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:  
  - **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
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** ($P_X(\text{Alt})$)

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** \((P_\chi(Alt))\)

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** ($P_\chi(\text{Alt})$)

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**

$B > C / -X$

$B > A$

$C > B$

$\underline{B > A / X}$
Blurring Process

- **Blurring Process**

- A historical model for explaining unnatural phenomena ($B > A / X$)
  1. A set of segments enters complementary distribution
  2. A sound change occurs that operates on the changed/unchanged subset of those segments
  3. Another sound change occurs that blurs the original complementary distribution

**Blurring Cycle**

- $B > C / X$
- $B > A$
- $C > B$
- $B > A / X$

**Blurring Chain**

- $B > C / X$
- $C > D$
- $D > A$
- $B > A / X$
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

<table>
<thead>
<tr>
<th>Blurring Cycle</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>D &gt; Z / [−nas] d &gt; ɒ/ [−nas]</td>
<td>Avestan *band dasa</td>
</tr>
<tr>
<td>D &gt; T</td>
<td>d &gt; t</td>
</tr>
<tr>
<td>Z &gt; D</td>
<td>ɒ &gt; d</td>
</tr>
<tr>
<td>Yaghnobi vant das</td>
<td></td>
</tr>
</tbody>
</table>

Gašper Beguš
Harvard University — begus@fas.harvard.edu
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

<table>
<thead>
<tr>
<th>Blurring Cycle</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>D &gt; Z / [–nas]</td>
<td>Avestan  band  dasa</td>
</tr>
<tr>
<td>d &gt; ū/[–nas]</td>
<td>Sogdian   band  ūasa</td>
</tr>
<tr>
<td>D &gt; T</td>
<td>Yaghnobi  vant  *ūasa</td>
</tr>
<tr>
<td>Z &gt; D</td>
<td>Yaghnobi  vant  das</td>
</tr>
<tr>
<td>Z &gt; D</td>
<td>Yaghnobi  vant  das</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Blurring Cycle</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>D &gt; Z / [−nas] d &gt; ŀ / [−nas]</td>
<td>Avestan band dasa</td>
</tr>
<tr>
<td>D &gt; T            d &gt; t</td>
<td>Sogdian βand ņasa</td>
</tr>
<tr>
<td>Z &gt; D            ņ &gt; d</td>
<td>Yaghnobi vant *ňasa</td>
</tr>
<tr>
<td></td>
<td>Yaghnobi vant das</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Blurring Cycle</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>D &gt; Z / [−nas] d &gt; ḍ/[−nas]</td>
<td>Avestan band dasa</td>
</tr>
<tr>
<td>D &gt; T           d &gt; t</td>
<td>Sogdian βand ḍasa</td>
</tr>
<tr>
<td>Z &gt; D           ḍ &gt; d</td>
<td>Yaghnobi vant *ḍasa</td>
</tr>
<tr>
<td></td>
<td>Yaghnobi vant dasa</td>
</tr>
</tbody>
</table>

Gašper Beguš
Harvard University — begus@fas.harvard.edu
A Formal Model of Phonological Typology
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* $(P_\chi(Alt))$

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** \( (P_\chi(Alt)) \)

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** $(P_\chi(Alt))$

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 $P_\chi$ 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

$$P(T_1) = \frac{P(A_1)P(A_2) \ast \ldots \ast P(A_n)}{n!}$$

- Binomial sample: $P_\chi$ 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 $P_\chi$ 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 $P_\chi$ 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 languages 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 $P_{\chi}$ 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 $P_{\chi}(PND)$ and $P_{\chi}(FV)$.

- *Bootstrapped historical probabilities*
  
  $P_{\chi}(PND) = [0.02\%, 0.11\%]$  
  $P_{\chi}(FV) = [0.00\%, 0.02\%]$
Comparison of $P_{\chi}$

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

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

- Do we predict PND to be attested?
- Yes, difference between $P_{\chi}(\text{PND})$ and $P(1/200)$ is $[-0.05\%, 2.47\%]$
- But: $P_{\chi}(\text{FV})$ and $P(1/200)$ is $[-0.003\%, 2.86\%]$
- $P_{\chi}(\text{FV})$ 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

- $P_\chi$ 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 ($\sigma^2$) in the prior
- White (2017) differentiates weights ($\mu$) 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

<table>
<thead>
<tr>
<th>/ND/</th>
<th>*NT</th>
<th>IDENT-IO(voice)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [NT]</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>☇ b. [ND]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 greater 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 $M$ (restricted CON):
  - $w(\mathcal{F}) >> w(M)$: $P([\text{NAT}]) = P([\text{UNNAT}]) = 0.5$
  - $w(M) > w(\mathcal{F})$: $P([\text{NAT}]) > P([\text{UNNAT}])$
Tarma Quechua

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

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

Berawan

<table>
<thead>
<tr>
<th>BB</th>
<th>LJ</th>
<th>LTn</th>
<th>LTu</th>
</tr>
</thead>
<tbody>
<tr>
<td>#_</td>
<td>#_</td>
<td>#_</td>
<td>#_</td>
</tr>
<tr>
<td>V_V</td>
<td>V_V</td>
<td>V_V</td>
<td>V_V</td>
</tr>
</tbody>
</table>

voice

- voiceless
- voiced
A formal model of typology

- To allow all constraints into $\text{CON}$ is not a solution: severe overgeneration
- Different $\sigma^2$ (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 $\text{CON}$
- Encodes historical probabilities (CB) and learnability differences (AB)
- MaxEnt probability distribution over candidates
A formal model of typology

- $w_\chi$ (CB) and $\sigma^2$ (AB)
- $\sigma^2$ 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)
- \( \sigma^2 \) 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)

- $\sigma^2$ 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)
- \(\sigma^2\) of PND and PNV should be equal (also under symmetric P-map approach)
- \(\Delta w_\chi(PND) = 7.66; \Delta w_\chi(PNV) = 1.36\)
- 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

<table>
<thead>
<tr>
<th></th>
<th>IDENT-IO $w_\chi = 10$</th>
<th>*NT $w_\chi = 8.64$</th>
<th>$H_\chi$</th>
<th>$P_\chi$</th>
<th>Typol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [NT]</td>
<td>−1</td>
<td>−10</td>
<td>.795</td>
<td>.924</td>
<td></td>
</tr>
<tr>
<td>b. [ND]</td>
<td>−1</td>
<td>−8.64</td>
<td>.205</td>
<td>.076</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>IDENT-IO $w_\chi = 10$</th>
<th>*ND $w_\chi = 2.34$</th>
<th>$H_\chi$</th>
<th>$P_\chi$</th>
<th>Typol.</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. [ND]</td>
<td>−1</td>
<td>−10</td>
<td>.99953</td>
<td>≈ .9975</td>
<td></td>
</tr>
<tr>
<td>b. [NT]</td>
<td>−1</td>
<td>−2.34</td>
<td>.00047</td>
<td>≈ .0025</td>
<td></td>
</tr>
</tbody>
</table>
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)
- But also $\omega_\chi$: 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
- $\Delta w(*D#, \text{IDENT-IO}(voi)) < \Delta w(*D#, \text{IDENT-IO}(nas))$
- 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 $\sigma^2$: P-map vs. structural complexity  
  - Are natural constraints innate and unnatural constructed during acquisition (cf. Hayes 1999)? Does this affect $\sigma^2$?
References


References


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.


References


References


References


References


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.