Unnatural and Lexically Gradient Phonology

Gašper Beguš\textsuperscript{1}  Aleksei Nazarov\textsuperscript{2}

\textsuperscript{1}Harvard University  
\textit{begus@fas.harvard.edu}

\textsuperscript{2}University of Huddersfield  
\textit{a.nazarov@hud.ac.uk}

5\textsuperscript{th} Annual Meeting on Phonology  
New York University  
September 15-17, 2017
Gradient phonotactics

Two aspects of OT widely discussed:

- How to represent **gradient** phonotactic restrictions

- How to represent **unnatural** processes

No systematic treatment of the intersection: **unnatural gradient** phonotactics

- Can gradient phonotactic restrictions operate in the phonetically unnatural direction?

- Tarma Quechua stop voicing
Naturalness

- A new division of naturalness

<table>
<thead>
<tr>
<th>Phonetic tendencies are enforced by</th>
<th>contradicted by</th>
</tr>
</thead>
<tbody>
<tr>
<td>natural processes</td>
<td>✔</td>
</tr>
<tr>
<td>unmotivated processes</td>
<td>✗</td>
</tr>
<tr>
<td>unnatural processes</td>
<td>✗</td>
</tr>
</tbody>
</table>
Literature so far

- Unnatural categorical process: post-nasal devoicing, confirmed in Tswana with wug-tests (Coetzee and Pretorius 2010)
- Most other processes discussed are in fact *unmotivated*
- Some processes labeled as “unnatural” in Hayes and White (2013)
  - “No [θ,ð] before stressless rounded vowels”
    \[
    \begin{array}{c}
    +COR \\
    +cont \\
    -strid \\
    \end{array}
    \begin{array}{c}
    -stress \\
    +round \\
    \end{array}
    \]
  - “No [ʒ] before stressed vowel + obstruent”
    \[
    \begin{array}{c}
    +cont \\
    +voice \\
    \end{array}
    \begin{array}{c}
    +stress \\
    -son \\
    -ant \\
    \end{array}
    \]
Outline

1 Introduction

2 Background

3 Data

4 Origins

5 Synchronic implications

6 Conclusions
Tarma

- TQ a dialect of Quechua spoken in Tarma, Junín, Peru (Adelaar 1977, Puente Baldoceda 1977)
- **Distribution of $[\pm \text{voice}]$ in [DOR] and [LAB] stops**
- Adelaar (1977): $[+\text{voice}]$: intervocalically, post-consonantally, but not post-nasally
Data

- From Adelaar (1977):

  b, g / C____; C ≠ N
  
  b, g / V____V
  
  p, k / elsewhere

  #____ [pirwa]
  ___R, T [rikra]
  N____ [wampu]
  V____V [kuba]
  R, T___ [takba]

- Adelaar (1977) offers no further descriptions on the distribution
Data

- From Adelaar (1977):

  $b, g / C \_\_\_; C \neq N$

  $b, g / V \_\_V$

  $p, k / \text{elsewhere}$

  $
  \begin{array}{c|c}
    \# \_ \_ & \text{[pirwa]} \\
    \_ \_ \_R, T & \text{[rikra]} \\
    N \_ \_ & \text{[wampu]} \\
    V \_\_V & \text{[kuba]} \\
    R, T \_ \_ & \text{[takba]}
  \end{array}
  $

- Adelaar (1977) offers no further descriptions on the distribution.


