A Formal Model of Phonological Typology

Gašper Beguš

Harvard University
begus@fas.harvard.edu

35th West Coast Conference on Formal Linguistics
University of Calgary, Alberta, CA
April 28, 2017

Introduction

- **Phonological typology** — contentious
- Two approaches: (Moreton 2008)
  - Analytic bias
  - Channel bias

Introduction

- **Phonological typology** — contentious
- Two approaches: (Moreton 2008)
  - Analytic bias
  - Channel bias
- Empirical evidence in favor of both hypotheses exists

Introduction

- **Phonological typology** — contentious
- Two approaches: (Moreton 2008)
  - Analytic bias
  - Channel bias
- Empirical evidence in favor of both hypotheses exists
  - Processes that are typologically rare have been shown to be underlearned/require more input data to be learnt
  - Processes that are the result of phonologized phonetically motivated sound changes are also typologically frequent
- A mounting body of research acknowledges both AB and CB (Moreton 2008)
- *But*: very few attempts have been made to model AB and CB together or try to disambiguate the two

Aims

- Propose a new model of typology that admits both influences
- MaxEnt probability distribution over candidates
- Quantify CB and AB
- Propose a new model of typology within CB
- Provide grounds for disambiguating the two

Outline

1 Typology within CB

2 Analytic bias

3 A formal model of typology

4 Discussion and future directions

Outline

1 Typology within CB

2 Analytic bias

3 A formal model of typology

4 Discussion and future directions

Outline

1. Typology within CB
2. Analytic bias
3. A formal model of typology
4. Discussion and future directions

Outline

1 Typology within CB

2 Analytic bias

3 A formal model of typology

4 Discussion and future directions

Previous research

- Typology under CB not satisfactory: Rare alternations are infrequent because they are produced by rare sound changes (Blevins 2004)
- Kiparsky (2006, 2008) lists **several diachronic scenarios** — combinations of sound changes — that would yield final voicing, yet FV is never attested
- Against CHANNEL BIAS: CB wrongly predicts the typology
- Other models fails to yield directly implementable results (Moreton 2008, cf. Yu 2011, Cathcart 2015)

Historical Probabilities

- A new typology within CB

Historical Probabilities of Alternations **(P_χ(Alt))**

The probability that an alternation arises based on the number of sound changes required (MSCR) and their respective probabilities/rates.

Historical Probabilities

- A new typology within CB

Historical Probabilities of Alternations (P_χ(Alt))

The probability that an alternation arises based on the **number of sound changes required** (MSCR) and their respective probabilities/rates.

Historical Probabilities

- A new typology within CB

Historical Probabilities of Alternations (P_χ(Alt))

The probability that an alternation arises based on the number of sound changes required (MSCR) and their respective **probabilities**/rates.

On naturalness

- Number of sound changes?

On naturalness

- **Natural**: universal phonetic tendencies
- **Unmotivated**: lack phonetic motivations, but not against UPT
- **Unnatural**: operating in the opposite direction from UPT

- *Definition of Universal phonetic tendency (UPT)*
  UPTs are phonetic processes motivated by articulatory (or perceptual) mechanisms that passively and universally operate in speech production and are typologically common.

Post-Nasal Devoicing

- Unnatural alternations and phonotactic restrictions

Post-Nasal Devoicing

- Unnatural alternations and phonotactic restrictions
  - Yaghnobi (Xromov 1972)
  - Tswana and Shekgalagari (Solé *et al.* 2010)
  - Makhuwa and Bube (Janson 1991/1992, Janssens 1993)
  - Konyagi (Merrill 2015)
  - Sicilian and Calabrian (south Italian dial.) (Rohlfs 1949)
  - Murik, Buginese, and Land Dayak (Austronesian) (Blust 2009)

  - Intervocalic devoicing: Berawan and Kiput (Blust 2005, 2013)
  - Voicing after voiceless obstruents in Tarma Quechua (Adelaar 1977, Beguš and Nazarov 2017)

Blurring Process

- BLURRING PROCESS
- A historical model for explaining unnatural phenomena (B $>$ A / X)
  - (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

Blurring Process

- BLURRING PROCESS
- A historical model for explaining unnatural phenomena (B > A / X)
  - (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

