Revisiting ‘The role of features in phonological inventories’

Daniel McCloy,¹ Steven Moran²,³ & Richard Wright¹

¹Department of Linguistics, University of Washington
²Linguistics Department, University of Zurich
³Research Unit: Quantitative Language Comparison, University of Munich

2013 January 18
Summary of Clements (2009)

• “The role of features in phonological inventories”. In Raimy & Cairns (eds.) Contemporary views on architecture and representations.

  ◦ Feature bounding
  ◦ Marked feature values
  ◦ Feature economy
  ◦ Robustness
  ◦ Phonological enhancement
Summary of Clements (2009)

• “The role of features in phonological inventories”. In Raimy & Cairns (eds.) *Contemporary views on architecture and representations.*

  ◦ Feature bounding

  ◦ Marked feature values

  ◦ Feature economy

  ◦ Robustness

  ◦ Phonological enhancement
Overview

Background
   About PHOIBLE

Feature Bounding
   Definition
   Clements’s findings
   Results from PHOIBLE
   Summary

Marked feature values
   Definition
   Clements’s findings
   Results from PHOIBLE
   Summary

Discussion
   Acknowledgments
Overview

Background
  About PHOIBLE

Feature Bounding
  Definition
  Clements’s findings
  Results from PHOIBLE
  Summary

Marked feature values
  Definition
  Clements’s findings
  Results from PHOIBLE
  Summary

Discussion
  Acknowledgments
About phoible

- ~1600 phonological inventories (1298 unique)
  - Stanford Phonology Archive (SPA)\textsuperscript{4}
  - UCLA Phonological Segment Inventory Database (UPSID)\textsuperscript{13}
  - \textit{Systèmes alphabétiques des langues africaines} (AA)\textsuperscript{8}
  - Published phonological descriptions / grammars

- Entries include:
  - Symbolic representations of phonemes (superset of IPA)
  - Genealogical, geographic, & demographic data (Ethnologue,\textsuperscript{11} WALS\textsuperscript{9})
  - Vector of feature values for each phoneme
About phoible

• ~1600 phonological inventories (1298 unique)
  ◦ Stanford Phonology Archive (SPA)\(^4\)
  ◦ UCLA Phonological Segment Inventory Database (UPSID)\(^{13}\)
  ◦ *Systèmes alphabétiques des langues africaines* (AA)\(^8\)
  ◦ Published phonological descriptions / grammars

• Entries include:
  ◦ Symbolic representations of phonemes (superset of IPA)
  ◦ Genealogical, geographic, & demographic data (Ethnologue,\(^{11}\) WALS\(^9\))
  ◦ Vector of feature values for each phoneme
Guiding principles of **phoible** development (1 of 2)

- As true to original description as possible (required several additions to IPA)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>D tap</td>
<td>distinguished from flap r following Maddieson(^{12})</td>
</tr>
<tr>
<td>a(^_{f}) fortis</td>
<td></td>
</tr>
<tr>
<td>a(^{l}) lenis</td>
<td></td>
</tr>
<tr>
<td>a(^{r}) frictionalized approximant ⇒ fricative; clicks with fricated anterior release</td>
<td></td>
</tr>
<tr>
<td>a(^{n}) non-strident</td>
<td></td>
</tr>
<tr>
<td>a(^{h}) half-long</td>
<td></td>
</tr>
<tr>
<td>a(^{H}) epiglottalized</td>
<td></td>
</tr>
<tr>
<td>a(^{g}) glottalized used with voiceless consonants, or wherever source implies something other than “creaky”</td>
<td></td>
</tr>
</tbody>
</table>
Guiding principles of *phoible* development (1 of 2)

- As true to original description as possible (required several additions to IPA)

<table>
<thead>
<tr>
<th>D</th>
<th>tap</th>
<th>distinguished from flap /r/ following Maddieson$^{12}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>aᵳ</td>
<td>fortis</td>
<td></td>
</tr>
<tr>
<td>aᵯ</td>
<td>lenis</td>
<td></td>
</tr>
<tr>
<td>aₙ</td>
<td>frictionalized</td>
<td>approximant $\Rightarrow$ fricative; clicks with fricated anterior release</td>
</tr>
<tr>
<td>a</td>
<td>non-strident</td>
<td></td>
</tr>
<tr>
<td>a'</td>
<td>half-long</td>
<td></td>
</tr>
<tr>
<td>aᴴ</td>
<td>epiglottalized</td>
<td></td>
</tr>
<tr>
<td>aʔ</td>
<td>glottalized</td>
<td>used with voiceless consonants, or wherever source implies something other than “creaky”</td>
</tr>
</tbody>
</table>
Guiding principles of Phoible development (2 of 2)

