Unnatural phenomena and gradient phonotactics

Gašper Beguš

Harvard University
begus@fas.harvard.edu

phoNE
New York University
April 9, 2016
Introduction

- Can gradient phonotactics be “unnatural”?
- What is unnatural?
- A model for explaining unnaturalness and typology
- How to capture (unnatural) gradience theoretically?
Gradient phonotactics

- Two aspects of OT widely discussed: how to handle **naturalness** and **gradience** (Frisch et al. 2004, Antilla 2008, Coetzee and Pater 2008)

- Little has been said about intersection of the two: **unnatural gradient phonotactics**

- Gradience implies naturalness?

- **Question:** Can gradience be unnatural?

- **Claim:** unnatural gradient phonotactics exists

- Tarma Quechua stop voicing

- Berawan dialects stop devoicing
Naturalness

- **Natural**: phonetically grounded
- **Unmotivated**: lack phonetic motivations
- **Unnatural**: operating in the opposite direction from universal phonetic tendency
Naturalness

- Traditionally, unnatural all processes that lack phonetic motivation
- Most studies on phonotactics involve unmotivated restrictions (Hayes and White 2013, Albright 2009)
- Likewise, studies on gradience only include natural processes
- No cases of unnatural gradient phonotactics so far
Outline

1 Introduction

2 Unnatural gradient phonotactics

3 Typology

4 Theory

5 Conclusion
Tarma

- Tarma Quechua, a dialect of Quechua spoken in Tarma, Junín, Peru (Adelaar 1977, Puente Baldoceda 1997, Nazarov 2008)

- **Stop voicing**

- Proto-Quechua, Pre-Tarma only voiceless stops

- Voicing occurs: intervocalically, post-consonantally, but not post-nasally

- Bilabial, velar series undergo voicing, dental remain voiceless
Data

- From Adelaar 1977 and Nazarov 2008
  \[ p, k > b, g / C \quad ; \quad C \neq N \]
  \[ p, k > b, g / V \quad V \]
- Examples:

<table>
<thead>
<tr>
<th></th>
<th>*pirwa</th>
<th>pirwa</th>
</tr>
</thead>
<tbody>
<tr>
<td>N__</td>
<td>*wampu-</td>
<td>wampu</td>
</tr>
<tr>
<td>V__V</td>
<td>*kupa-</td>
<td>kupa</td>
</tr>
<tr>
<td>Y,R,S,T__</td>
<td>*kipu-</td>
<td>kipu</td>
</tr>
<tr>
<td></td>
<td>*takpa-</td>
<td>takba</td>
</tr>
</tbody>
</table>
- In some words: lexical variation, not productive
Data

- Distribution of voicing
- Native vocabulary analyzed by Nazarov (2008)
Data

- Post-nasally voiced stops universally preferred
  (Hayes and Stivers 2000)
- After voiceless stops voiced stops universally dispreferred
  (voice disagree)
- Intervocally, voiced stops universally preferred
- Gradience goes in the **opposite direction**!
Data

- More voicing after T than after Y, R (Nazarov 2008)

<table>
<thead>
<tr>
<th></th>
<th>Y,R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiced</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>Voiceless</td>
<td>14</td>
<td>6</td>
</tr>
</tbody>
</table>

However, this is not significant $p < 0.10$
Data

- Post-consonantally voicing after (Nazarov 2008):
  \( t, \text{št}, \text{št}, k, s, j, x, l, \text{št}, r, j, w \)

