Soft Typology of Coda Place of Articulation Distributions Requires Synchronic Constraints

Charlie O’Hara
University of Southern California

Workshop on Emergence of Universals
February 19, 2018
Crucial goal of linguistic theory is to explain generalizations observed across typology.

- If a pattern is unattested, either:
  - The pattern is banned by Universal Grammar.
  - The pattern is unlikely to be innovated or retained due to other factors.
Emergent Biases

Why would a pattern be less likely than another pattern?

- **Inductive Bias:** It is more difficult to learn, even with pristine training data.

- **Channel Bias:** It is more difficult to learn, due to asymmetries in mistransmission of training data.
Structure vs. Substance

Biases can be divided based on their effects:

- **Structural Bias**—A bias for featurally simple patterns. i.e. A pattern based on one feature is preferred to one based on two.

- **Substantive Bias**—Bias based on the substance of the features rather than abstract complexity. i.e. One pattern based on only one feature is preferred to another based only on some other feature.
Structure vs. Substance

Biases can be divided based on their effects:

- **Structural Bias**—A bias for featurally simple patterns. i.e. A pattern based on one feature is preferred to one based on two.

- **Substantive Bias**—Bias based on the substance of the features rather than abstract complexity. i.e. One pattern based on only one feature is preferred to another based only on some other feature.

<table>
<thead>
<tr>
<th>SIMPLE</th>
<th>COMPLEX</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Coda</em></td>
<td><em>[Coda+Dorsal]</em></td>
</tr>
<tr>
<td>kV  pV  tV &gt; kV  pV  tV</td>
<td></td>
</tr>
<tr>
<td>Vk  Vp  Vt</td>
<td>Vk  Vp  Vt</td>
</tr>
</tbody>
</table>
Structure vs. Substance

Biases can be divided based on their effects:

- **Structural Bias**—A bias for featurally simple patterns. i.e. A pattern based on one feature is preferred to one based on two.

- **Substantive Bias**—Bias based on the substance of the features rather than abstract complexity. i.e. One pattern based on only one feature is preferred to another based only on some other feature.

```
SIMPLE
*Coda
kV \ pV \ tV > kV \ pV \ tV
Vk \ Vp \ Vt
```

```
SIMPLE
*Dorsal

Vk \ Vp \ Vt
```
Structurally Biased Phonology

How do these biases interact?

- A strong hypothesis: (Pater & Moreton (2012)’s Structurally Biased Phonology)

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Channel</td>
<td>?</td>
<td>✓</td>
</tr>
</tbody>
</table>

- Structural inductive bias:
  - Appears similarly in other domains of pattern-learning (Moreton & Pertsova, 2014)
  - Is well documented in artificial grammar learning experiments (Moreton & Pater, 2012)

- No Substantive Inductive Bias:
  - Less evidence that substantive bias emerges in artificial grammar learning (Moreton & Pater, 2012; Glewwe, 2017)
  - Much of substantive bias can be explained through channel bias, without requiring any innate biases (Blevins, 2004)
Substantive Bias in place of articulation inventories

Typology of the interaction of place of articulation and syllable-position.

- This domain shows two types of biases:
  - **Structural**: Patterns based on just syllable-position are better attested than conjunctions of syllable-position and place of articulation.
  - **Anti-Structural**: Patterns based on conjunctions of syllable-position and place of articulation are better attested than those defined only with place of articulation.

```
  SIMPLE  COMPLEX  SIMPLE
  *Coda    *[Coda+Dorsal]  *Dorsal

  kV  pV  tV  >  kV  pV  tV  >  kV  pV  tV
  Vk  Vp  Vt  Vk  Vp  Vt  Vk  Vp  Vt
```

Substantive Bias in place of articulation inventories

Typology of the interaction of place of articulation and syllable-position.

- This domain shows two types of biases:
  - **Structural:** Patterns based on just syllable-position are better attested than conjunctions of syllable-position and place of articulation.
  - **Anti-Structural:** Patterns based on conjunctions of syllable-position and place of articulation are better attested than those defined only with place of articulation.