BLURRING CYCLE

B > C / –X

B > A

C > B

B > A / X

Blurring Process

- BLURRING PROCESS
- A historical model for explaining unnatural phenomena (B $>$ A / X)
  (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

BLURRING CYCLE	BLURRING CHAIN
B $>$ C / $-$X	B $>$ C / X
B $>$ A	C $>$ D
C $>$ B	D $>$ A
B $>$ A / X	B $>$ A / X

Blurring Process

- Post-nasal devoicing (Dickens 1984), Hyman (2001)
- In all twelve cases in which PND is reported as a sound change or synchronic alternation, it arises though a combination of 3 sound changes
-

Blurring Cycle		**Example**		
		Avestan	**band**	**dasa**
D > Z / [−nas]	d > ð/[−nas]__	Sogdian	βand	ðasa
D > T	d > t	Yaghnobi	vant	*ðasa
Z > D	ð > d	Yaghnobi	vant	das

Blurring Process

- Post-nasal devoicing (Dickens 1984), Hyman (2001)
- In all twelve cases in which PND is reported as a sound change or synchronic alternation, it arises though a combination of 3 sound changes
-

Blurring Cycle		**Example**		
		Avestan	band	dasa
D > Z / [−nas]	d > ð/[−nas]__	Sogdian	**βand**	**ðasa**
D > T	d > t	Yaghnobi	vant	*ðasa
Z > D	ð > d	Yaghnobi	vant	das

Blurring Process

- Post-nasal devoicing (Dickens 1984), Hyman (2001)
- In all twelve cases in which PND is reported as a sound change or synchronic alternation, it arises though a combination of 3 sound changes
-

Blurring Cycle		**Example**		
		Avestan	band	dasa
D > Z / [−nas]	d > ð/[−nas]__	Sogdian	βand	ðasa
D > T	d > t	Yaghnobi	**vant**	***ðasa**
Z > D	ð > d	Yaghnobi	vant	das

Blurring Process

- Post-nasal devoicing (Dickens 1984), Hyman (2001)
- In all twelve cases in which PND is reported as a sound change or synchronic alternation, it arises though a combination of 3 sound changes
-

Blurring Cycle		**Example**		
		Avestan	band	dasa
D > Z / [−nas]	d > ð/[−nas]__	Sogdian	βand	ðasa
D > T	d > t	Yaghnobi	vant	*ðasa
Z > D	ð > d	Yaghnobi	**vant**	**das**

Minimal number of sound changes

- "Telescoping", concatenation of sound changes explains unmotivated processes (Wang 1968, Hyman 2001)
- For UNNATURAL processes, we need a special scenario ("Blurring Process")

Minimal number of sound changes

Minimal Sound Change Requirement (MSCR; Beguš 2016)

Natural processes arise through a single sound change. Minimally two sound changes have to operate in combination for an unmotivated process to arise. Minimally three sound changes have to operate in combination for an unnatural process to arise.

- *A scale of decreased probabilities*
  $P_\chi(\text{natural}) < P_\chi(\text{unmotivated}) < P_\chi(\text{unnatural})$

Minimal number of sound changes

- Sound change: a change in one feature in a given environment (Picard 1994)
- B > A / X

Minimal number of sound changes

- Sound change: a change in one feature in a given environment (Picard 1994)
- B $>$ A / X
- B $>$ A / X by definition impossible
- Two sound change? Let B change to C, where B and C differ in one feature, but, to be sure, a different feature from the one that separates A and B. From this point, it is impossible for an unnatural sound change to arise without a third sound change: C cannot develop directly to A, since the two segments differ in two features: feature F_1, which distinguishes A and B, and feature F_2, which distinguishes B and C.
- By definition, two sound changes are required in order to change two features $\rightarrow$ unnatural processes require altogether at least three sound changes.

Probabilities

Historical Probabilities of Alternations (P_χ(Alt))

The probability that an alternation arises based on the number of sound changes required (MSCR) and their respective probabilities/rates.

Probabilities

Historical Probabilities of Alternations (P_χ(Alt))

The probability that an alternation arises based on the **number of sound changes required** (MSCR) and their respective probabilities/rates.