<table>
<thead>
<tr>
<th>Pre-Tarma</th>
<th>Tarma</th>
<th>Pre-Tarma</th>
<th>Tarma</th>
</tr>
</thead>
<tbody>
<tr>
<td>aštka</td>
<td>aštga</td>
<td>maštka</td>
<td>maštga</td>
</tr>
<tr>
<td>aštspa</td>
<td>aštšba</td>
<td>arku</td>
<td>argu</td>
</tr>
<tr>
<td>arku</td>
<td>argu</td>
<td>kutštka</td>
<td>kutšga</td>
</tr>
<tr>
<td>awkis</td>
<td>awgis</td>
<td>lušpi</td>
<td>lušbi</td>
</tr>
<tr>
<td>ajpa</td>
<td>ajba</td>
<td>pilpa</td>
<td>pilba</td>
</tr>
<tr>
<td>ţfaspu</td>
<td>ţfasbu</td>
<td>luxpi</td>
<td>luxbi</td>
</tr>
<tr>
<td>ţšilpi</td>
<td>ţšilbi</td>
<td>mutki</td>
<td>mutgi</td>
</tr>
<tr>
<td>takpa</td>
<td>takba</td>
<td>tikpa</td>
<td>tikba</td>
</tr>
<tr>
<td>lutpi</td>
<td>lutbi</td>
<td>tikpi</td>
<td>tikbi</td>
</tr>
</tbody>
</table>
Another locus of unnaturalness: **kb, tb, tg**

Intervocalic stop-stop cluster: **VC₁C₂V** when **C₂ = [b] or [g]**

<table>
<thead>
<tr>
<th>TT</th>
<th>TD</th>
<th>DT</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>66</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Clusters that agree in voice preferred

Pre-voicing preferred

Tarma Quechua gradience in the **opposite direction**
Data
Data

<table>
<thead>
<tr>
<th></th>
<th>No Voicing</th>
<th>Voicing</th>
</tr>
</thead>
<tbody>
<tr>
<td>N_</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>V_V</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Y,R,S,T_</td>
<td>100%</td>
<td>0%</td>
</tr>
</tbody>
</table>

\[ p < 0.00001 \]
[aku]
[agu]
Unnatural phenomena and gradient phonotactics
[ebi]
Unnatural phenomena and gradient phonotactics
[arba]
Unnatural phenomena and gradient phonotactics
[atbi]
Unnatural phenomena and gradient phonotactics
Unnatural phenomena and gradient phonotactics

[asga]
Yaghnobi

- A model for explaining unnatural alternations

**BLURRING CYCLE**

- $B > C / Z$
- $B > A$
- $C > B$

**BLURRING CHAIN**

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

- Xromov (1972, 128): \(D > T/N\)

<table>
<thead>
<tr>
<th>Yaghnobi</th>
<th>Sogdian</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>'γantum</td>
<td>γandum</td>
<td>‘wheat’</td>
</tr>
<tr>
<td>ʃi'kampa</td>
<td>əʃkamb</td>
<td>‘stomach’</td>
</tr>
<tr>
<td>sank(a)</td>
<td>sang</td>
<td>‘stone’</td>
</tr>
<tr>
<td>ran'kiːna</td>
<td>rang</td>
<td>‘color’</td>
</tr>
<tr>
<td>un'kuʃt</td>
<td>anguʃt</td>
<td>‘finger’</td>
</tr>
<tr>
<td>'ʧintir</td>
<td>ʧǝndǝr</td>
<td>postp.</td>
</tr>
<tr>
<td>-ant</td>
<td>-and</td>
<td>3rd pl.</td>
</tr>
</tbody>
</table>
Direct historical evidence for such a proposal: **Sogdian**

**Stage 1** is directly attested:

<table>
<thead>
<tr>
<th>Language</th>
<th>Rule</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proto-Iranian</td>
<td>D &gt; D / [-nas]__</td>
<td>*band</td>
</tr>
<tr>
<td>Sogdian</td>
<td>D &gt; D</td>
<td>βand</td>
</tr>
<tr>
<td>Yaghnobi</td>
<td>D &gt; T</td>
<td>vant</td>
</tr>
</tbody>
</table>
Tswana, Shekgalagari

- Solé *et al.* (2010):