<table>
<thead>
<tr>
<th>SIMPLE</th>
<th>COMPLEX</th>
<th>SIMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Coda</td>
<td>*[Coda+Dorsal]</td>
<td>*Dorsal</td>
</tr>
<tr>
<td>kV pV tV</td>
<td>kV pV tV</td>
<td>kV pV tV</td>
</tr>
<tr>
<td>Vk Vp Vt</td>
<td></td>
<td>Vk Vp Vt</td>
</tr>
</tbody>
</table>
Substantive Bias in place of articulation inventories

Typology of the interaction of place of articulation and syllable-position.

- This domain shows two types of biases:
  - **Structural**: Patterns based on just syllable-position are better attested than conjunctions of syllable-position and place of articulation.
  - **Anti-Structural**: Patterns based on conjunctions of syllable-position and place of articulation are better attested than those defined only with place of articulation.

```
SIMPLE  COMPLEX  SIMPLE
*Coda    *[Coda+Dorsal]  *Dorsal
kV  pV  tV  >  kV  pV  tV  >  kV  pV  tV
Vk  Vp  Vt  Vk  Vp  Vt  Vk  Vp  Vt
```
Substantive Bias in place of articulation inventories

Typology of the interaction of place of articulation and syllable-position.

- **This domain shows two types of biases:**
  - **Structural:** Patterns based on just syllable-position are better attested than conjunctions of syllable-position and place of articulation.
  - **Anti-Structural:** Patterns based on conjunctions of syllable-position and place of articulation are better attested than those defined only with place of articulation.

<table>
<thead>
<tr>
<th>SIMPLE</th>
<th>COMPLEX</th>
<th>SIMPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Coda</em></td>
<td><em>[Coda+Dorsal]</em></td>
<td><em>Dorsal</em></td>
</tr>
<tr>
<td>kV</td>
<td>pV</td>
<td>tV</td>
</tr>
<tr>
<td>Vk</td>
<td>Vp</td>
<td>Vt</td>
</tr>
</tbody>
</table>
Explaining lack of No-Dorsals

The anti-structural bias against the *Dorsal pattern must be substantive.

- This bias does emerge from a model that allows for both substantive and structural inductive bias.
- This bias does not emerge from a model where structural bias is inductive bias, and substantive bias is channel bias.

Do we need substantive inductive bias?

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>Channel</td>
<td>?</td>
<td>✓</td>
</tr>
</tbody>
</table>
Explaining lack of No-Dorsals

The anti-structural bias against the *Dorsal pattern must be substantive.

- This bias does emerge from a model that allows for both substantive and structural inductive bias.
- This bias does not emerge from a model where structural bias is inductive bias, and substantive bias is channel bias.

Do we need substantive inductive bias?

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
<td>✓?</td>
</tr>
<tr>
<td>Channel</td>
<td>?</td>
<td>✓</td>
</tr>
</tbody>
</table>
Explaining lack of No-Dorsals

The anti-structural bias against the *Dorsal pattern must be substantive.

- This bias *does* emerge from a model that allows for both substantive and structural inductive bias.

- This bias *does not* emerge from a model where structural bias is inductive bias, and substantive bias is channel bias.

Do we need substantive inductive bias?

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Channel</td>
<td>✉</td>
<td>✓</td>
</tr>
</tbody>
</table>
Explaining lack of No-Dorsals

The anti-structural bias against the *Dorsal pattern must be substantive.

- This bias does emerge from a model that allows for both substantive and structural inductive bias.
- This bias does not emerge from a model where structural bias is inductive bias, and substantive bias is channel bias.

Do we need substantive inductive bias?

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
<td>✓?</td>
</tr>
<tr>
<td>Channel</td>
<td>?</td>
<td>✓</td>
</tr>
</tbody>
</table>
Typological Survey

Development of the Word Edge Consonant Database (WECD)

- **172 Languages Total**
- Word-initial and word-final consonants recorded.
- Languages with no consonants (of any sort) in word-final position were not included.
- **Focus on 77 languages with just [k p t] initially.**
  - Ignoring languages that lack [p] but have [b] due to voice-place effects (Hayes & Steriade, 2004).