Probabilities

Historical Probabilities of Alternations (P_χ(Alt))

The probability that an alternation arises based on the number of sound changes required (MSCR) and their respective **probabilities**/rates.

Probabilities

Historical Probabilities of Alternations (P_χ(Alt))

The probability that an alternation arises based on the number of sound changes required (MSCR) and their respective probabilities/rates.

- Probabilities/rates?

Probability of a sound change

- **Historical probabilities**: BOOTSTRAPPING SOUND CHANGES (BSC)
- Estimates of P_χ not a trivial task: we propose to estimate sound change probabilities using bootstrapping (Efron 1979): a sample of successes (languages in the sample with a sound change S_1) and failures (languages in the sample without S_1)
- Probability of a sound change A_1

$$P_\chi(A_1) = \frac{\text{number of languages with sound change } A_1}{\text{number of languages surveyed}}$$

Historical Probabilities

- If an alternation A_x requires $n > 1$ sound changes to arise ($n \geq 2$ for unmotivated, $n \geq 3$ for unnatural)
- Joint probabilities of n number of sound changes

$$P(T_1) = \frac{P(A_1)P(A_2) * ... * P(A_n)}{n!}$$

- Binomial sample: P_χ is bootstrapped from a product of probabilities based on the number of successes and failures (divided by $n!$ to account for the ordering of sound changes)

Assumptions

- Assumptions: sample is representative and well-balanced
- Occurrence of sound change properly counted: occurrence in the proto-language vs. daughter languages
- Independence of sound change
  - A_1 and A_2 independent
  - $P(A_1|\text{phonemic inventory})$, **but** some of (in)dependence captured by context
  - In the current state of the field, we lack sufficiently accurate estimates of sound change probabilities to be able to estimate their conditional probabilities
  - For practical purposes, we can disregard the phonemic inventory and generalize $P(A_1/\_\_X)$
  - If two sound changes occur in two related languages, they are considered independent

Applications

- What can we do with historical probabilities?
  1. Compare two alternations and perform inferential statistics: is historical probability of alternation X significantly more frequent than Y?
  2. Predict attestedness/unattestedness in a given sample
  3. Go below "zero probabilities"
  4. Identify historically equiprobable processes
  5. Use P_χ to encode historical or Channel Bias in a typological model

Applications

- What can we do with historical probabilities?
  1. Compare two alternations and perform inferential statistics: is historical probability of alternation X significantly more frequent than Y?
  2. Predict attestedness/unattestedness in a given sample
  3. Go below "zero probabilities"
  4. Identify historically equiprobable processes
  5. **Use P_χ to encode historical or Channel Bias in a typological model**

PND vs. FV

- Post-nasal devoicing (PND) and final voicing (FV)
- Kümmel (2007) surveys approximately 200 languages
- Blurring cycle for PND:
  - 56 lang. fricativization of voiced stops
  - 11 languages unconditioned devoicing
  - 37 languages unconditioned occlusion
- Final voicing (FV) (from Kiparsky 2006):
- The only scenario in Kiparsky (2006) that results in *alternation* and involves less than four sound changes
  - 10 languages degemination
  - 45 languages post-vocalic voicing
  - 3 languages final degemination

Historical Probabilities

- Estimates of P_χ were bootstrapped using the boot package (Canty and Ripley 2016, Davison and Hinkley 1997) in R (R Core Team). The following 95% adjusted bootstrap percentile (BC_a) intervals were calculated for P_χ(PND) and P_χ(FV).
- *Bootstrapped historical probabilities*
  $P_\chi(\text{PND}) = [0.02\%, 0.11\%]$
  $P_\chi(\text{FV}) = [0.00\%, 0.02\%]$

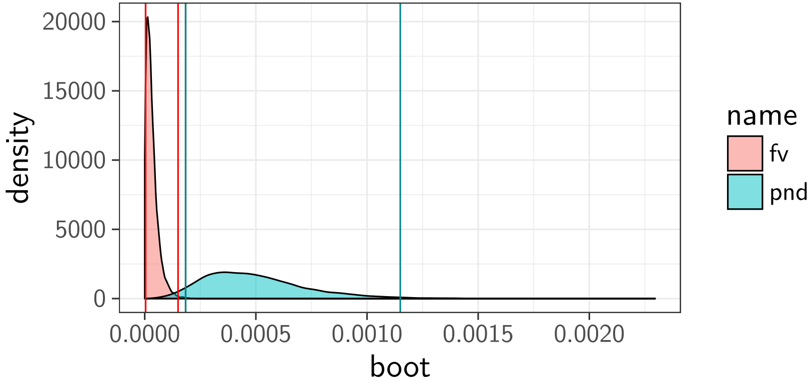
Comparison of P_χ

- We can test whether FV is significantly different from PND
- $P_\chi(\text{PND}) - P_\chi(\text{FV}) = [0.02\%, 0.11\%]$