<table>
<thead>
<tr>
<th>No N-prefix</th>
<th>N-prefix</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>χu-pak-a</td>
<td>χu-m-pak-a</td>
<td>‘to praise’</td>
</tr>
<tr>
<td>χu-tut-a</td>
<td>χu-n-tut-a</td>
<td>‘to respect’</td>
</tr>
<tr>
<td>χu-cúb-á</td>
<td>χu-ŋ-cúb-á</td>
<td>‘to beat’</td>
</tr>
<tr>
<td>χu-kεl-a</td>
<td>χu-ŋ-kεl-a</td>
<td>‘to show’</td>
</tr>
<tr>
<td>χu-bón-á</td>
<td>χu-m-pón-á</td>
<td>‘to see’</td>
</tr>
<tr>
<td>χu-duz-a</td>
<td>χu-n-duz-a</td>
<td>‘to annoint’</td>
</tr>
<tr>
<td>χu-jís-a</td>
<td>χu-ŋ-cís-a</td>
<td>‘to feed’</td>
</tr>
<tr>
<td>χu-at-a</td>
<td>χu-ŋ-gat-a</td>
<td>‘like’</td>
</tr>
<tr>
<td>χu-hinp-a</td>
<td>χu-m-pʰinp-a</td>
<td>‘defeat’</td>
</tr>
</tbody>
</table>
Tswana, Shekgalagari

- Dialectal evidence (Solé et al. 2010)
- Tswana has systems with fricativization except after nasals and systems with unconditioned devoicing
- Combination of these two dialects → post-nasal devoicing

<table>
<thead>
<tr>
<th></th>
<th>#ba</th>
<th>aba</th>
<th>mba</th>
</tr>
</thead>
<tbody>
<tr>
<td>devoicers</td>
<td>#pa</td>
<td>apa</td>
<td>mpa</td>
</tr>
<tr>
<td>leniters</td>
<td>#βa</td>
<td>aβa</td>
<td>mba</td>
</tr>
<tr>
<td>post-nasal devoi.</td>
<td>#ba</td>
<td>aba</td>
<td>mpa</td>
</tr>
</tbody>
</table>

- Crucially, fricatives got occluded back to stops
A model for explaining unnatural alternations

**BLURRING CYCLE**
- \( B > C / Z \)
- \( B > A \)
- \( C > B \)

**BLURRING CHAIN**
- \( B > C / X \)
- \( C > D \)
- \( D > A \)
## Explanation of Tarma Quechua

<table>
<thead>
<tr>
<th>BLURRING CHAIN</th>
<th>Tarma Quechua</th>
</tr>
</thead>
<tbody>
<tr>
<td>B &gt; C / X</td>
<td>T &gt; S / Y,R,S,T__, V__V</td>
</tr>
<tr>
<td>C &gt; D</td>
<td>S &gt; Z</td>
</tr>
<tr>
<td>D &gt; A</td>
<td>Z &gt; D</td>
</tr>
</tbody>
</table>
Data

- **p < 0.00001**
Phonetics

- Stops surface as fricatives sometimes
Unnatural phenomena and gradient phonotactics
[usbi]
[agii]
Phonetics

- This indicates that there was a stage in the development with **voiced fricatives**
- Occlusion to stops not operated categorically
Intervocalic Devoicing

- Intervocalic (or postvocalic) voicing is common
  \[ T > D / V\text{(V)} \]

- The most frequent type of lenition (alongside spirantization, approximatization and others)
  Kaplan (2010)

- 26 of 153 languages have intervocalic lenition
Intervocalic Devoicing

- Good phonetic rationale
- **Passive voicing**: voicing into closure of intervocalic voiceless stops (e.g. German)
  
  Jessen (1998)
- Articulation
  
  Keating (1986)
- P-map: minimal change to achieve phonotactics: $T \rightarrow D$
  
  Kaplan (2010), Steriade (2010)
- Initial devoicing common (cf. English, Yidin)
Intervocalic Devoicing

- Very common sound change
- Reported in over 40 languages

Kümmel (2007)

Intervocalic voicing is a universal **phonetic tendency**.
Intervocalic Devoicing

- Intervocalic *devoicing* is the opposite process
  \[ D > T / V__(V) \]

- **Unattested** as a synchronic phonological process

- It would operate against the universal phonetic tendency of voicing intervocalic voiceless stops

- P-map: spirantization is perceptually less salient than devoicing in intervocalic position

Steriade (2001), Kaplan (2010)
Berawan dialects

- In Berawan dialects labials and velars undergo devoicing, as reported in Blust (2013).
- Labials additionally change place of articulation (but only intervocalically!)
- Analysis on the basis of description in Burkhardt (2014)
Berawan dialects