**Data**

- From Adelaar (1977):
  
  \[
  \begin{align*}
  &b, g / C\____; C \neq N \\
  &b, g / V\____V \\
  &p, k / \text{elsewhere} \\
  &\#\____ / [pirwa] \\
  &\____R, T / [rikra] \\
  &N\____ / [wampu] \\
  &V\____V / [kuba] \\
  &R, T\____ / [takba]
  \end{align*}
  \]

- Adelaar (1977) offers no further descriptions on the distribution.
Data

- Lexical, phonetic, and morphophonological analysis
- Unnatural gradient phonotactic restrictions
Data

- Distribution of voicing
- Native vocabulary from Adelaar (1977)

Counts:

- All tokens with [DOR] or [LAB] in TQ vocabulary (Adelaar 1977)
- 1199 tokens: 910 in native vocabulary, 289 in loans from Spanish
- Each data point was annotated for presence or absence of voicing, place of articulation of the stop (labial or velar), and position in the word
# Data

Counts:

<table>
<thead>
<tr>
<th></th>
<th>#</th>
<th>N</th>
<th>V</th>
<th>R</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>voiced</td>
<td>7</td>
<td>7</td>
<td>99</td>
<td>72</td>
<td>68</td>
</tr>
<tr>
<td>voiceless</td>
<td>276</td>
<td>67</td>
<td>134</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>% voiced</td>
<td>2.5</td>
<td>9.5</td>
<td>42.5</td>
<td>84.7</td>
<td>86.1</td>
</tr>
</tbody>
</table>
Data

- Logistic regression model:

|                | Est.  | SE    | z value | Pr(>|z|) |
|----------------|-------|-------|---------|----------|
| (Intercept)    | -0.045| 0.172 | -0.260  | 0.7952   |
| V__V vs. R__   | 2.044 | 0.332 | 6.164   | 0.0000   |
| V__V vs. T__   | 2.155 | 0.353 | 6.101   | 0.0000   |
| V__V vs. N__   | -1.884| 0.421 | -4.478  | 0.0000   |
| V__V vs. #__   | -3.437| 0.407 | -8.437  | 0.0000   |
| velar vs. labial | -0.502| 0.214 | -2.344  | 0.0191   |
Data

A graph showing the percentage of voiced sounds across different positions.

- Position: #_, N_, V_, R_, T_
- Y-axis: % voiced
- X-axis: Position

Gašper Beguš, Aleksei Nazarov
Harvard University and University of Huddersfield
Unnatural and Lexically Gradient Phonology
Data

The diagram shows the percentage of voiced sounds in different positions. The x-axis represents the position, and the y-axis represents the percentage of voiced sounds. The positions are labeled as `#`, `N`, `V V`, `R`, and `T`. The data points are depicted with red markers, and error bars indicate the variability.
Data

A graph shows the percentage of voiced sounds across different positions. The x-axis represents positions such as '#', 'N', 'V', 'R', and 'T', and the y-axis represents the percentage voiced. The data points are marked with error bars, indicating variability within the data. The graph illustrates a trend where the percentage of voiced sounds increases as the position moves from left to right.
## Data

<table>
<thead>
<tr>
<th>Universal tendencies for [+voice]</th>
<th>Observed significant trends in TQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>T</strong>__ &lt; <strong>V</strong>__<strong>V</strong></td>
<td><strong>V</strong><strong><strong>V</strong> &lt; <strong>T</strong></strong></td>
</tr>
<tr>
<td><strong>T</strong>__ &lt; <strong>N</strong>__</td>
<td><strong>N</strong>__ &lt; <strong>V</strong><strong><strong>V</strong> &lt; <strong>T</strong></strong></td>
</tr>
</tbody>
</table>
Data

- Another locus of unnaturalness: TT sequences

<table>
<thead>
<tr>
<th>1\textsuperscript{st} member</th>
<th>2\textsuperscript{nd} member</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Labial</td>
</tr>
<tr>
<td>t</td>
<td>lutbi</td>
</tr>
<tr>
<td>ţaj</td>
<td>/</td>
</tr>
<tr>
<td>ţs</td>
<td>aţşba</td>
</tr>
<tr>
<td>k</td>
<td>takba</td>
</tr>
<tr>
<td>s</td>
<td>tţasbu</td>
</tr>
<tr>
<td>j</td>
<td>kaţbi</td>
</tr>
<tr>
<td>x</td>
<td>saxbi</td>
</tr>
<tr>
<td>l</td>
<td>tţilbi</td>
</tr>
<tr>
<td>r</td>
<td>karba</td>
</tr>
<tr>
<td>j</td>
<td>ajba</td>
</tr>
<tr>
<td>w</td>
<td>kawbu</td>
</tr>
</tbody>
</table>
**Data**