1. That allow maximally three supralaryngeal places of articulation for stops.
Typological Survey

Development of the Word Edge Consonant Database (WECD)

- 172 Languages Total
- Word-initial and word-final consonants recorded.
- Languages with no consonants (of any sort) in word-final position were not included.
- Focus on 77 languages with just [k p t] initially.¹
- Ignoring languages that lack [p] but have [b] due to voice-place effects (Hayes & Steriade, 2004).

¹That allow maximally three supralaryngeal places of articulation for stops
Markedness Hierarchies

Languages in my typological study show tendencies that replicate predictions made elsewhere about these scales.

- Onset vs. coda \(^2\)
  \[ CV \succ VC \]

- Place of articulation \(^3\)
  \[ \text{Coronal} \succ \frac{10}{13} \quad \text{Labial} \succ \frac{12}{14} \quad \text{Dorsal} \]

---

\(^2\) (Jakobson & Halle, 1956; Kingston, 1985; Goldsmith, 1990)

\(^3\) (de Lacy, 2006; Kean, 1975; Lombardi, 2001)
Results of Typological Survey

Four structurally distinct word-final inventories are available for languages with all of [k p t] word initially.

<table>
<thead>
<tr>
<th></th>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Codas</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td>1 Coda</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td>2 Codas</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td>3 Codas</td>
<td>tV</td>
<td>pV</td>
</tr>
</tbody>
</table>
No Coda Pattern

<table>
<thead>
<tr>
<th>Codas</th>
<th>Onset</th>
<th>Coda</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
<tr>
<td>1 Coda</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
<tr>
<td>2 Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
<tr>
<td>3 Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
</tbody>
</table>

- Example: Italian

[tasto] button  [pasto] meal  [kasto] chaste
[kasat]  [kasap]  [kasak]
Only 1 Coda Pattern

<table>
<thead>
<tr>
<th>Onset</th>
<th>Coda</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Codas</td>
<td>tV</td>
</tr>
<tr>
<td>1 Coda</td>
<td>tV</td>
</tr>
<tr>
<td>2 Codas</td>
<td>tV</td>
</tr>
<tr>
<td>3 Codas</td>
<td>tV</td>
</tr>
</tbody>
</table>

Example: Finnish

[telata] *to paint* [pelata] *to play* [kelata] *to wind*

[keot] *anthills* *[keop] *[keok]*

---

4 Nimboran (Anceaux, 1965) has only [p] word finally.
# 2 Codas Pattern

<table>
<thead>
<tr>
<th>No Codas</th>
<th>Onset</th>
<th>Coda</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
<tr>
<td>1 Coda</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
<tr>
<td>2 Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
<tr>
<td>3 Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
</tr>
</tbody>
</table>

- Example: Movima (Haude, 2006)\(^5\)

  - \(['tanna]\) *I cut*
  - \([\text{pεnna}]\) *my landing place*
  - \([\text{tʃu:'hat}]\) *palm tree*
  - \([\text{ku:'dup}]\) *flea*
  - \([\text{kanan}]\) *your food*
  - \(*[\text{ku:'dʊk}]\) *flea*

---

\(^5\) Korowai (Enk & Vries, 1997) and Navaho Haile (1926 (1974)) show \([\text{k p}]\) and \([\text{k t}]\) respectively.
## 3 Codas Pattern

<table>
<thead>
<tr>
<th>No Codas</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>X</th>
<th>X</th>
<th>X</th>
<th>27</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Coda</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>Vt</td>
<td>X</td>
<td>X</td>
<td>2</td>
</tr>
<tr>
<td>2 Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>Vt</td>
<td>Vp</td>
<td>X</td>
<td>5</td>
</tr>
<tr>
<td>3 Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>Vt</td>
<td>Vp</td>
<td>Vk</td>
<td>43</td>
</tr>
</tbody>
</table>

- **Example:** English

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[pat]</td>
<td>pot</td>
<td>[pap]</td>
<td>pop</td>
<td>[pak]</td>
<td>pock</td>
</tr>
</tbody>
</table>
### Structural Bias

There is a soft generalization favoring the patterns with either all or none of the codas.