- Unique feature-value vector for each phoneme as described in source (regardless of within-language contrasts)

- Example: feature-value vectors should distinguish
  - S (English)
  - S̪ (Spanish)
  - S̻ (Basque\textsuperscript{15})
  - S̺ (Galician\textsuperscript{16})
Guiding principles of PHOIBLE development (2 of 2)

- Unique feature-value vector for each phoneme as described in source (regardless of within-language contrasts)

- Example: feature-value vectors should distinguish
  - s (English)
  - s (Spanish)
  - s̪ (Basque)
  - s̻ (Galician)

McCloy, Moran & Wright (UW/UZ/UM) Revisiting Clements 2009 2013.01.18 7 / 29
Features in PHOIBLE

- **Currently 37 features**
  - Mostly follows Hayes\(^{10}\) and Moisik & Esling\(^{14}\)

- Hierarchical organization: parent node \([-\text{value}]\) ⇒ child node \([0\text{value}]\)
  - All \([-\text{coronal}]\) segments are \([0\text{anterior}, 0\text{distributed}, 0\text{strident}]\)
  - All \([-\text{dorsal}]\) segments are \([0\text{high}, 0\text{low}, 0\text{front}, 0\text{back}]\)
  - All \([-\text{labial}]\) segments are \([0\text{round}, 0\text{labiodental}]\)
  - 0 values treated as not contrasting with either + or −

- Contour segments: ordered tuple values for certain features
  - Example: \(k\xi\)’ (velar ejective with lateral release)
    - has feature values \([-\text{sonorant}], [-,+\text{continuant}], [-\text{nasal}], [-,+\text{lateral}]\), etc.
    - found in Zulu (Bantoid, Niger-Congo)
Features in phoible

• Currently 37 features
  ◦ Mostly follows Hayes\textsuperscript{10} and Moisik & Esling\textsuperscript{14}

• Hierarchical organization: parent node [−value] ⇒ child node [0value]
  ◦ All [−coronal] segments are [0anterior, 0distributed, 0strident]
  ◦ All [−dorsal] segments are [0high, 0low, 0front, 0back]
  ◦ All [−labial] segments are [0round, 0labiodental]
  ◦ 0 values treated as not contrasting with either + or −

• Contour segments: ordered tuple values for certain features
  ◦ Example: $k\dot{\xi}$ (velar ejective with lateral release)
    ◦ has feature values [−sonorant], [−,+continuant], [−nasal], [−,+lateral], etc.
    ◦ found in Zulu (Bantoid, Niger-Congo)
Features in phoible

- Currently 37 features
  - Mostly follows Hayes\textsuperscript{10} and Moisik & Esling\textsuperscript{14}

- Hierarchical organization: parent node [−value] ⇒ child node [0value]
  - All [−coronal] segments are [0anterior, 0distributed, 0strident]
  - All [−dorsal] segments are [0high, 0low, 0front, 0back]
  - All [−labial] segments are [0round, 0labiodental]
  - 0 values treated as not contrasting with either + or −

- Contour segments: ordered tuple values for certain features
  - Example: $k_l'$ (velar ejective with lateral release)
    - has feature values [−sonorant], [−,+continuant], [−nasal], [−,+lateral], etc.
    - found in Zulu (Bantoid, Niger-Congo)
Phonemes in phoible

- ~2000 distinct segments (~1000 occur in only one language)

  - ʂ (non-strident voiceless retroflex fricative)
    - found in Sa'ban (Malayo-Polynesian, Austronesian)

  - ũ (nasalized creaky high back round vowel)
    - found in Mambay (Adamawa, Niger-Congo)

  - ʈ (glottalized voiceless retroflex stop)
    - found in Siona (Tucanoan)

  - ɬ (simultaneous alveolar/velar voiceless lateral fricative)
    - found in Axluslay/Nivaclé (Matacoan)
Phonemes in PHOIBLE

- ~2000 distinct segments (~1000 occur in only one language)

  - ʂ (non-strident voiceless retroflex fricative)
    - found in Sa'ban (Malayo-Polynesian, Austronesian)