- Second-element stops (labial and velar) are significantly more frequently voiced (as opposed to voiceless) in clusters with a voiceless first element in TQ native vocabulary ($\beta = 1.8$, $z = 5.6$, $p < 0.0001$)

<table>
<thead>
<tr>
<th></th>
<th>TT</th>
<th>TD</th>
<th>DT</th>
<th>DD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Count</strong></td>
<td>11</td>
<td>68</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Percent</strong></td>
<td>13.9%</td>
<td>86.1%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

- All effects thus far remain even if we add loanwords to the models.
Data

- Phonetic analysis
- Recordings by Willem Adelaar, analyzed in Praat (Boersma and Weenink 2015)
Unnatural and Lexically Gradient Phonology
[akba]
[ukba]
After fricatives
[asba]
[asga]
Data

- After nasals
- Unaspirated
[ampa]
[aŋki]
Productivity

- **Alternating suffixes**
  - -ba/-pa ‘genitive’
  - -bax/-pax ‘purposive’
  - -bita/-pita ‘procedentive’
  - -bis/-pis ‘even, too’
Productivity

- **Alternating suffixes**
  - -ba/-pa ‘genitive’
  - -bax/-pax ‘purposive’
  - -bita/-pita ‘procedentive’
  - -bis/-pis ‘even, too’

- **Intervocalic**
  wawxi-gi-ba wayi-n
  ‘the house of your brother’

- **Post-nasal**
  wayi-n-pa pasa-un
  ‘we’re going to walk by way of his house’

- **Post-obstruent**
  tamya-ya-n nuqa-ntik-baq
  ‘it is raining now for us’ (Creider 1968:12-13)
Productivity

- **Alternating suffixes**
  - -ba/-pa ‘genitive’
  - -bax/-pax ‘purposive’
  - -bita/-pita ‘procedentive’
  - -bis/-pis ‘even, too’

- **Intervocalic**
  - wawxi-gi-*ba* wayi-n
  ‘the house of your brother’

- **Post-nasal**
  - wayi-n-pa pasa-un
  ‘we’re going to walk by way of his house’

- **Post-obstruent**
  - tamya-ya-n nuqa-ntik-baq
  ‘it is raining now for us’ (Creider 1968:12-13)
Productivity

- **Alternating suffixes**
  - -ba/-pa ‘genitive’
  - -bax/-pax ‘purposive’
  - -bita/-pita ‘procedentive’
  - -bis/-pis ‘even, too’

- **Intervocalic**
  wawxi-gi-ba wayi-n
  ‘the house of your brother’

- **Post-nasal**
  wayi-n-pa pasa-un
  ‘we’re going to walk by way of his house’

- **Post-obstruent**
  tamya-ya-n nuqa-ntik-baq
  ‘it is raining now for us’
  (Creider 1968:12-13)
Productivity

- **Alternating suffixes**
  - -ba/-pa ‘genitive’
  - -bax/-pax ‘purposive’
  - -bita/-pita ‘procedentive’
  - -bis/-pis ‘even, too’

- **Intervocalic**
  - wawxi-gi-ba wayi-n
    ‘the house of your brother’

- **Post-nasal**
  - wayi-n-pa pasa-un
    ‘we’re going to walk by way of his house’

- **Post-obstruent**
  - tamya-ya-n nuqa-ntik-*baq*
    ‘it is raining now for us’

(Creider 1968:12-13)
Productivity

- Loanwords:
  - Sp. *cuculi* > kuguli: ‘white-winged dove’
  - Sp. *cotpe* > kutbi ‘an animal from the mountains’
  - Sp. *sauco* > sawgu ‘magic tree’
  - Sp. *vaca* > waːga ‘cow’

- In two loanwords, a Spanish voiced intervocalic stop devoices to a TQ voiceless stop (data from Adelaar 1977).
  - Sp. *taruga* > taruka ‘deer’
  - Sp. *dios se lo pague* > jusulpaːki ‘thank you’
Outline

1. Introduction
2. Background
3. Data
4. Origins
5. Synchronic implications
6. Conclusions
Origins of TQ stop voicing

- How did this phonotactic restriction arise?