<table>
<thead>
<tr>
<th>Coda</th>
<th>Onset</th>
<th></th>
<th>Coda</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No Codas</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>x</td>
</tr>
<tr>
<td>1 Coda</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>Vt</td>
</tr>
<tr>
<td>2 Coda</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>Vt</td>
</tr>
<tr>
<td>3 Coda</td>
<td>tV</td>
<td>pV</td>
<td>kV</td>
<td>Vt</td>
</tr>
</tbody>
</table>

Simplicity bias predicts these generalizations.
Anti-Structural Bias

Simplicity also predicts that the pattern that bans one place in all positions will be well attested.

\[*[\text{Coda}+\text{Dorsal}]\]

\[\begin{array}{ccc}
kV & pV & tV \\
\forall k & Vp & Vt \\
\end{array}\]

\[*\text{Dorsal}\]

\[\begin{array}{ccc}
kV & pV & tV \\
\forall k & Vp & Vt \\
\end{array}\]

\[*[\text{Dorsal}\lor\text{Coda}]\]

\[\begin{array}{ccc}
kV & pV & tV \\
\forall k & Vp & Vt \\
\end{array}\]

- No languages that lack one of [p t k] initially allows any of them finally
Anti-Structural Bias

Simplicity also predicts that the pattern that bans one place in all positions will be well attested.

\[ *[\text{Coda} + \text{Dorsal}] \quad \Rightarrow \quad *[\text{Dorsal}] \quad < \quad *[\text{Dorsal} \lor \text{Coda}] \]

- No languages that lack one of [p t k] initially allows any of them finally

<table>
<thead>
<tr>
<th>Initial</th>
<th>Final</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>p t ?</td>
<td>∅</td>
<td>Tahitian, Vanimo, Wutung, Xavante</td>
</tr>
<tr>
<td>p t</td>
<td>∅</td>
<td>Nouri</td>
</tr>
<tr>
<td>k p ?</td>
<td>∅</td>
<td>Hawaiian, Yellowknife Chipewyan, Colloquial Samoan</td>
</tr>
<tr>
<td>k t ?</td>
<td>?</td>
<td>Ayulta Mixtec</td>
</tr>
</tbody>
</table>

Typological Generalization

Bias against losing a place of articulation
Results of Typological Survey

Structural

<table>
<thead>
<tr>
<th>Stops</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>kV pV tV</td>
<td>43</td>
<td>5</td>
</tr>
<tr>
<td>Vk Vp Vt</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Anti-Structural

<table>
<thead>
<tr>
<th>Stops</th>
<th>Simple</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td>kV pV tV</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Vk Vp Vt</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>
Modeling

To see whether the substantive bias can be explained through inductive or channel biases three simulations were run.

- A model with substantively biased constraints in the grammar (O’Hara, 2017, 2018, submitted)
- A model with no substantive bias
- A model with substantive bias only in the channel.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
</tr>
<tr>
<td>Channel</td>
<td>✗</td>
</tr>
</tbody>
</table>
Modeling

To see whether the substantive bias can be explained through inductive or channel biases three simulations were run.

- A model with substantively biased constraints in the grammar (O'Hara, 2017, 2018, submitted)
- A model with no substantive bias
- A model with substantive bias only in the channel.

<table>
<thead>
<tr>
<th></th>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Channel</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
Modeling

To see whether the substantive bias can be explained through inductive or channel biases three simulations were run.

- A model with substantively biased constraints in the grammar (O'Hara, 2017, 2018, submitted)
- A model with no substantive bias
- A model with substantive bias only in the channel.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
</tr>
<tr>
<td>Channel</td>
<td>✗</td>
</tr>
</tbody>
</table>
Modeling

To see whether the substantive bias can be explained through inductive or channel biases three simulations were run.

- A model with substantively biased constraints in the grammar (O’Hara, 2017, 2018, submitted)
- A model with no substantive bias
- A model with substantive bias only in the channel.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Substance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inductive</td>
<td>✓</td>
</tr>
<tr>
<td>Channel</td>
<td>x</td>
</tr>
</tbody>
</table>
Generational Transmission

- Rate of attestation is related to stability across generations.\(^6\)
- Generational Stability Model:
  - A stable run is one which is closer to \(P\) than any other pattern after \(x\) generations of \(y\) iterations.