  - ũ (nasalized creaky high back round vowel)
    - found in Mambay (Adamawa, Niger-Congo)

  - ź (glottalized voiceless retroflex stop)
    - found in Siona (Tucanoan)

  - ɬ (simultaneous alveolar/velar voiceless lateral fricative)
    - found in Axluslay/Nivaclé (Matacoan)
Overview

Background
  About PHOIBLE

Feature Bounding
  Definition
  Clements’s findings
  Results from PHOIBLE
  Summary

Marked feature values
  Definition
  Clements’s findings
  Results from PHOIBLE
  Summary

Discussion
  Acknowledgments
Feature bounding: Definition

- Mathematical relationship between segments & features
  - minimum \( \lceil \log_2(n) \rceil \) binary features needed to distinguish \( n \) phonemes
  - Linguistic features rarely orthogonal; actual number of features needed often much higher
Feature bounding: Clements’s findings from UPSID

- **Observation**: coronals in UPSID restricted to 4-way place contrasts at most

- **Prediction**: $[\pm \text{anterior}] \times [\pm \text{distributed}]$ might be enough to capture all (within-language) contrasts

- **Finding**: phonological approach accounts for all coronal contrasts in UPSID
  - Exceptions (Albanian & !Xóõ) rely on secondary features (velarization/pharyngealization & affrication, respectively)
Feature bounding: Clements’s findings from UPSID

- Observation: coronals in UPSID restricted to 4-way place contrasts at most

- Prediction: \([\pm \text{anterior}] \times [\pm \text{distributed}]\) might be enough to capture all (within-language) contrasts

- Finding: phonological approach accounts for all coronal contrasts in UPSID
  - Exceptions (Albanian & !Xóõ) rely on secondary features (velarization/pharyngealization & affrication, respectively)
Feature bounding: Clements’s findings from UPSID

• Observation: coronals in UPSID restricted to 4-way place contrasts at most

• Prediction: $[\pm \text{anterior}] \times [\pm \text{distributed}]$ might be enough to capture all (within-language) contrasts

• Finding: phonological approach accounts for all coronal contrasts in UPSID
  ◦ Exceptions (Albanian & !Xóô) rely on secondary features (velarization/pharyngealization & affrication, respectively)
Feature bounding: Results from PHOIBLE

### Four-way coronal stop contrasts

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Language</th>
<th>Genus, Root</th>
<th>Features needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>t̪ t̠ t̟ t̠</td>
<td>Eastern Arrernte Western Arrarnta Yanyuwa</td>
<td>Pama-Nyungan, Australian</td>
<td></td>
</tr>
<tr>
<td>t̪ t̠ t̟ c</td>
<td>Alyawarra Diyari Yolngu Kalkatungu</td>
<td>Pama-Nyungan, Australian</td>
<td>anterior, distributed</td>
</tr>
<tr>
<td>t̪ t̠ t̟ c̟</td>
<td>Tira Heiban, Niger-Congo</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t̪ t̠ t̟ c̟</td>
<td>Nunggubuyu</td>
<td>Nunggubuyu, Australian</td>
<td></td>
</tr>
<tr>
<td>t̪ t̠ t̟ c</td>
<td>Garawa</td>
<td>Garawan, Australian</td>
<td>anterior, distributed, dorsal</td>
</tr>
<tr>
<td>t̪ t̠ c̟ c</td>
<td>Quechan</td>
<td>Yuman, Hokan</td>
<td>θ &amp; C both [-ant +dist +dors]</td>
</tr>
</tbody>
</table>
## Feature bounding: Results from PHOIBLE