<table>
<thead>
<tr>
<th>Context</th>
<th>Voicing</th>
<th>Labial</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pre-TQ</td>
<td>TQ</td>
</tr>
<tr>
<td>#__</td>
<td>X</td>
<td>*pirwa</td>
<td>pirwa</td>
</tr>
<tr>
<td>N__</td>
<td>X</td>
<td>*wampu</td>
<td>wampu</td>
</tr>
<tr>
<td>V__V</td>
<td>✓</td>
<td>*kupa</td>
<td>kupa</td>
</tr>
<tr>
<td>R,T__</td>
<td>✓</td>
<td>*takpa</td>
<td>takpa</td>
</tr>
</tbody>
</table>

- The most intriguing aspect about this hypothetical sound change is that this unnatural voicing operates gradiently rather than categorically with different rates of application across different environments.
Origins of TQ stop voicing

- A diachronic device for explaining unnatural processes: Blurring Process (Beguš 2016)
- A > B / X natural
- B > A / X unnatural
  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

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

Blurring Chain

\[
\begin{align*}
B & > C \ / \ X \\
C & > D \\
D & > A
\end{align*}
\]
Origins of TQ stop voicing

- **Blurring Chain in Tarma Quechua**

  \[ T > S \quad / \quad [-\text{nas},-\#] \quad / \quad p > \phi \quad / \quad [-\text{nas},-\#] \]

  \[ S > Z \quad / \quad ___V \quad / \quad \phi > \beta \quad / \quad ___V \]

  \[ Z > D \quad / \quad ___V \quad / \quad \beta > b \]

- **Blurring Chain in Tarma Quechua**

<table>
<thead>
<tr>
<th>#__</th>
<th>V__V</th>
<th>N__</th>
<th>T__</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>pirwa</td>
<td>kupa</td>
<td>wampu</td>
</tr>
<tr>
<td>2.</td>
<td>pirwa</td>
<td>kuφa</td>
<td>wampu</td>
</tr>
<tr>
<td>3.</td>
<td>pirwa</td>
<td>kuβa</td>
<td>wampu</td>
</tr>
<tr>
<td>4.</td>
<td>pirwa</td>
<td>kuba</td>
<td>wampu</td>
</tr>
</tbody>
</table>
Distribution

![Graph showing distribution of voiced sounds across positions.

- % voiced y-axis
- Position x-axis with positions labeled #_, N_, V_V, R_, T_.

The graph illustrates the increase in voiced sounds from left to right, with the highest percentage occurring in the T_ position.

Gasper Beguš, Aleksei Nazarov
Harvard University and University of Huddersfield

Unnatural and Lexically Gradient Phonology

34 / 53
Origins of TQ stop voicing

- **Support from dialectal data:**

  - *aptay* → [haxʷtay]
  - *upyay* → [uxyay]

- **Fricativization of voiceless stops in Cusco Quechua**

- **Aspiration and fricativization in Imbabura Quechua**

  - Proto-Quechua *paki* → *phaki* → Imbabura Quechua [faki]
  - Proto Quechua *qipa* → *kipa* → Imbabura Quechua [xipa]
[ubi]
[atbi]
Outline

1. Introduction
2. Background
3. Data
4. Origins
5. Synchronic implications
6. Conclusions
Synchronic implications

- Deriving typology one of the main advantages of OT
- Harmonic Grammar (HG) with numerically weighted constraints well-suited for gradient processes (Pater 2009)
- MaxEnt: Probability distribution over candidates (Goldwater and Johnson 2003)
- Problem that HG approach faces: the derivation of unnatural processes
Synchronic implications