- **g (from r)** $\rightarrow$ **k / V__V**
  - *bigiu* $\rightarrow$ *bikiw*
  - *gigiəq* $\rightarrow$ *giki?
  - *magi* $\rightarrow$ *maki*
  - *igιŋ* $\rightarrow$ *ikiŋ*
  - *ugat* $\rightarrow$ *kit*

- **b** $\rightarrow$ **k / V__V**
  - *abiŋ* $\rightarrow$ *akiŋ*
  - *bibi* $\rightarrow$ *biki*
  - *bəliwiŋ* $\rightarrow$ *bəlikiŋ*
  - *bibuy* $\rightarrow$ *bikuy*
  - *dibiŋ* $\rightarrow$ *dikin*
Long Terawan

- Word-initially, stops remain voiced

| *gəm      | >  | gəm         |
| *gigun    | >  | gikuŋ       |
| *gimot    | >  | gimok       |
| *gitaq    | >  | gitaʔ       |
| *bitok    | >  | bitok       |
| *buliən   | >  | bulin       |
| *busak    | >  | busek       |
| *buttan   | >  | buten       |
### Berawan

<table>
<thead>
<tr>
<th>Dialect</th>
<th>Labials</th>
<th>Velars</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>#p</td>
<td>VpV</td>
</tr>
<tr>
<td>Batu Belah</td>
<td>52</td>
<td>10</td>
</tr>
<tr>
<td><em>p-value</em></td>
<td></td>
<td>0.1168</td>
</tr>
<tr>
<td>Long Teru</td>
<td>46</td>
<td>13</td>
</tr>
<tr>
<td><em>p-value</em></td>
<td></td>
<td>0.023</td>
</tr>
<tr>
<td>Long Jegan</td>
<td>49</td>
<td>10</td>
</tr>
<tr>
<td><em>p-value</em></td>
<td></td>
<td>0.1157</td>
</tr>
<tr>
<td>Long Terawan</td>
<td>41</td>
<td>11</td>
</tr>
<tr>
<td><em>p-value</em></td>
<td></td>
<td>0.004</td>
</tr>
</tbody>
</table>

Unnatural phenomena and gradient phonotactics

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

Batu Belah

Labial | Alveolar | Velar
--- | --- | ---
 Voiced | Voiceless | variable

% distribution:
- Labial: 0%, 25%, 50%, 75%, 100%
- Alveolar: 0%, 25%, 50%, 75%, 100%
- Velar: 0%, 25%, 50%, 75%, 100%
Berawan

Long Teru

Labial | Alveolar | Velar
--- | --- | ---
# | V-V | # | V-V | # | V-V

variable

Voiced | Voiceless
Berawan

Long Jegan

- **Labial**
  - # - V_V
  - 0% - 25% - 50% - 75% - 100%
- **Alveolar**
  - # - V_V
  - 0% - 25% - 50% - 75% - 100%
- **Velar**
  - # - V_V
  - 0% - 25% - 50% - 75% - 100%

Variable:
- **Voiced**
- **Voiceless**
Berawan

Long Terawan

Labial

Alveolar

Velar

variable

Voiced

Voiceless

Gašper Beguš
Harvard University — begus@fas.harvard.edu
Unnatural phenomena and gradient phonotactics
Crucial problem

- \(*b > k / V__V*

  **Scenario 1**  **Scenario 2**
  1. \(b > p\)  \(b > g\)
  2. \(p > k\)  \(g > k\)

- Problem: if devoicing occurs before change of place, why no change of the original *p?*
  - \(*hapuy > apuy, not **akuy*

- If change of place first, why only word-internally?
Berawan dialects (Blust 2005)

- Berawan stops

\[ D > T / V___V \]

\[
\begin{array}{ccc}
\# & V & V \\
*b & b & k \\
*d & d & r \\
*g & g & k \\
\end{array}
\]
Berawan dialects (Blust 2005)

- Blurring chain

<table>
<thead>
<tr>
<th>BLURRING CHAIN</th>
<th>Berawan</th>
</tr>
</thead>
<tbody>
<tr>
<td>B &gt; C / X</td>
<td>D &gt; Z / V__V</td>
</tr>
<tr>
<td>C &gt; D</td>
<td>Z &gt; S</td>
</tr>
<tr>
<td>D &gt; A</td>
<td>S &gt; T</td>
</tr>
</tbody>
</table>
Typology