### Attested coronal stop contrast types (1 of 2)

<table>
<thead>
<tr>
<th>Contrast type (à la Clements 2009)</th>
<th>Contrast</th>
<th>Sample language</th>
<th>#lx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>{apical / nonapical} anterior</td>
<td>t̺ t̻</td>
<td>Nez Perce (Sahaptian, Penutian)</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>t t̺</td>
<td>Maung (Iwaidjan, Australian)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>t̺ t̺</td>
<td>Didinga (Surmic, Nilo-Saharan)</td>
<td>1</td>
</tr>
<tr>
<td>{apical / nonapical} posterior</td>
<td>t̺ c̟</td>
<td>Siraiki (Iranian, Indo-European)</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>t̺ t̺</td>
<td>Bardi (Nyulnyulan, Australian)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>t̺ c̟</td>
<td>Alawa (Maran, Australian)</td>
<td>3</td>
</tr>
<tr>
<td>apical {anterior / posterior}</td>
<td>t̺ c̟</td>
<td>Iai (Oceanic, Austronesian)</td>
<td>38</td>
</tr>
<tr>
<td>nonapical {anterior / posterior}</td>
<td>t̺ c̟</td>
<td>Chrau (Bahnaric, Austro-Asiatic)</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>t̺ t̺</td>
<td>Ngiyambaa (Pama-Nyungan, Australian)</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>t̺ c̟</td>
<td>Nunggubuyu (Nunggubuyu, Australian)</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>t̺ c̟</td>
<td>Kunjen (Pama-Nyungan, Australian)</td>
<td>1</td>
</tr>
</tbody>
</table>
# Feature bounding: Results from PHOIBLE

## Attested coronal stop contrast types (2 of 2)

<table>
<thead>
<tr>
<th>Contrast type (à la Clements 2009)</th>
<th>Contrast</th>
<th>Sample language</th>
<th>#lx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>apical anterior / nonapical posterior</td>
<td>( \text{t} ) ( \text{c} )</td>
<td>Shekgalagari (Bantoid, Niger-Congo)</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>( \text{t} ) ( \text{t̟} )</td>
<td>Amuzgo (Amuzgoan, Oto-Manguean)</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>( \text{t} ) ( \text{c̟} )</td>
<td>Campa (Arawakan)</td>
<td>4</td>
</tr>
<tr>
<td>nonapical anterior / apical posterior</td>
<td>( \text{t̻} ) ( \text{t̠} )</td>
<td>Bagirmi (Bongo-Bagirmi, Nilo-Saharan)</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>( \text{t̻} ) ( \text{t} )</td>
<td>Punjabi (Indic, Indo-European)</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast type (à la Clements 2009)</th>
<th>Contrast</th>
<th>Sample language</th>
<th>#lx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>nonapical {anterior / posterior}? (( \text{t̻} ) vs ( \text{c} ) according to Furby)</td>
<td>( \text{t} ) ( \text{c} )</td>
<td>Garawa (Garawan, Australian)</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast type (à la Clements 2009)</th>
<th>Contrast</th>
<th>Sample language</th>
<th>#lx.</th>
</tr>
</thead>
<tbody>
<tr>
<td>“palatal” vs “pre-palatal” according to Halpern</td>
<td>( \text{c̟} ) ( \text{c} )</td>
<td>Quechan (Yuman, Hokan)</td>
<td>1</td>
</tr>
</tbody>
</table>
## Three- and four-way coronal fricative contrasts

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Language</th>
<th>Genus, Root</th>
<th>Features needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>θ ʃ</td>
<td>26 languages</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ʃ ʃ</td>
<td>20 languages</td>
<td>various</td>
<td></td>
</tr>
<tr>
<td>ʃ ʂ</td>
<td>3 languages</td>
<td>anterior, distributed</td>
<td></td>
</tr>
<tr>
<td>ʃ ʃ Basque</td>
<td>Basque</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ʃ ʃ Chimborazo Quichua</td>
<td>Quechuan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ʃ ʃ Serrano</td>
<td>Takic, Uto-Aztecan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ ʃ Galician</td>
<td>Romance, Indo-European</td>
<td></td>
<td></td>
</tr>
<tr>
<td>θ ʃ</td>
<td>Berta</td>
<td>Berta, Nilo-Saharan</td>
<td>anterior, distributed, strident</td>
</tr>
<tr>
<td>θ ʃ Libyan Arabic</td>
<td>Semitic, Afro-Asiatic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Feature bounding: Summary

- Clements’s generalization about coronals holds across >99% of languages.

- Most violations resolved with other common features (dorsal, strident).

