
- Consequence: near-geminates avoided without OCP-place
(Baković 2005: 292)
- OCP-place still used for phonotactic antigemination
(Pierrehumbert 1993; Frisch et al. 2004; Coetzee & Pater 2008)

Today's talk

1. similarity avoidance without assimilation in Croatian allomorph selection

2. OCP-place effect in procedural antigemination (needed in both types)
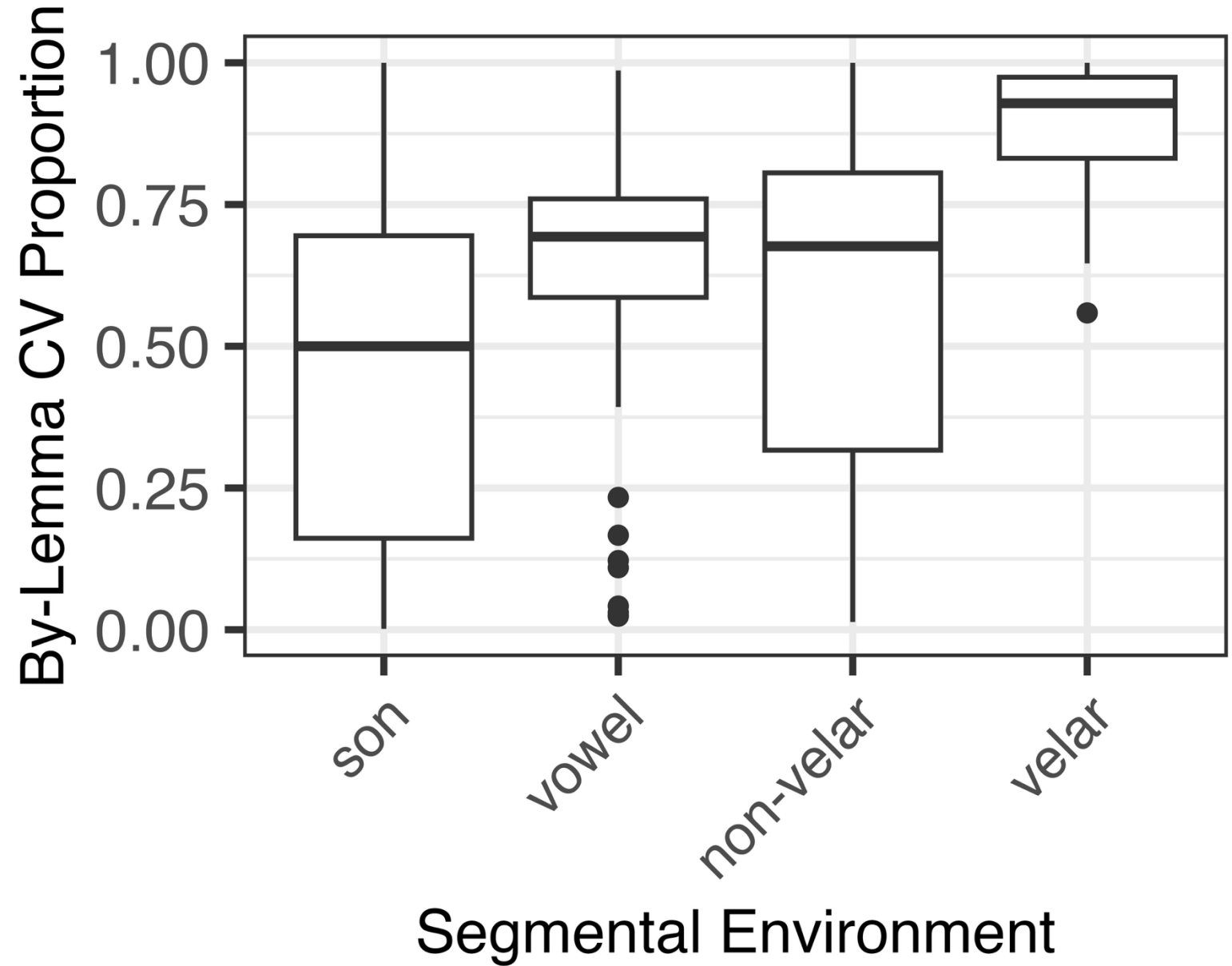
Mobile *a*-alternations in Croatian