- A combination of sound changes appears unconstrained
- A single sound change is constrained, combination unconstrained
- How to derive typology?
Minimal number of sound changes

- Note that unnatural processes always require three operating sound changes
- $B > A / X$
- $B > C / X$
- $C > A$

At least three sound changes
- While it is theoretically possible $B > C$ and $C > A$, the last sound change would necessarily involve two features to change, which in fact means two sound changes as $C$ differs from $B$ in one feature, which cannot be the feature producing unnaturalness, and $A$ differs from $B$ in one feature, which has to be the feature producing unnaturalness in a given context

- Crucial assumption: sound change is a change in one feature in a given environment
Typology

- Rare phenomena produced by rare sound changes
  (Blevins 2008)
- Morphological phenomena rare because of complex history
  (Harris 2003, 2008)

\[ P(A \cap B) = P(A)P(B) \leq P(A) \]

\[ P(A \cap B \cap C) = .1 \times .1 \times .1 = .001 \]
Typology

- **Natural**
- **Unmotivated**: Two sound changes, less frequent (telescoping, Wang 1968)
- **Unnatural**: At least three sound changes, least frequent

- Low probability not the only reason
- As soon as an B → A / X arises, the inverse universal phonetic tendency A > B / X will begin operating against it
- Confirmed by Tswana
A problem for theory?

- It is generally agreed upon that gradient phonotactics has to be encoded in the grammar (Coetzee and Pater 2008, Antilla 2008)
- Various approaches for capturing gradience theoretically
- **The problem:** how to derive a system in which unnatural element is more common?
A problem for theory?

- **Harmony** (HG) can be transformed to percentages, but given richness of the base, we cannot derive a system in which the unnatural element is more frequent.

- **Faith** and a markedness constraint *X

- If equal weights, $P(\text{UNNAT}) < .5$

- If either Faith or *X have greater weights:
  a) If Faith > *X: $P(\text{UNNAT}) = .5$
  b) If Faith < *X: $P(\text{UNNAT}) < .5$
A problem for theory?

- Should CON be restricted?
- *−X
- DISAGREE
- *VDV
- *ND

**Problem:** how to encode these are rare?
A new proposal

- A new proposal: Inherently Weighted Constraints
- Both *X and *−X
- Constraints weighted, subject to normal distribution
- Derives such systems and encodes typology
A new proposal

\[ W_1 - W_2 = \Delta W = 2 \operatorname{erf}^{-1}(1 - 2P(C_2 \gg C_1)) \]

**Inherent Weights of Two Constraints**

![Graph showing the inherent weights of two constraints with values W1 = 1.6784 and W2.]
A new proposal

\[ W_1 - W_2 = \Delta W = 2 \text{erf}^{-1}(1 - 2P(C_2 \gg C_1)) \]

Inherent Weights of Two Constraints

\begin{itemize}
  \item \( W_1 - W_2 = 3.29 \)
  \item \( W_2 \)
  \item \( W_1 \)
\end{itemize}
Outline

1 Introduction
2 Unnatural gradient phonotactics
3 Typology
4 Theory
5 Conclusion

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

Unnatural phenomena and gradient phonotactics
Conclusion

- Gradient phonotactics can be unnatural
- Sound change restricted: combinations result in unnatural processes
- **Blurring chain**
- Theoretical implications: a new constraint architecture required
Beguš, Gašper. 2015a. Post-nasal devoicing as a sound change. Presentation at the 89th Annual LSA Meeting, Portland, OR.
———. 2015b. Intervocalic Devoicing in Kiput and Berawan Dialects. Presentation at the 22nd AFLA, McGill University.
References


References


References


References


Thank you!

* I would like to thank Kevin Ryan, Maria Polinsky, and Aleksei Nazarov for their useful comments and Willem Adelaar for his kind permission to analyze his recordings. Special thanks goes also to Asia Center at Harvard University for supporting my research. All mistakes are my own.