Comparison of P_χ

- We can test whether FV is significantly different from PND
- $P_\chi(PND) - P_\chi(FV) = [0.02\%, 0.11\%]$
- Yes, $P_\chi(PND)$ is significantly higher than $P_\chi(FV)$.

Comparison of P_χ



Prediction of attestedness

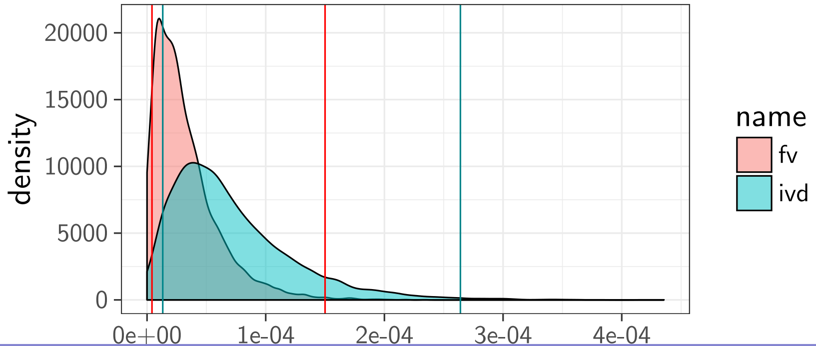
- Do we predict PND to be attested?
- Yes, difference between P_χ(PND) and P(1/200) is [-0.05%, 2.47%]
- But: P_χ(FV) and P(1/200) is [-0.003%, 2.86%]
- P_χ(FV) is not significantly smaller than probability of being attested once in the sample, but [-0.003%, 2.86%] and it is significantly different from PND

Going "below zero"

- Why not just calculate probabilities of observed sound changes?
- BSC "reaches below zero"
- Can differentiate unattested processes

Going "below zero"

- FV and intervocalic devoicing both unattested (cf. Yu 2004, recently Sula in Bloyd 2017)
- According to BSC, FV is predicted to be significantly less probable than IVD, but the difference is very small [0.001%, 0.016%]



Input to a typological model

- P_χ using BSC and MSCR as an input to a typological model: CB contribution

Outline

1 Typology within CB

2 Analytic bias

3 A formal model of typology

4 Discussion and future directions

Analytic bias

- Numerous studies experimentally confirm that some alternations are underlearned (Moreton and Pater 2012)
- These differences can be encoded in MaxEnt models of phonological learning: regularization term
- Wilson (2006) differentiates variance (σ^2) in the prior
- White (2017) differentiates weights (μ) in the prior
- Effectively, this means that some processes require mode input data to be learnt
- Both determine parameters in the prior by P-map related perceptual distance measures as independent inputs (Steriade 2001)

Analytic bias

- The evidence for AB is strongest when testing alternations that target fewer vs. more features (structural bias)
- Less robust results when testing alternations that target a single feature where one direction is phonetically natural and typologically common and the other is unnatural and rare (substantive bias) (Moreton and Pater 2012)
- Two studies specifically tested learnability of PND and IVD compared to their natural counterparts and found no effect (Seidl et al. 2007, Do et al. 2016)
- This suggests that a model that admits both AB and CB will perform better

Outline

1 Typology within CB

2 Analytic bias

3 A formal model of typology

4 Discussion and future directions

A formal model of typology

- Propose a model that models the two influences together
- Typological predictions have long been a strength of Optimality Theory (OT)
- **Primary device** to derive typological predictions: Restricted CON that admits no unnatural constraints

/ND/	*NT	IDENT-IO(voice)
a. [NT]	*!	*
☞b. [ND]		

A formal model of typology

- Classical OT *undergenerates*: PND (*ND) attested (Coetzee and Pretorius 2010)
- But so does HG (Beguš and Nazarov 2017)

HG undergenerates

Natural Gradience Bias (NGB)

HG with restricted CON predicts that the probability of the natural feature value in a given environment is always equal or grater than the probability of the unnatural value in a given environment.