- OT with restricted CON: factorial typology, unnatural processes unattested (a desired prediction for final voicing)
- HG: An additional aspect of the predictive power of HG under the restricted CON hypothesis that has gone largely unnoticed in the literature:
Synchronic implications

- OT with restricted $\text{CON}$: factorial typology, unnatural processes unattested (a desired prediction for final voicing)
- HG: An additional aspect of the predictive power of HG under the restricted $\text{CON}$ hypothesis that has gone largely unnoticed in the literature:
  - Natural elements in a given environment will always be more frequent than unnatural ones
Synchronic implications

- OT with restricted $\text{CON}$: factorial typology, unnatural processes unattested (a desired prediction for final voicing)
- HG: An additional aspect of the predictive power of HG under the restricted $\text{CON}$ hypothesis that has gone largely unnoticed in the literature:
  - Natural elements in a given environment will always be more frequent than unnatural ones
- If we allow only natural constraints into $\text{CON}$, we can only derive systems with gradient phonotactic restrictions in which the natural element in a given context is more frequent than the unnatural element
Synchronic implications

- E.g. **Final Voicing**
- Restr. **Con**: *D# ✓ *T# ✗
- Let us assume that all inputs have a uniform prior probability
- HG: $P(/T#/) = P(/D#/) = 0.5$
- If the faithfulness constraint ($\mathcal{F}$) Ident-IO(oui) has a positive infinite weight and the markedness constraint ($\mathcal{M}$) *D# has a finite weight, the phonotactic probabilities of [T#] and [D#] ($P([T#])$ and $P([D#])$) are both 0.5.
- If the markedness constraint is weighted finitely lower than, or even higher than the faithfulness constraint, the phonotactic probability of [T#] will be greater than that of [D#]
Synchronic implications

- With restricted $\text{CON}$, no weighting exists that would yield a system in which the unnatural feature value has a greater posterior probability than the natural one in a given context.

- $w(F) - w(M) = \infty$: $P(\text{nat}) = P(\text{unnat}) = 0.5$
- $w(F) - w(M) < \infty$: $P(\text{nat}) > P(\text{unnat})$
Synchronic implications

Natural Gradience Bias (NGB)

HG with restricted $\text{CON}$ 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.
NGB correctly predicts the major typological trend with regard to gradient phonotactic restrictions: all cases reported previously indeed operate in the natural direction.

- As tacit phonotactic knowledge obtained from experiments, e.g. Albright 2009


Synchronic implications

- However, the Tarma Quechua systems of stop voicing presented in this paper suggest that HG with restricted $\text{CON}$ undergenerates.

- The natural constraints $\text{*NT}$ and $\text{*T[-voice]}$ will not be able to give $[\text{NT}]$ a higher probability than $[\text{ND}]$, or $[\text{TD}]$ a higher probability than $[\text{TT}]$.

- This, in turn suggests, that $\text{CON}$ must contain some unnatural Markedness constraints.

- Other such cases: Berawan (Beguš and Nazarov 2017)

- If we admit all constraints into $\text{CON}$, how to encode rarity of some processes?
Conclusions and future directions

- A case of *unnatural* gradient phonotactic restriction
- Lexical counts, phonetic analysis, signs of productivity
- Unnatural gradient phonotactic restrictions find natural origin: Blurring Chain
- Synchronic implications: NGB
- A challenge to restricted $CON$
- Further experimental work
- Other such cases


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


Thank you!

* We would like to thank Willem Adelaar, Adam Albright, Gaja Jarosz, Jay Jasanoff, Sasha Lubotsky, Joe Pater, and Kevin Ryan for their useful comments on various versions of this work. All mistakes are our own. This research is partially funded by Mind Brain Behavior interfaculty initiative at Harvard University.