- Only remaining violation (Quechan ζ vs c) possibly reanalyzable as t̠ vs c, which [dorsal] serves to distinguish.
Overview

Background
  About PHOIBLE

Feature Bounding
  Definition
  Clements’s findings
  Results from PHOIBLE
  Summary

Marked feature values
  Definition
  Clements’s findings
  Results from PHOIBLE
  Summary

Discussion
  Acknowledgments
Marked feature values: Definition

- **Marked feature value**: fails to occur in some languages, and complementary value never fails to occur
  - Example: [+nasal] consonants are missing from some languages (21 in PHOIBLE)
  - All languages have [−nasal] consonants
  - Therefore [+nasal] is marked (for consonants)

- **Marked segment**: segment that exhibits a marked feature

- **Clements’s predictions**:
  - Languages w/ marked segments will have larger inventories
  - Within a language, marked segments < unmarked segments (constrained to segments for which the marked feature matters)
Marked feature values: Definition

- Marked feature value: fails to occur in some languages, and complementary value never fails to occur
  - Example: [+nasal] consonants are missing from some languages (21 in PHOIBLE)
  - All languages have [−nasal] consonants
  - Therefore [+nasal] is marked (for consonants)

- Marked segment: segment that exhibits a marked feature

- Clements’s predictions:
  - Languages w/ marked segments will have larger inventories
  - Within a language, marked segments < unmarked segments (constrained to segments for which the marked feature matters)
Marked feature values: Definition

- Marked feature value: fails to occur in some languages, and complementary value never fails to occur
  - Example: [+nasal] consonants are missing from some languages (21 in PHOIBLE)
  - All languages have [−nasal] consonants
  - Therefore [+nasal] is marked (for consonants)

- Marked segment: segment that exhibits a marked feature

- Clements’s predictions:
  - Languages w/ marked segments will have larger inventories
  - Within a language, marked segments < unmarked segments (constrained to segments for which the marked feature matters)
Marked feature values: Clements’s findings from upsID

- [+sonorant]
- [+continuant]
- [+nasal]
- [+strident]
- [+posterior]
- [+lateral]
- [+spread glottis]

- [+constricted glottis]
- [+round]
- [+high]
- [+low]
- [+front]
- [+labial]
- [+dorsal]
### Marked feature values for vowels (1 of 2)

<table>
<thead>
<tr>
<th>Marked feat.</th>
<th>What is rare?</th>
<th>Langs. w/o marked feat.</th>
<th>% langs. w/ unmarked segs. &gt; marked segs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>([-\text{voice})]()</td>
<td>Voicing contrast</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>([+\text{coronal})]()</td>
<td>Rhotic vowels</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>([+\text{atr}]/[+\text{rtr})]()</td>
<td>Pharyngeal/ATR contrast</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>([+\text{short})]()</td>
<td>Length contrast</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>(\star[+\text{nasal})]()</td>
<td>Nasal contrast</td>
<td>77%</td>
<td>93%</td>
</tr>
<tr>
<td>(\star[+\text{long})]()</td>
<td>Length contrast</td>
<td>66%</td>
<td>80%</td>
</tr>
</tbody>
</table>

\(\star\) denotes findings mentioned in Clements 2009\(^3\)

3 of the 19 languages requiring [short] also require [long], constituting a three-way length contrast. The languages are South Central Dinka (Nilotic, Nilo-Saharan), Ndut-Falor (Northern Atlantic, Niger-Congo), and Hopi (Hopi, Uto-Aztecan).
### Marked feature values for vowels (1 of 2)

<table>
<thead>
<tr>
<th>Marked feat.</th>
<th>What is rare?</th>
<th>Langs. w/o marked feat.</th>
<th>% langs. w/ unmarked segs. &gt; marked segs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[−voice]</td>
<td>Voicing contrast</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>[+coronal]</td>
<td>Rhotic vowels</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>[+atr]/[+rtr]</td>
<td>Pharyngeal/ATR contrast</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>[+short]</td>
<td>Length contrast</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>★ [+nasal]</td>
<td>Nasal contrast</td>
<td>77%</td>
<td>93%</td>
</tr>
<tr>
<td>★ [+long]</td>
<td>Length contrast</td>
<td>66%</td>
<td>80%</td>
</tr>
</tbody>
</table>

★ denotes findings mentioned in Clements 2009

3 of the 19 languages requiring [short] also require [long], constituting a three-way length contrast. The languages are South Central Dinka (Nilotic, Nilo-Saharan), Ndut-Falor (Northern Atlantic, Niger-Congo), and Hopi (Hopi, Uto-Aztecan).
Marked feature values: Results from *phoible*