- free variation between C-final and CV-final allomorphs
1. proclitics: [s]~[sa] 'with', [k]~[ka] 'towards',
2. suffix: [-ɔːg]~[-ɔːga] 'gen.sg'

    (1)   a. [k prɔmjɛni]   'towards the change' (24%)
                b. [ka prɔmjɛni]   'towards the change' (76%)

- this study: corpus data from `hrWaC 2.0` (100,018 bigrams, 8,339 unique lemmas)
- similar phenomena in other Slavic languages:
    1. Polish (Pająk & Baković 2010)
    2. Russian (Linzen et al. 2013)

Mobile *a*-alternation

1. phonological effects
   (markedness constraints, prosodic alignment)

2. morpheme-specificity
   (CV rates vary across morphemes: s~sa 'with' vs. k~ka 'towards')

3. lexical exceptions
   (CV rates vary across lemmas)

phonological factors

By-Lemma CV Proportion vs Segmental Environment (son, vowel, non-velar, velar)

1. [k=ra]: unmarked
2. [k=a]: Align(ω,L,σ,L)
3. [k=ta]: *k+obstruent
4. [k=ka]: NoGeminate

phonological factors: logistic regression

```
brm(realization ~ environment + (1 | lemma))
```

Environment (baseline: sonorant)	Estimate	l-95% CI	u-95% CI
(Intercept)	-0.39	-.69	-0.10
vowel	0.36	0.28	0.43
non-velar obstruent	0.23	0.17	0.29
velar obstruent	1.67	1.54	1.80

phonological factors: missing comparisons

```
emmeans(model, ~ environment, type = "response")
```

Comparison	Odds Ratio	95% CI
vowel vs. non-velar obs	1.14	[1.06, 1.21]
vowel vs. **velar obs**	0.27	[0.24, 0.31]
non-velar obs vs. **velar obs**	0.24	[0.21, 0.27]

<u>velar obs</u> > vowel > non-velar obs > sonorant

antigemination environments:
which velars?