\(^6\) Staubs (2014); Kirby (2017); Hughto (2018)
Generational Transmission

- Rate of attestation is related to stability across generations.\(^6\)
- Generational Stability Model:
  - A stable run is one which is closer to P than any other pattern after \(x\) generations of \(y\) iterations.

\[\begin{align*}
P &\rightarrow \text{Inductive Bias} \\
&\rightarrow \text{Channel Bias} \\
\end{align*}\]

---

\(^6\) Staubs (2014); Kirby (2017); Hughto (2018)
Learners

I model learners using MaxEnt grammars\textsuperscript{7} with the Perceptron algorithm\textsuperscript{8}.

- Learners have some fixed set of constraints in their grammar.
- Probabilities are assigned to output candidates based on weighted violation of the constraints.
- Upon observing data that would not match the learner’s data, update constraint weights according to the difference in violation profiles.
- Initialized with markedness constraints high, and faithfulness low.\textsuperscript{9}

\textsuperscript{7} Goldwater & Johnson (2003)
\textsuperscript{8} Rosenblatt (1958); Boersma & Pater (2016)
\textsuperscript{9} Jesney & Tessier (2011)
Substantive Bias can enter the system in two ways:

- Through constraint set:
  - Substantively biased set:
    - \(*k, *kp, *kpt, NoCoda, Onset, Max*
  - Non-Substantively biased set:
    - \(*k, *p, *t, *kpt, NoCoda, NoOnset, Max*

- Through asymmetries in misperception:
  - Substantively biased channel:
    - Mutate the data presented to a learner.
  - Un-biased channel:
    - Let learners receive pristine data.
Substantively biased constraints

With substantively biased constraints, both the structural, and anti-structural bias are captured.\textsuperscript{10}

<table>
<thead>
<tr>
<th></th>
<th>kV</th>
<th>pV</th>
<th>tV</th>
<th>Simple</th>
<th>(\frac{42}{50} = 84%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural</td>
<td>Vk</td>
<td>Vp</td>
<td>Vt</td>
<td>Complex</td>
<td>(\frac{24}{50} = 48%)</td>
</tr>
<tr>
<td>Anti-Structural</td>
<td>Vk</td>
<td>Vp</td>
<td>Vt</td>
<td>Simple</td>
<td>(\frac{0}{50} = 0%)</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td>Vp</td>
<td>Vt</td>
<td>Complex</td>
<td>(\frac{46}{50} = 92%)</td>
</tr>
</tbody>
</table>

\textsuperscript{10} Simulations run for 25 generations of 3200 iterations at learning rate .05.
Substantively biased constraints

**Speed vs. Distance**

- Starting distance = Initial grammatical bias
- Speed = expected rate of change of all violated constraints

![Graph showing relationships between variables](image.png)
Substantively biased constraints

**Speed vs. Distance**

- Starting distance = Initial grammatical bias
- Speed = expected rate of change of all violated constraints
Substantively biased constraints

**Speed vs. Distance**

- Starting distance = Initial grammatical bias
- Speed = expected rate of change of all violated constraints
No Substantive Bias

- Neutral constraint set and no misperception\textsuperscript{11}
- Stability results confirm structural bias towards No-Dorsals.

\begin{itemize}
  \item \textbf{Structural}
    \begin{itemize}
      \item Simple \[ \frac{20}{20} = 100\% \]
      \item Complex \[ \frac{5}{20} = 25\% \]
    \end{itemize}
  \item \textbf{Anti-Structural}
    \begin{itemize}
      \item Simple \[ \frac{18}{20} = 90\% \]
      \item Complex \[ \frac{5}{20} = 25\% \]
    \end{itemize}
\end{itemize}

\textsuperscript{11}Simulations run for 20 generations of 2000 iterations at learning rate .05
Substance-Free Grammar

- Starting distance is equal for all forms
- Speeds can also differ based on rates of misperception

\[
\begin{align*}
&kV \quad pV \quad tV \quad Vk \quad Vp \quad Vt \\
&\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad \downarrow
\end{align*}
\]
Substance-Free Grammar

- Starting distance is equal for all forms
- Speeds can also differ based on rates of misperception

\[
\begin{align*}
&v_k \\
&v_p \\
&v_t \\
&v_k \\
&v_p \\
&v_t \\
&v_k \\
&v_p \\
&v_t \\
&v_k \\
&v_p \\
&v_t
\end{align*}
\]
Substance-Free Grammar