### Marked feature values for vowels (2 of 2)

<table>
<thead>
<tr>
<th>Marked feat.</th>
<th>What is rare?</th>
<th>Langs. w/o marked feat.</th>
<th>% langs. w/ unmarked segs. &gt; marked segs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>★ [+labial]</td>
<td>Lack of rounded vowels</td>
<td>1.2%</td>
<td>90%</td>
</tr>
<tr>
<td>★ [+high]</td>
<td>Lack of high vowels</td>
<td>&lt;1%</td>
<td>81%</td>
</tr>
<tr>
<td>★ [+front]</td>
<td>Lack of front vowels</td>
<td>&lt;1%</td>
<td>87%</td>
</tr>
<tr>
<td>[+back]</td>
<td>Lack of back vowels</td>
<td>&lt;1%</td>
<td>88%</td>
</tr>
<tr>
<td>★ [+low]</td>
<td>Lack of low vowels</td>
<td>&lt;1%</td>
<td>99%</td>
</tr>
</tbody>
</table>

★ denotes findings mentioned in Clements 2009\(^3\)
## Marked feature values: Results from PHOIBLE

### Marked feature values for consonants (1 of 2)

<table>
<thead>
<tr>
<th>Marked feat.</th>
<th>What is rare?</th>
<th>Langs. w/o marked feat.</th>
<th>% langs. w/ unmarked segs. &gt; marked segs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+epiConstr]</td>
<td>epiglottal consonants /ʰ ʢ/</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>[+fortis]</td>
<td>plain/fortis contrast</td>
<td>&gt;99%</td>
<td>100%</td>
</tr>
<tr>
<td>[+click]</td>
<td>clicks</td>
<td>99%</td>
<td>100%</td>
</tr>
<tr>
<td>[+syllabic]</td>
<td>syllabic/nonsyllabic contrast</td>
<td>98%</td>
<td>100%</td>
</tr>
<tr>
<td>[+low]</td>
<td>pharyngeal consonants /ʰ ʢ/</td>
<td>96%</td>
<td>100%</td>
</tr>
<tr>
<td>[+long]</td>
<td>length contrast</td>
<td>94%</td>
<td>100%</td>
</tr>
<tr>
<td>[+rLrx] (ejct)</td>
<td>ejactive consonants</td>
<td>89%</td>
<td>100%</td>
</tr>
<tr>
<td>[-high]</td>
<td>uvular/pharyngeal(ized) cons.</td>
<td>84%</td>
<td>99%</td>
</tr>
<tr>
<td>[+lLrx] (impl)</td>
<td>implosive consonants</td>
<td>81%</td>
<td>100%</td>
</tr>
<tr>
<td>[+tap]</td>
<td>tap/flap consonants</td>
<td>72%</td>
<td>100%</td>
</tr>
<tr>
<td>★ [+constrGlot]</td>
<td>glottalized/creaky/ejective cons.</td>
<td>54%</td>
<td>100%</td>
</tr>
<tr>
<td>[+trill]</td>
<td>trilled consonants</td>
<td>52%</td>
<td>100%</td>
</tr>
</tbody>
</table>

★ denotes findings mentioned in Clements 2009³
## Marked feature values for consonants (2 of 2)

<table>
<thead>
<tr>
<th>Marked feat.</th>
<th>What is rare?</th>
<th>Langs. w/o marked feat.</th>
<th>Pct.</th>
<th>Count</th>
<th>% langs. w/ unmarked segs. &gt; marked segs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>[+labiodent]</td>
<td>lack of labiodentals</td>
<td>38%</td>
<td>487</td>
<td></td>
<td>99%</td>
</tr>
<tr>
<td>★ [+spreadGlot]</td>
<td>lack of aspirated cons. or /h/</td>
<td>31%</td>
<td>400</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>★ [−anterior]</td>
<td>lack of retroflex &amp; palatal cons.</td>
<td>14%</td>
<td>177</td>
<td></td>
<td>94%</td>
</tr>
<tr>
<td>[+lateral]</td>
<td>lack of laterals</td>
<td>13%</td>
<td>166</td>
<td></td>
<td>100%</td>
</tr>
<tr>
<td>[+back]</td>
<td>lack of velar/uvular(ized) cons.</td>
<td>11%</td>
<td>137</td>
<td></td>
<td>93%</td>
</tr>
<tr>
<td>★ [+strident]</td>
<td>lack of coronal fricatives/affricates</td>
<td>5.6%</td>
<td>73</td>
<td></td>
<td>82%</td>
</tr>
<tr>
<td>[+front]</td>
<td>lack of palatals &amp; fronted velars</td>
<td>5.2%</td>
<td>67</td>
<td></td>
<td>85%</td>
</tr>
<tr>
<td>[+delayedRel]</td>
<td>lack of fricatives</td>
<td>3.2%</td>
<td>42</td>
<td></td>
<td>70%</td>
</tr>
<tr>
<td>★ [+nasal]</td>
<td>lack of nasal cons.</td>
<td>1.6%</td>
<td>21</td>
<td></td>
<td>99%</td>
</tr>
<tr>
<td>★ [−voice]</td>
<td>lack of voiceless cons.</td>
<td>&lt;1%</td>
<td>6</td>
<td></td>
<td>76%</td>
</tr>
<tr>
<td>★ [+sonorant]</td>
<td>lack of approximants &amp; nasals</td>
<td>&lt;1%</td>
<td>1</td>
<td></td>
<td>87%</td>
</tr>
<tr>
<td>★ [+labial]</td>
<td>lack of labial &amp; rounded cons.</td>
<td>&lt;1%</td>
<td>1</td>
<td></td>
<td>97%</td>
</tr>
<tr>
<td>★ [+dorsal]</td>
<td>lack of dorsals</td>
<td>&lt;1%</td>
<td>1</td>
<td></td>
<td>95%</td>
</tr>
</tbody>
</table>