- strong avoidance of [k] before velars
- 3 velars in Croatian: k, g, x
- [k]-avoidance in [k=k], /k=g/ → [g=g] expected
- **no assimilation in [k=x]**
- **no avoidance expected before x!**
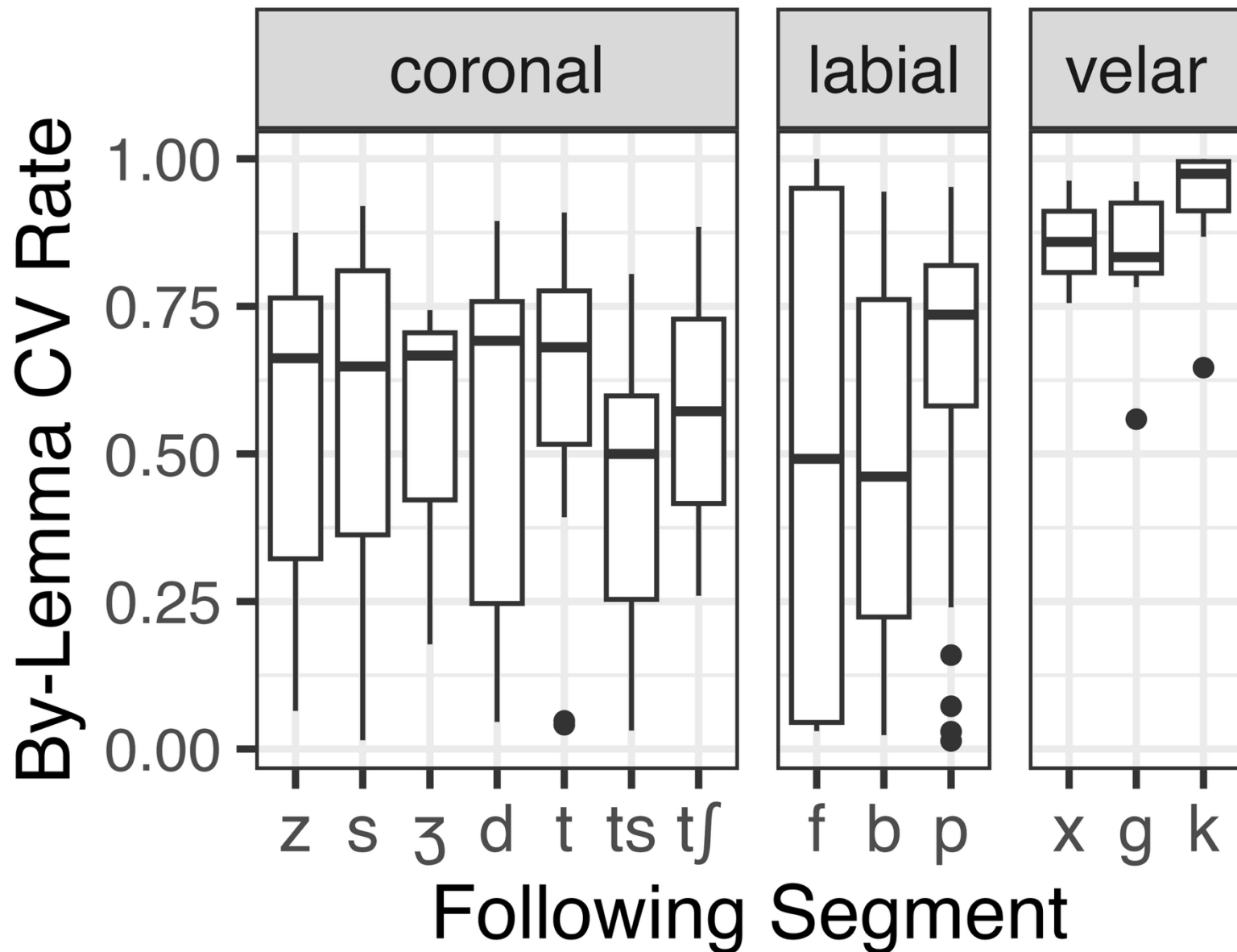
antigemination environments:
which velars?

- C~CV before obstruent-initial words
  (60,722 bigrams, 4,699 lemmas)
- `brm(realization ~ initial_seg + (1 | lemma))`
- `initial_seg:` reference level set to "x"
- **negative** effects for all non-velars relative to x
  (non-velars *less likely* than [x] to trigger CV realization)
- **positive** effects for velar stops relative to x
  (*more* likely than x to trigger CV realization)

{labials, coronals} < x < {k,g}

taking stock

- **Finding**: similarity avoidance in a **non-assimilating cluster** (k=x) in corpus data, BUT
- less so than in assimilating clusters (k=k, k=g)

alternative: phonotactics

- alternative: [k=x] is avoided because it's phonotactically undesirable
- Indeed, [kx] is impossible word-initially and severely underrepresented word-internally
- other k-obstruent clusters are avoided as well
- [k=x] still avoided more robustly than other phonotactically banned clusters
- **place cooccurrence effect**

maxent analysis

- variation; subtle, gradient phonological effects on CV
- maxent Harmonic Grammar: Harmony → probability
  (Johnson & Goldwater 2003; Hayes & Wilson 2008)
- superior to other stochastic frameworks
  (Zuraw & Hayes 2017; Smith & Pater 2020; Flemming 2021; Hayes 2022)
- `maxent.ot` R package (Mayer et al. 2024)
- two models compared:
  1. baseline
  2. OCP
- goal: determine whether OCP-place is needed

