Historical Probabilities of Alternations

Gašper Beguš
Harvard University
begus@fas.harvard.edu

53rd Annual Meeting of the Chicago Linguistic Society
University of Chicago
May 26, 2017
Introduction

- **Phonological typology** — contentious
- **Two approaches:**
  - **Analytic bias**
  - **Channel bias**

(Moreton 2008)
Introduction

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

(Moreton 2008)
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

- Many acknowledge both AB and CB (Moreton 2008)
Questions

- Problems with typology within CB
- Unattestedness of unnatural alternations has been an argument in favor of the Analytic Bias approach to typology (Kiparsky 2006)
- 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
- Numerous diachronic scenarios can yield unnatural alternations; yet they are not attested
Previous research

- Typology under CB not satisfactory: Rare alternations are infrequent because they are produced by rare sound changes (Blevins 2004)

- Moreton (2008) attempts to quantify phonetic precursors, but phonetic precursors and sound change do not always align

- The most elaborate attempt thus far in Cathcart (2015) lacks MSCR, fails to discriminate alternations from static phonotactic restrictions, and is computationally too demanding and the model

- Fails to yield directly implementable results
Aims

- Propose a quantitative model of typology within Channel Bias
Historical Probabilities of Alternations \( P_X(A) \)

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

- A new typology within CB
Historical Probabilities

- A new typology within CB

**Historical Probabilities of Alternations** ($P_{\chi}(A)$)

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 (Pχ(A))**

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

1 Introduction

2 Number of sound changes

3 Probabilities

4 Examples

5 Discussion and conclusion
Outline

1. Introduction

2. Number of sound changes

3. Probabilities

4. Examples

5. Discussion and conclusion
Outline

1. Introduction
2. Number of sound changes
3. Probabilities
4. Examples
5. Discussion and conclusion
Outline

1. Introduction
2. Number of sound changes
3. Probabilities
4. Examples
5. Discussion and conclusion
Outline

1. Introduction
2. Number of sound changes
3. Probabilities
4. Examples
5. Discussion and conclusion
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.
Naturalness

- **Unnatural**: operating against universal phonetic tendency (UPT)

<table>
<thead>
<tr>
<th>natural</th>
<th>unnatural</th>
</tr>
</thead>
<tbody>
<tr>
<td>T &gt; D / N__</td>
<td>D &gt; T / N__</td>
</tr>
<tr>
<td>T &gt; D / V__V</td>
<td>D &gt; T / V__V</td>
</tr>
<tr>
<td>D &gt; T / __#</td>
<td>T &gt; D / __#</td>
</tr>
</tbody>
</table>
Models

- $A > B / X = \text{UPT}$
- How do we get $B > A / X$?
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)

- Hypercorrection
Blurring Process

- **Blurring Process**
- A historical model for explaining unnatural phenomena
  - (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
  (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**

\[
B > C \rightarrow -X \\
B > A \\
C > B \\
B > A \rightarrow X
\]
**Blurring Process**

- **Blurring Process**
- A historical model for explaining unnatural phenomena

(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**

\[
\begin{align*}
B & > C / \neg X \\
B & > A \\
C & > B \\
\hline
B & > A / X
\end{align*}
\]

**Blurring Chain**

\[
\begin{align*}
B & > C / X \\
C & > D \\
D & > A \\
\hline
B & > A / X
\end{align*}
\]
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

<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; ḍ/[−nas]___</td>
<td>Sogdian ḍand ḍasa</td>
</tr>
<tr>
<td>D &gt; T</td>
<td>Yaghnobi vant *ḍasa</td>
</tr>
<tr>
<td>d &gt; t</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]</td>
<td>Avestan band dasa</td>
</tr>
<tr>
<td>D &gt; T</td>
<td>Sogdian .band .dasa</td>
</tr>
<tr>
<td>Z &gt; D</td>
<td>Yaghnobi vant *dasa</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; ɗ/[–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>ɗ &gt; d</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Gašper Beguš
Harvard University — begus@fas.harvard.edu
Historical Probabilities of Alternations
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; ڦ/[−nas]_</td>
<td>Sogdian βand Ḯasa</td>
</tr>
<tr>
<td>D &gt; T</td>
<td>Yaghnobi vant * שקיבל</td>
</tr>
<tr>
<td>d &gt; t</td>
<td>Yaghnobi vant das</td>
</tr>
<tr>
<td>Z &gt; D</td>
<td></td>
</tr>
<tr>
<td>Ḧ &gt; d</td>
<td></td>
</tr>
</tbody>
</table>
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)
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")