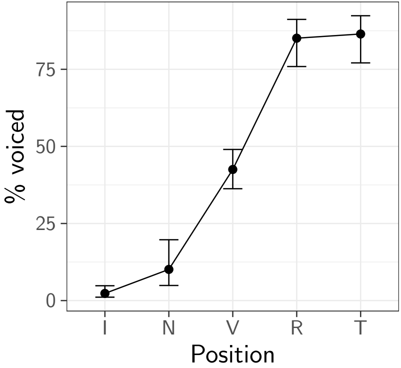
- Inputs /NAT/ and /UNNAT/ have a uniform prior probability under a Bayesian interpretation of Richness of the Base

(Beguš and Nazarov 2017)

- If we admit only natural $\mathcal{M}$ (restricted CON):
  - $w(\mathcal{F}) >> w(\mathcal{M})$: P([NAT]) = P([UNNAT]) = 0.5
  - $w(\mathcal{M}) > w(\mathcal{F})$: P([NAT]) > P([UNNAT])

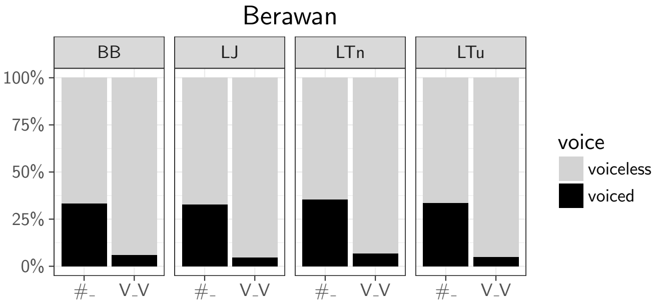
Tarma Quechua

- *N[+voice], *T[-voice] (Beguš and Nazarov 2017)



Berawan

- *V[+voice]V (Beguš and Nazarov 2017)



A formal model of typology

- To allow all constraints into CON is not a solution: severe overgeneration
- Different σ^2 (Wilson 2006) fails to encode rarity of PND, FV, IVD (where no differences in learnability or P-map)
- A model of typology should
  - Generate all attested alternations
  - Encode that some alternations are less frequent than others
  - Encode why some alternations are dispreferred

A formal model of typology

- A new model admits all constraints into CON
- Encodes historical probabilities (CB) and learnability differences (AB)
- MaxEnt probability distribution over candidates

A formal model of typology

- w_χ (CB) and σ^2 (AB)
- σ^2 adopted from Wilson (2006)

$$\Delta w_\chi = -\log\left(\frac{P_\chi}{1 - P_\chi}\right)$$

- Crucially, when modeling acquisition, speakers have no access to historical weights or probabilities
- When modeling typology, both AB and CB contribute to the typology
- Generates all attested patterns, but encodes that some are rare (due to historical probabilities and learnability)

Disambiguation

- New model performs better:
- No difference in learnability experiments between PND and PNV (Do et al. 2016)
- σ^2 of PND and PNV should be equal

Disambiguation

- New model performs better:
- No difference in learnability experiments between PND and PNV (Do et al. 2016)
- σ^2 of PND and PNV should be equal (also under symmetric P-map approach)

Disambiguation

- New model performs better:
- No difference in learnability experiments between PND and PNV (Do et al. 2016)
- σ^2 of PND and PNV should be equal (also under symmetric P-map approach)
- $\Delta w_\chi(\text{PND}) = 7.66$; $\Delta w_\chi(\text{PNV}) = 1.36$
- Typologically, clear difference
  - 15 languages of 197 surveyed feature PNV as alternation (Locke 1983)
  - Only one case of PND (Coetzee and Pretorius 2010) in all surveys available to me

A formal model of typology

/NT/	IDENT-IO $w_\chi = 10$	*NT $w_\chi = 8.64$	H_χ	P_χ	Typol.
a. [NT]		-1	-10	.795	.924
b. [ND]	-1		-8.64	.205	.076