baseline model

constraint	definition	type
*Mobile-a	Penalizes [ka]	markedness
BeMoraic	Penalizes [k]	markedness
NoGeminate	Penalizes [k=k], [g=g]	markedness
*StopObs	Penalizes [k=t], [k=s], [k=k] etc.	markedness
Align(ω,L,σ,L)	Penalizes [k=a], [k=i], [k=u] etc.	markedness
Agree-voice	Penalizes [k=b], [k=d], [k=g] etc.	markedness
Ident-voice	Penalizes /k=d/ → [g=d] etc.	faithfulness
*Hiatus	Penalizes [ka=a], [ka=i] etc.	markedness

maxent analysis

- OCP model: Baseline constraints + OCP-dorsal

  (2)   OCP-dorsal
        Assess a violation for every instance of adjacent dorsal obstruents.

- **observed probability**: sum of by-lemma candidate rates by environment

- *nebo* 'sky': 31% [ka], 69% [k]; *nov* 'new': 61% [ka], 39% [k]

- k=n: .69 + .39 = 1.08; ka=n: .31 + .61 = .92

- rationale: suppress the contribution of high-frequency lemmas (*token-heavy model)

model comparison

- OCP-dorsal significantly improves model fit:
  $\chi^2(1) = 5.52, p < .05$ *

| metric	score	interpretation
Δloglik $\text{loglik}_{\text{Baseline}} - \text{loglik}_{\text{OCP}}$	2.76	> 0: OCP improves model fit
ΔAIC $\text{AIC}_{\text{Baseline}} - \text{AIC}_{\text{OCP}}$	-3.52	< −2: OCP significantly improves model fit
ΔAICc $\text{AICc}_{\text{Baseline}} - \text{AICc}_{\text{OCP}}$	-3.51	< −2: OCP significantly improves model fit

desirable prediction: x vs. {k, g} contrast

- OCP model not only captures the avoidance of [k=x], but also explains why avoidance before [x] is weaker than before {k, g}

- propensity for [ka] (ascending):

sonorant < non-velar obs < velar fricative < velar stop

- cumulativity (Jäger & Rosenbach 2006; Pater 2009):
  - [k=x] only violates OCP-dorsal
  - [k=k], [g=g] violate both OCP-dorsal and NoGeminate

deriving the scale

k=r ⊂ k=x ⊂ k=k
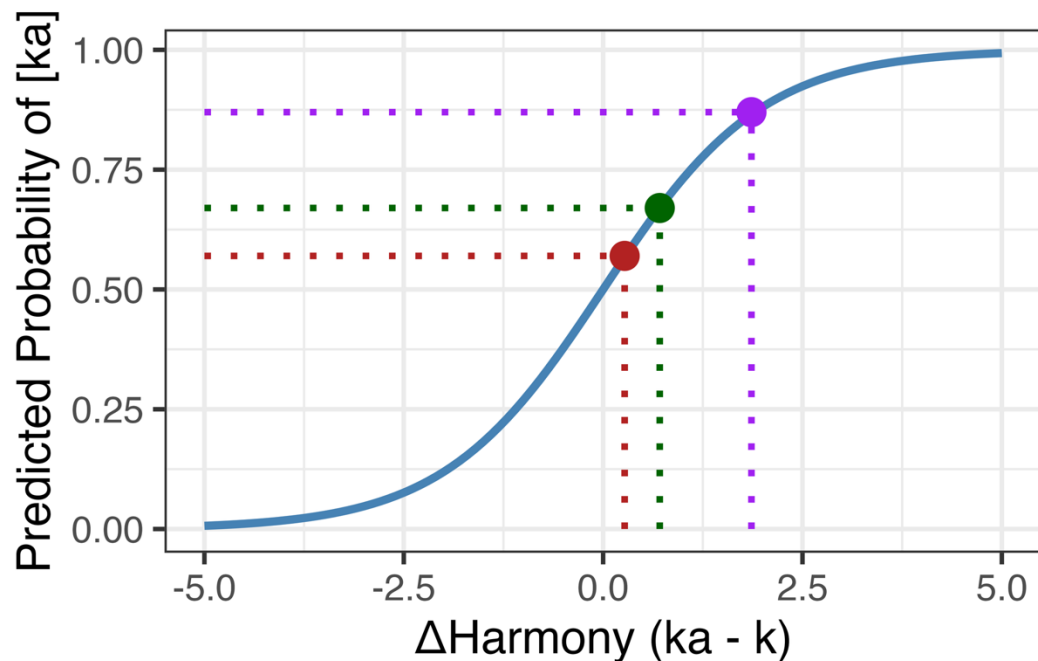
/k~ka=r/	*α	HAVE-μ	OCP-dor	NoGEM			
	1.12	1.39	.44	1.15	H	eH	p
a. k=r		−1			−1.39	.25	.43
b. ka=r	−1				−1.12	.33	**.57**