■ The primary diachronic device for explaining unnaturalness
  
  **HYPERCORRECTION**
  
  (Ohala 1981)

■ **Blurring Process** a new diachronic device for explaining unnatural alternations
Minimal number of sound changes

- **Blurring process** also allows us to establish the minimal number of sound changes required for unnatural processes to arise.
- Unnatural alternations always require three operating sound changes.
Minimal number of sound changes

- Sound change: change in a single feature value (Picard 1994)
- $B > A / X$
Minimal number of sound changes

- Sound change: change in a single feature value (Picard 1994)
- B > A / X
- B > A / X by definition impossible
Minimal number of sound changes

- B > C / X
- C > A

Two sound changes? 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.
Minimal number of sound changes

- B > C / X
- C > A
- C ~ A:
Minimal number of sound changes

- B > C / X
- C > A
- C ~ A: $\phi_1$ (A ~ B)
Minimal number of sound changes

- B > C / X
- C > A
- C ∼ A: φ₁ (A ∼ B) and φ₂ (B ∼ C)
Minimal number of sound changes

- B > C / X
  C > A
- C ~ A: $\phi_1 (A \sim B)$ and $\phi_2 (B \sim C)$, $\phi_1 \neq \phi_2$
Minimal number of sound changes

- **B > C / X**
  
- **C > A**

- C ~ A: \( \phi_1 (A \sim B) \) and \( \phi_2 (B \sim C) \), \( \phi_1 \neq \phi_2 \)

- 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
Minimal number of sound changes

- At least three sound changes for unnatural alternations

**Minimal Sound Change Requirement (MSCR)**

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*
  \[ P_\chi(\text{natural}) < P_\chi(\text{unmotivated}) < P_\chi(\text{unnatural}) \]
Historical Probabilities

Historical Probabilities of Alternations ($P_\chi(A)$)

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

**Historical Probabilities of Alternations** \( (P_X(A)) \)

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

Historical Probabilities of Alternations ($P_{\chi}(A)$)

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

1 Introduction

2 Number of sound changes

3 Probabilities

4 Examples

5 Discussion and conclusion
Model

- The model can be quantified
- **Bootstrapping Sound Changes**
Poisson Stochastic Process


\[ P(T_1) = \int_0^t f_1 t_1 dt_1 \times \int_{t_1}^t f_2 t_2 dt_2 \times \ldots \times \int_{t_{n-1}}^t f_n t_n dt_n \]

where

\[ f_i = \lambda_i e^{-\lambda_i t} \]
Probability of a sound change

- We can disregard temporal dimension and model sound change probabilities from a given sample as a binary outcome model if we stay within the sample

\[ P_{\chi}(S_1) = \frac{\text{number of languages with sound change } S_1}{\text{number of languages surveyed}} \]
**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 $S_1$

$$P_\chi(S_1) = \frac{\text{number of languages with sound change } S_1}{\text{number of languages surveyed}}$$
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
  
  $$P(A_x) = \frac{P(S_1)P(S_2) \ast \ldots \ast P(S_n)}{n!}$$

- 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
  - $S_1$ and $S_2$ independent
  - $P(S_1|\text{phonemic inventory})$, 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 $P(S_1/\_\_X)$
  - 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
Outline

1 Introduction

2 Number of sound changes

3 Probabilities

4 Examples

5 Discussion and conclusion
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
- $P_\chi(PND) - P_\chi(FV) = [0.02\%, 0.11\%]$
Comparison of $P_\chi$

- 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 $P_\chi$
Prediction of attestedness

- Do we predict PND to be attested?
- Yes, difference between $P_\chi$ (PND) and $P(1/200)$ is [-0.05%, 2.47%]
- But: $P_\chi$ (FV) and $P(1/200)$ is [-0.003%, 2.86%]
- $P_\chi$ (FV) is not significantly smaller than probability of being attested once in the sample, but the $BC_a$ CI is [-0.003%, 2.86%] and it is significantly different from PND
Going “below zero”

