A maximum entropy model of phonological typology

Gašper Beguš
Harvard University
begus@fas.harvard.edu

HARVARD UNIVERSITY

NELS 48
University of Iceland
Oct 27–29, 2017

Outline

One of the most contested debates in phonology concerns identifying factors that affect typology. The **Analytic Bias** approach (AB) claims that biases in learning affect the typology, while the **Channel Bias** approach (CB) assumes phonetic precursors and directionality of sound change affect the typology (Moreton 2008). Empirical evidence in favor of both hypotheses exists:

- processes that are typologically rare have been shown to be underlearned;
- processes that are the result of phonologized phonetically motivated sound changes are also typologically frequent.

An increasing body of work acknowledges both influences (Moreton 2008), but few attempts have been made to model them together or try to disambiguate the two. This paper aims to fill this gap and proposes a model that unifies the two influences as well as provides grounds for disambiguating AB and CB.

Introduction

- A new model of typology within CB
- A subdivision of naturalness
  - **Natural**: phonetically motivated
  - **Unmotivated**: lack phonetic motivations
  - **Unnatural**: operate against universal phonetic tendencies

Unnatural alternations & sound changes		
Post-nasal devoicing	Yaghnobi	Xromov (1972)
	Tswana and Shekgalagari	Solé et al. (2010)
	Makhuwa and Bube	Janson (1991/1992)
		Janssens (1993)
	Konyagi	Merill (2016)
	Sicilian and Calabrian	Rohlfs (1949)
	Murik, Buginese	Blust (2013)
Unnatural phonotactics		
Intervocalic devoicing	Berawan (and Kiput)	Blust (2005,2013)
Post-obstruent stop voicing	Tarma Quechua	Adelaar (1977)

Historical Probabilities

- A new model for explaining unnatural phenomena: *Blurring Process*
  a. A set of segments enters complementary distribution
  b. A sound change occurs that operates on the changed/unchanged subset of those segments
  c. Another sound change occurs that blurs the original complementary distribution
- All cases of unnatural processes explained by *Blurring Process*

Blurring Cycle	*Blurring Chain*
B > C / −X	B > C / X
B > A	C > D
C > B	D > A
B > A / X	B > A / X

Minimal Sound Change Requirement (MSCR)

Minimally three sound changes have to operate in combination for an unnatural process to arise. Minimally two sound changes have to operate in combination for an unmotivated process to arise.

Stage	1. >	2. >	3. >	4.
ϕ_1	+	+	−	−
ϕ_2	+	$-/\varnothing$	$-/\varnothing$	+

How does the unnatural change in the direction $+\phi_1 > -\phi_1$ arise? According to the definition, sound change cannot produce $+\phi_1 > -\phi_1$ in a single step, given the constant value of ϕ_2. The change from $+\phi_1 > -\phi_1$, however, *can* be phonetically motivated with different values of ϕ_2 (e.g. in a Blurring Chain) or when $[+\phi_1, +\phi_2]$ only appears in a given environment, which means the context becomes irrelevant ($\varnothing$) for evaluating naturalness (as is the case in the Blurring Cycle). In other words, we cannot change $[+\phi_1, +\phi_2]$ to $[-\phi_1, +\phi_2]$, but it is possible that $+\phi_1 > -\phi_1$ is motivated under $-\phi_2$ or $\varnothing$ (motivated under a different context). This means that first, a sound change that targets ϕ_2 has to operate and change its value, either in a given environment X (Blurring Chain) or in the elsewhere condition −X (Blurring Cycle). Under the changed ϕ_2, the $+\phi_1 > -\phi_1$ can be motivated (which is the second sound change in the Blurring Process). Finally, in order for the change $+\phi_1 > -\phi_1$ to appear unnatural, the value of ϕ_2 has to change to the initial stage (the third sound change)

Historical Probabilities of Alternations (P_χ)

The probability that an alternation arises based on the **number of sound changes required (MSCR)** and their **respective probabilities** that can be estimated from samples of sound changes.