/ND/	IDENT-IO $w_\chi = 10$	*ND $w_\chi = 2.34$	H_χ	P_χ	Typol.
a. [ND]	-1		-10	.99953	$\approx .9975$
b. [NT]		-1	-2.34	.00047	$\approx .0025$

Disambiguation

- The model provides grounds for disambiguation
- We now *can* and *should* calculate the CB contribution (based on BSC) and AB contribution (based on learnability experiments)
- Overlap: typologically rare processes more difficult to learn *and* not the result of common sound change
- AB might influence CB and historical probabilities if it is assumed that sound change is primarily influenced by learning
- This "duplication" problem is avoided in the case of unnatural processes: we argue that a **combination of at least three sound changes** is required for any unnatural alternation (MSCR)
- The fact that sound changes need to operate in combination means that CB (transmission) has to independently influence historical probabilities of unnatural processes

Outline

1 Typology within CB

2 Analytic bias

3 A formal model of typology

4 Discussion and future directions

Future directions

- Too Many / Too Few Solution problem
- P-map (Steriade 2001)
- *D# $\gg$ IDENT-IO(voi), IDENT-IO(nas)
- *NT $\gg$ IDENT-IO(voi), IDENT-IO(nas), MAX (Pater 1999)
- Rarity encoded by σ^2 (Wilson 2006, cf. White 2017)
- But also w_χ: evidence that, although not unnatural, final nasalization arises through a combination of sound changes (Beguš and Nazarov 2017)
- Noon (Merrill 2015), Austronesian (Blust 2005, 2016): blurring process
- Δw(*D#, IDENT-IO(voi)) $<$ Δw(*D#, IDENT-IO(nas))
- Further learnability experiments are needed (cf. Albright and Do 2017)

Future directions

- The primary **role of AB** to reduce alternations?
  - Dispreference for alternations is a robust result in artificial grammar learning experiments
  - Future directions: identify historically equiprobable processes and test learnability differences that result in typology
- Modeling **phonological learning**:
  - Determining variance σ^2: P-map vs. structural complexity
  - Are natural constraints innate and unnatural constructed during acquisition (cf. Hayes 1999)? Does this affect σ^2?

References

Adelaar, Willem F. H. 1977. *Tarma Quechua: Grammar, texts, dictionary.* Ph. D. dissertation, University of Amsterdam.

Albright, Adam and Youngah Do. 2017.A substantive bias for perceptually minimal alternations in Artificial Grammar learning. Presentation at *14th Old World Conference on Phonology* in Düsseldorf, Germany on February 20-22, 2017.

Beguš, Gašper. 2016. *Post-Nasal Devoicing and a Probabilistic Model of Phonological Typology*. Ms., Harvard University.

Beguš, Gašper and Aleksei Nazarov. 2017. *Lexicon against naturalness: Unnatural gradient phonotactic restrictions and their origins*. Ms., Harvard University.

Blevins, Juliette. 2004. *Evolutionary Phonology.* Cambridge: CUP.

Bloyd, Tobias. 2017. Synchronic intervocalic fortition in Sula: a counter-universal. Presentation at *LSA 2017 Annual Meeting* in Austin, TX on January 5-8, 2017.

Blust, Robert. 2005. Must sound change be linguistically motivated? *Diachronica* 22 (2): 219–269.

References

Blust, Robert. 2013. *The Austronesian Languages*. Canberra: Asia-Pacific Linguistics.

Blust, Robert. 2016. Austronesian against the world: where the P-map ends. Presentation at the *42nd Berkeley Linguistic Society* in Berkeley, CA on February 5-6, 2016.

Canty, Angelo and Brian Ripley. 2016. *boot: Bootstrap R (S-Plus) Functions*. R package version 1.3-18.

Cathcart, Chundra. 2015. *NELS 45: Proceedings of the Forty-Fifth Annual Meeting of the North East Linguistic Society*, edited by Thuy Bui and Deniz zyldz. Volume 1. 145-150. Amherst: GLSA.

Coetzee, Andries W. and Rigardt Pretorius. 2010. Phonetically grounded phonology and sound change: The case of Tswana labial plosives. *Journal of Phonetics* 38: 404-421.