- 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%]
Going “below zero”
Equiprobable processes

- **Equiprobable processes**

- **Analytic Bias** approach (AB): biases in learning affect the typology

- **Channel Bias** approach (CB) assumes phonetic precursors and transmission of language affect the typology

  - Empirical evidence in favor of both
  - Processes that are more difficult to learn are also typologically more rare
  - Processes that are the result of phonologized phonetically motivated sound changes are also typologically frequent

- **Identify equiprobable alternations** for experimental testing
Equiprobable processes

- Test learnability of processes that are equiprobable historically.
- If an alternation $A_1$ is typologically more common than $A_2$, but $P_{\chi}(A_1)$ and $P_{\chi}(A_2)$ are not significantly different, we can test the learnability of $A_1$ and $A_2$. If biases in learnability are confirmed, we would get a direct confirmation of $AB$ without the “duplication” problem outlined above.
Outline

1  Introduction

2  Number of sound changes

3  Probabilities

4  Examples

5  Discussion and conclusion
Discussion

- A new quantitative model of typology within CB
  - Historical probability of FV is small enough for the process not to be attested in a given sample, historical probability PND is significantly higher
  - Based on BSC, Channel Bias would thus correctly predict the observed typology, even in the case that has traditionally been used against CB.

- Comparison of historical probabilities of alternations with inference
- Prediction of attestedness
- Quantification “below zero”
- Identification of historically equiprobable alternations
A formal model of typology

- Many acknowledge 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
- A typological model that models both AB and CB with MaxEnt
A formal model of typology

- $\sigma^2$ in the prior (Wilson 2006) in MaxEnt models encodes some processes are underlearned
- Unnatural processes often equally learnable (Seidl et al. 2007, Do et al. 2016)
- Substantive bias?
- $w_\chi$ (CB) and $\sigma^2$ (AB)

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

- Crucially, when modeling acquisition, speakers have no access to historical weights or probabilities
- When modeling typology, both AB and CB contribute to the typology
A formal model of typology

- A model with historical weights performs better: it generates all attested patterns, but encodes that some are rare (due to historical probabilities and learnability)
- Example: No difference in learnability experiments between PND and PNV (Do et al. 2016)
- $\sigma^2$ of PND and PNV should be equal
A formal model of typology

- A model with historical weights performs better: it generates all attested patterns, but encodes that some are rare (due to historical probabilities and learnability)
- Example: 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)
A formal model of typology

- A model with historical weights performs better: it generates all attested patterns, but encodes that some are rare (due to historical probabilities and learnability)
- Example: 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>/NT/</th>
<th>IDENT-IO</th>
<th>*NT</th>
<th>$H_\chi$</th>
<th>$P_\chi$</th>
<th>Typol.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w_\chi = 10$</td>
<td>$w_\chi = 8.64$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. [NT]</td>
<td></td>
<td>$-1$</td>
<td>$-10$</td>
<td>.795</td>
<td>.924</td>
</tr>
<tr>
<td>b. [ND]</td>
<td></td>
<td>$-1$</td>
<td>$-8.64$</td>
<td>.205</td>
<td>.076</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>/ND/</th>
<th>IDENT-IO</th>
<th>*ND</th>
<th>$H_\chi$</th>
<th>$P_\chi$</th>
<th>Typol.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$w_\chi = 10$</td>
<td>$w_\chi = 2.34$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. [ND]</td>
<td></td>
<td>$-1$</td>
<td>$-10$</td>
<td>.99953</td>
<td>$\approx .9975$</td>
</tr>
<tr>
<td>b. [NT]</td>
<td></td>
<td>$-1$</td>
<td>$-2.34$</td>
<td>.00047</td>
<td>$\approx .0025$</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
Future directions

- Phonotactics — MSCR limited to alternations, but unnatural phonotactics result from “blurring process” as well
  
  (Beguš and Nazarov 2017)

- Further predictions: unmotivated alternations involving multiple features predicted to be less frequent compared to those requiring less features

- Larger, balanced samples

- Artificial grammar learning experiments


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.