Bootstrapping Sound Changes (BSC)

- Historical Probability of a process can be estimated with a statistical technique *stratified non-parametric bootstrap* (Efron 1979)

$$P_\chi(S_i) = \frac{\text{number of languages with sound change } S_i}{\text{number of languages surveyed}}$$

$$P_\chi(T_j) = \frac{\prod_{i=1}^{n} P_\chi(S_i)}{n!}$$

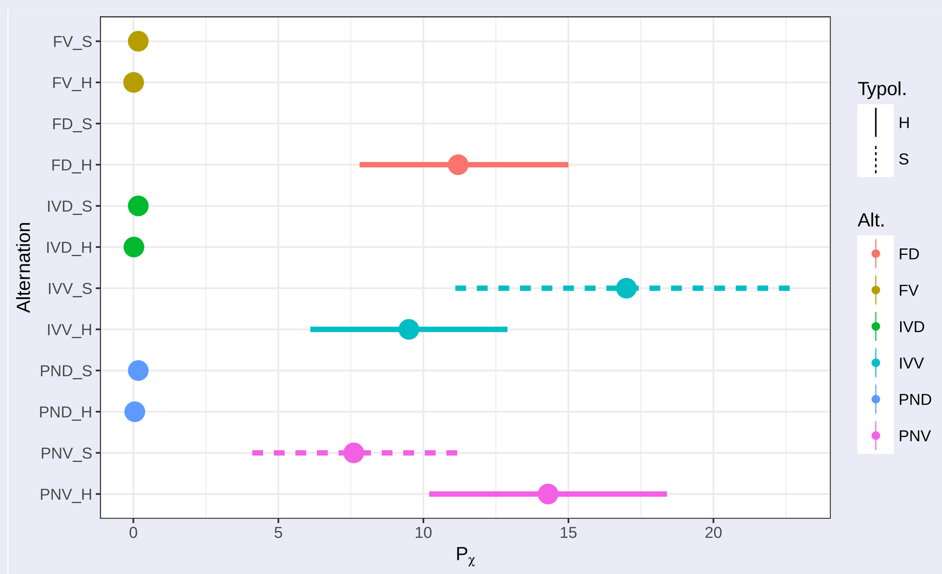
- Estimates of P_χ with 95% adjusted bootstrap percentile (BC_a) intervals were bootstrapped using the *boot* package (Canty and Ripley 2016, Davison and Hinkley 1997) in R (R Core Team).

- Blurring Process identified for PND, FV, and PNV; counts from Kümmel (2007)

Alternation	Sound change	Count	Surveyed
PND	D > Z / [−nas]/V__(V)	97	294
	D > T	18	263
	Z > D	27	216
IVD	D > Z / V__ (V)	83	294
	Z > S	7	216
	S > T	34	248
FV	T: > T / __#	6	294
	T > D / V__	32	294
	T: > T	27	≈88

- *Bootstrapped Historical Probabilities*

		95% BC_a CI		95% Profile CI	
A_k	P_χ	Lower	Upper	Lower	Upper
PNV	14.3	10.2	18.4	10.6	18.6
PND	0.05	0.02	0.09	—	
IVV	9.5	6.1	12.9	6.5	13.2
IVD	0.02	0.008	0.05	—	
FD	11.2	7.8	15.0	8.0	15.2
FV	0.01	0.004	0.03	—	



Implications of BSC

- Estimate P_χ for any synchronic process
- Compare P_χ of two processes (attested or unattested) and perform inferential statistics on the comparison $P_\chi(\text{PND}) - P_\chi(\text{FV})$: $95\%BC_a\ [0.01\%, 0.07\%]$; significantly different with $\alpha = 0.05$
- Predict attestedness of a given process
- Use P_χ for a combined model of typology