Davison, A. C. and D. V. Hinkley. 1997. *Bootstrap Methods and Their Applications.* Cambridge: CUP.

Dickens, Patrick J. 1984. The History of So-Called Strengthening in Tswana. *Journal of African Languages and Linguistics* 6:97–125.

References

Do, Youngah, Elizabeth Zsiga, and Jonathan Havenhill. 2016. Naturalness and frequency in implicit phonological learning. Talk presented at the 90th *Annual Meeting of the Linguistic Society of America*, Washington, DC, January 7-10.

Efron, Bradley. 1979. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* 7(1): 1-26.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly (eds.), *Functionalism and Formalism in Linguistics*, Volume I: General Papers, pp. 243-285. Amsterdam: John Benjamins, Amsterdam.

Hyman, Larry M. 2001. The Limits of Phonetic Determinism in Phonology: *NC Revisited. In *The Role of Speech Perception in Phonology*, edited by Elizabeth Hume and Keith Johnson, 141-186. San Diego, CA: Academic Press.

Janson, Tore (1991/1992). Southern Bantu and Makua. *Sprache und Geschichte in Afrika* 12/13, 1–44.

References

Janssens, Baudouin. 1993. Doubles réflexes consonantiques: quatre études sur le bantou de zone A (bubi, nen, bafia, ewondo) (Unpublished doctoral dissertation). Université libre de Bruxelles, Faculté de Philosophie et Lettres, Bruxelles.

Kiparsky, Paul. 2006. Amphichronic program vs. Evolutionary Phonology. *Theoretical Linguistics* 32(2): 217-236.

Kiparsky, Paul. 2008. Universals constrain change, change results in typological generalizations. In *Linguistic universals and language change*, edited by Jeff Good. 23-53. Oxford: Oxford University Press.

Kümmel, Martin. 2007. *Konsonantenwandel*. Wiesbaden: Reichert.

Locke, John. 1983. *Phonological Acquisition and Change*. New York: Academic Press.

Merrill, John. 2015. Nasalization as a repair for voiced obstruent codas in Noon. In *LSA Annual Meeting Extended Abstracts* 6:14-1.

Merrill, John. 2016. *A Historical Account of the Fula and Sereer Consonant Mutation and Noun Class Systems.* University of California, Berkeley, Ms.

References

Moreton, Elliott. 2008. Analytic bias and phonological typology. *Phonology* 25(1): 83127.

Moreton, E., and J. Pater. 2012. Structure and substance in artificial-phonology learning. Part I & Part II. *Language and Linguistic Compass* 6(11): 686-701 and 702-718.

Picard, Marc. 1994. *Principles and methods in historical phonology: From Proto-Algonkian to Arapaho*. Montreal: McGill-Queen's Press-MQUP.

R Core Team 2016. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rohlfs, Gerhard. 1949. *Lautlehre*. Vol. 1 of Historische Grammatik der Italienischen Sprache und ihrer Mundarten. Bern: Francke.

Seidl, Amanda, Eugene Buckley, and Alejandrina Cristia. 2007. Complexity trumps naturalness. Talk presented at the 81st *Annual Meeting of the Linguistic Society of America*, Anaheim, CA, January 4-7.

Solé, Maria-Josep, Larry M. Hyman, and Kemmonye C. Monaka. 2010. More on Post-Nasal Devoicing: The Case of Shekgalagari. *Journal of Phonetics* 38 (4): 299–319.

References

Steriade, Donca. 2001. *The Phonology of Perceptibility Effects: The P-map and its Consequences for Constraint Organization.* UCLA, Ms.

Wang, William S-Y. 1968. Vowel Features, Paired Variables, and the English Vowel Shift. *Language* 44:695-708.

Wilson, Colin. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science* 30: 945-982.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Language* 93(1): 1-36.

Xromov, Albert. 1972. *Jagnobskij jazyk*. Moskva: Nauka.

Yu, Alan C. L. 2004. Explaining final obstruent voicing in Lezgian: Phonetics and history. *Language* 80(1): 73-97.

Yu, Alan, C. L. 2011. On measuring phonetic precursor robustness: a response to Moreton. *Phonology* 28(3): 491-518.

Thank you!

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