- Starting distance is equal for all forms
- Speeds can also differ based on rates of misperception

\[ kV \quad pV \quad tV \quad Vk \quad Vp \quad Vt \]
Substance-Free Grammar

- Starting distance is equal for all forms
- Speeds can also differ based on rates of misperception

\[ kV \quad pV \quad tV \quad V_k \quad V_p \quad V_t \]
No Substantive Bias

Substance-Free Grammar

- Starting distance is equal for all forms
- Speeds can also differ based on rates of misperception

\[
\begin{align*}
&\downarrow & &\downarrow & &\downarrow & &\downarrow & &\downarrow \\
kV & pV & tV & V_k & V_p & V_t
\end{align*}
\]
Substance-Free Grammar

- Starting distance is equal for all forms
- Speeds can also differ based on rates of misperception

\[ \text{kV} \downarrow \text{pV} \downarrow \text{tV} \downarrow \text{Vk} \downarrow \text{Vp} \downarrow \text{Vt} \downarrow \]
Between the teacher and learners, there is a probability of mistransmission for each sound.

VC][CV Confusion Matrix
English Speakers
(adapted from Redford & Diehl (1999))

<table>
<thead>
<tr>
<th></th>
<th>p</th>
<th>t</th>
<th>k</th>
<th>∅</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>0.96</td>
<td>0.0095</td>
<td>0.015</td>
<td>0.016</td>
</tr>
<tr>
<td>t</td>
<td>0.039</td>
<td>0.934</td>
<td>0.010</td>
<td>0.016</td>
</tr>
<tr>
<td>k</td>
<td>0.021</td>
<td>0.015</td>
<td>0.948</td>
<td>0.016</td>
</tr>
<tr>
<td>Final</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p</td>
<td>0.697</td>
<td>0.099</td>
<td>0.135</td>
<td>0.070</td>
</tr>
<tr>
<td>t</td>
<td>0.083</td>
<td>0.766</td>
<td>0.060</td>
<td>0.091</td>
</tr>
<tr>
<td>k</td>
<td>0.030</td>
<td>0.042</td>
<td>0.885</td>
<td>0.043</td>
</tr>
</tbody>
</table>
### Results: Substantively biased channel

With substantively biased channel, neither the structural, nor anti-structural bias are captured.$^{12}$

<table>
<thead>
<tr>
<th></th>
<th>kV</th>
<th>pV</th>
<th>tV</th>
<th>Simple</th>
<th>5/50 = 10%</th>
<th>Complex</th>
<th>37/50 = 74%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural</td>
<td>kV</td>
<td>pV</td>
<td>tV</td>
<td>Simple</td>
<td>31/50 = 62%</td>
<td>Complex</td>
<td>31/50 = 62%</td>
</tr>
<tr>
<td>Anti-Structural</td>
<td>kV</td>
<td>pV</td>
<td>tV</td>
<td>Simple</td>
<td>31/50 = 62%</td>
<td>Complex</td>
<td>0/50 = 0%</td>
</tr>
</tbody>
</table>

$^{12}$ Simulations run for 10 generations of 4000 iterations at learning rate .05
What error rates would be needed?

To ensure that these results are not due to some specific property of the confusion matrix used above, I searched across different rates of misperception.

- Two challenges:
  - Capture simplicity bias for syllable-position.
  - Capture anti-structural bias for place of articulation.
Substantively biased channel

What error rates would be needed?

To ensure that these results are not due to some specific property of the confusion matrix used above, I searched across different rates of misperception.

- Two challenges:
  - Capture simplicity bias for syllable-position.
  - Capture anti-structural bias for place of articulation.
What error rates would be needed?

To ensure that these results are not due to some specific property of the confusion matrix used above, I searched across different rates of misperception.

- Two challenges:
  - Capture simplicity bias for syllable-position.
  - Capture anti-structural bias for place of articulation.

$\sim$Actual Confusion Matrix Data
What error rates would be needed?

To ensure that these results are not due to some specific property of the confusion matrix used above, I searched across different rates of misperception.