Analytic Bias

- Structurally complex alternations are more difficult to learn(Moreton and Pater 2012)
- For naturalness opposing results reported, considerably less consistency
- MaxEnt encodes learnability differences: Wilson (2006), Do and Albright (2016) report significant differences in learnability when structural complexity is controlled for
- Others do not: (Pycha et al. 2003, Wilson 2003, Kuo 2009, Skoruppa and Peperkamp 2011, via Moreton and Pater 2012a,b, Seidl et al. 2007, Do et al. 2016, Glewwe 2017)
- Wilson (2006) differentiates variance (σ^2), while White (2017) differentiates weights (μ) to encode differences in learnability

A New Proposal

- Both AB and CB affect typology (Moreton 2008)
- Most models exclude one of the two factors: rare attempts to model both AB and CB together
- Unnatural alternations are considerably rarer (7.6% vs. approximately 0.2% for PNV vs. PND, based on Locke, 1983, reported in Hayes and Stivers, 2000, and our survey)
- Artificial grammar learning experiments suggest that the two processes, PNV and PND, are equally learnable (Do et al. 2016)

Historical Weigths

- To encode that some processes are rare due to the number of sound changes they require and their respective probabilities (CB), we introduce *Historical Weights* (w_χ) of different constraints
- Differences in Historical Weights between different faithfulness and markedness constraints can be directly derived from estimated Historical Probabilities using the BSC

$$\Delta w_\chi(\mathcal{M}, \mathcal{F}) = -\log\left(\frac{P_\chi(A_k)}{1 - P_\chi(A_k)}\right).$$

- Historical Weights:
  Δw_χ(*NT, Ident-IO(voi)) = 1.79
  Δw_χ(*ND, Ident-IO(voi)) = 7.66
- Comparison:

/NT/	Ident-IO $w_\chi = 10$	*NT $w_\chi = 8.21$	$\mathcal{H}$	P	Typol.
a. [NT]		−1	−8.21	.86	.924
b. [ND]	−1		−10	.14	.076
/ND/	Ident-IO $w_\chi = 10$	*ND $w_\chi = 2.34$	$\mathcal{H}$	P	Typol.
a. [ND]		−1	−2.34	.99953	≈ .9983
b. [NT]	−1		−10	.00047	≈ .0017

- Learners have no access to Historical Weights: when we model phonological learning, Historical Weights should be disregarded completely; only prior variance should determine differences in learning of different processes (in Wilson's 2006 terms). However, when we model typology, both prior variance and Historical Weights should affect the outcome.
- Identify AB-CB mismatches

Sound change	Alternation	P_χ	Lo.	Up.	Features	$P_{complex}$	P_χ
	No alternation	67.0	61.6	72.4	0		
D > Z / [−nas]__	D → Z / [−nas]__	33.0	27.6	38.4	1	↓	↓
D > T	Z → T / [+nas]__	1.1	0.7	1.8	2	↓	↓
Z > D	PND	0.05	0.03	0.09	1	↑	↓

Conclusions

Introducing Historical Weights into a typological model within the MaxEnt framework derives differences in the observed typology that would be left unexplained if only the AB influences would be admitted to the model. We adopt the MaxEnt model of learning and at the same time predict typology for processes that show no learnability differences. The combined typological model thus encodes that PND and PNV are equally learnable, but one has a higher Historical Probability. Both are derivable, but PND is correctly predicted to be considerably less frequent. Historical Probabilities closely match the observed typology (based on the survey in Locke (1983) and my estimates of typological frequency of PND).

For every synchronic alternation and its typological distribution, we now *can* and *should* calculate its Historical Probability (CB part) based on the proposed BSC, and its σ^2 (AB part) which should ideally be calculated on the basis of learnability experiments.

Future directions

- The paper leaves open the question of what determines the prior variance of different constraints and how exactly it affects the typology
- Disambiguation of AB and CB

* I would like to thank Kevin Ryan, Jay Jasanoff, and Adam Albright for their useful comments. All mistakes are my own. This research is supported by Mind Brain Behavior Interfaculty Initiative at Harvard University.

Poster: `goo.gl/fAY9BJ` References: `goo.gl/KxUPz9` (case-sensitive)