/k~ka=x/	*α	HAVE-μ	OCP-dor	NoGEM			
	1.12	1.39	.44	1.15	H	eH	p
a. k=x		−1	−1		−1.83	.16	.33
b. ka=x	−1				−1.12	.33	**.67**

/k~ka=k/	*α	HAVE-μ	OCP-dor	NoGEM			
	1.12	1.39	.44	1.15	H	eH	p
a. k=k		−1	−1	−1	−2.98	.05	.13
b. ka=k	−1				−1.12	.33	**.87**

more constraints, simpler theory!?

1. Baković (2005): assimilation-antigemination correlation
   essentially borne out in Croatian

2. [k=x] avoidance driven by OCP-place
   Croatian data suggest more constraints are needed

3. more constraints = consistent constraint set
   unifies the OT/HG accounts of phonotactic and procedural antigemination

4. inductionist approach (Hayes & Wilson 2008)
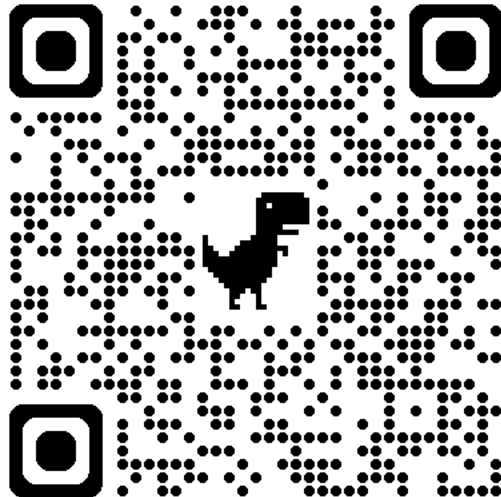   OCP-place induced as needed (language-specific)

further steps

- ongoing:
  experiment w/ nonce words

- future:
  1. full model
     (including other
     mobile *a*-morphemes)
  2. lexical exceptions
     (Nazarov & Smith 2022 learner)

Thank you!

aljosamilenkovic@g.harvard.edu
amilenkovic@umass.edu

Replication data:

appendix: results of the obstruent-only model (fixed effects)

Environment (baseline: x)	Estimate	l-95% CI	u-95% CI
Intercept	0.34	-0.12	0.81
tʃ	-0.41	-0.79	-0.02
ʒ	-0.84	-1.19	-0.49
ʃ	-0.41	-0.77	-0.05
b	-0.73	-0.99	-0.47
ts	-0.52	-0.85	-0.18
d	-0.42	-0.67	-0.16
f	-0.57	-0.87	-0.26
g	0.5	0.21	0.8
k	2.24	1.89	2.61
p	-0.34	-0.59	-0.1
s	-0.35	-0.59	-0.11
t	-0.5	-0.77	-0.24
z	-0.3	-0.57	-0.04