- **Two challenges:**
  - Capture simplicity bias for syllable-position.
  - Capture anti-structural bias for place of articulation.

\[
\begin{array}{cccccc}
pV & tV & Vp & Vt & pV & tV & Vp & Vt \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\end{array}
\]

**Needed Misperception Rates**
What error rates would be needed?

To ensure that these results are not due to some specific property of the confusion matrix used above, I searched across different rates of misperception.

- Two challenges:
  - Capture simplicity bias for syllable-position.
  - Capture anti-structural bias for place of articulation.

\[
\begin{array}{cccccc}
pV & tV & Vp & Vt & pV & Vp \\
\downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\end{array}
\]

40% Needed Misperception Rates
Problems with high rates of misperception

- An upper bound of 60% correct on coda consonants is problematic.
  - This rate of misperception is greater than what is observed in experiments I've seen.
  - Already, the rate of misperception introduced in the confusion matrices causes the instability of the pattern that allowed all of [t p k] in all positions.
    - Increasing the rate of misperception of coda consonants would only make that pattern harder to learn.
Conclusion

- There is an anti-structural bias against patterns that are defined simply on the place of articulation scale:
  - This is emergent with constraints defining the markedness hierarchies.
  - Not directly emergent from channel bias.

- Substantive inductive bias naturally captures the typology, in a way that channel bias cannot.
  - This suggests that substantive inductive bias is needed to capture this generalization.
  - Can syllable-position and place of articulation be treated as different types of features?

- BUT my model of channel bias was a very simple one
- Any ideas of a more elaborate/realistic model that might perform better?
Conclusion

- There is an anti-structural bias against patterns that are defined simply on the place of articulation scale:
  - This is emergent with constraints defining the markedness hierarchies.
  - Not directly emergent from channel bias.
- Substantive inductive bias naturally captures the typology, in a way that channel bias cannot:
  - This suggests that substantive inductive bias is needed to capture this generalization.
  - Can syllable-position and place of articulation be treated as different types of features?
  - BUT my model of channel bias was a very simple one
  - Any ideas of a more elaborate/realistic model that might perform better?
Thank You!

Special thanks go to:

- Karen Jesney, Rachel Walker,
- Audiences at USC: Hayeun Jang, Caitlin Smith, Dani Byrd, Louis Goldstein, Khalil Iskarous, Stephanie Shih, Jason Zevin, Monica Do, Jordan Ackerman, Yifan Yang, Yoona Yee.
- Audiences at AMP 2018, and LSA 2018, and everyone else whose discussion has helped me think about these issues: Bruce Hayes, Coral Hughto, Eleanor Glewewe, John Goldsmith, Andrew Lamont, Elliott Moreton, Joe Pater, Katya Pertsova, Betsy Pillion, Jason Riggle.


Hayes, Bruce, & Steriade, Donca. 2004. Introduction: The phonetic basis of phonological markedness. Pages 1–32 of: Hayes, Bruce, Kirchner, Robert, & Steriade, Donca (eds), *Phonetically-Based Phonology*. Cambridge: Cambridge University Press.


Works Cited III


### Languages

<table>
<thead>
<tr>
<th>Word-Final</th>
<th>Number</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>[p t]</td>
<td>3</td>
<td>Bahasa Indonesia, Itzaj Mayan, Movima</td>
</tr>
<tr>
<td>[t]</td>
<td>1</td>
<td>Finnish</td>
</tr>
<tr>
<td>[k p]</td>
<td>1</td>
<td>Korowai</td>
</tr>
<tr>
<td>[k t]</td>
<td>1</td>
<td>Navaho</td>
</tr>
<tr>
<td>[p]</td>
<td>1</td>
<td>Nimboran</td>
</tr>
<tr>
<td>[k]</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>77</strong></td>
<td></td>
</tr>
</tbody>
</table>
Marked-Faithfulness

- How do we capture gapped inventories?
  - [k p] - Korowai
  - [k t] - Navaho
  - [p] - Nimboran
- By including Marked-Faithfulness constraints (de Lacy, 2006), these can be captured.
  - Max-K, Max-KP, Max-KPT.
- Gapped inventories require high weightings of specific marked faithfulness constraints (i.e. Max-K), which are harder to learn, due to the low initial weight + low number of forms that would cause such an update.
- More in Future work.