★ denotes findings mentioned in Clements 2009³
Marked feature values: Summary

- Many more marked feature values than discussed by Clements

- [+round] not marked in PHOIBLE
  - 164 langs. lack [+round] consonants, but 2 langs. lack [−round] consonants (lack /p b m f v/, but have rounded/unrounded dorsals)

- Feature geometry/hierarchy has important implications for markedness results
Overview

Background
About PHOIBLE

Feature Bounding
Definition
Clements’s findings
Results from PHOIBLE
Summary

Marked feature values
Definition
Clements’s findings
Results from PHOIBLE
Summary

Discussion
Acknowledgments
Future directions

- Feature Economy
  - Feature set expansion to close remaining gaps
  - By-language dimensionality reduction to discover “optimal” feature sets

- Markedness
  - Alternative calculation based on cross-linguistic occurrence of features in economy-optimized, language-specific feature subsets

- Robustness, phonological enhancement

- PHOIBLE development
  - Expand language coverage
  - Allow alternative feature systems to be swapped in easily
  - Interface with lexical data (cognate identification, feature weighting)
Future directions

• Feature Economy
  ◦ Feature set expansion to close remaining gaps
  ◦ By-language dimensionality reduction to discover “optimal” feature sets

• Markedness
  ◦ Alternative calculation based on cross-linguistic occurrence of features in economy-optimized, language-specific feature subsets

• Robustness, phonological enhancement

• PHOIBLE development
  ◦ Expand language coverage
  ◦ Allow alternative feature systems to be swapped in easily
  ◦ Interface with lexical data (cognate identification, feature weighting)
Future directions

• Feature Economy
  ◦ Feature set expansion to close remaining gaps
  ◦ By-language dimensionality reduction to discover “optimal” feature sets

• Markedness
  ◦ Alternative calculation based on cross-linguistic occurrence of features in economy-optimized, language-specific feature subsets

• Robustness, phonological enhancement

• PHOIBLE development
  ◦ Expand language coverage
  ◦ Allow alternative feature systems to be swapped in easily
  ◦ Interface with lexical data (cognate identification, feature weighting)
Future directions

• Feature Economy
  ◦ Feature set expansion to close remaining gaps
  ◦ By-language dimensionality reduction to discover “optimal” feature sets

• Markedness
  ◦ Alternative calculation based on cross-linguistic occurrence of features in economy-optimized, language-specific feature subsets

• Robustness, phonological enhancement

• PHOIBLE development
  ◦ Expand language coverage
  ◦ Allow alternative feature systems to be swapped in easily
  ◦ Interface with lexical data (cognate identification, feature weighting)
References


Acknowledgments

• Many thanks to the organizers and participants of the CUNY Conference on the Feature!

• Thanks to contributors of PHOIBLE: Emily Bender, Morgana Davids, Scott Drellishak, David Ellison, Christopher Green, Richard John Harvey, Kelley Kilanski, Michael McAuliffe, Kevin Pittman, Brandon Plasters, Tristan Purvis, Cameron Rule, Daniel Smith, and Daniel Veja, as well as Marilyn Vihman for providing the Stanford Phonology Archive data.

• The development of PHOIBLE was partially funded by a grant from the University of Washington Royalty Research Fund.