Rare Hard-To-Learn Patterns Stably Learned Due To Language-Specific Lexical Frequencies

Charlie O’Hara
University of Southern California

Stanford University
September 22, 2018
Introduction

A major goal of phonological theory is to develop a model that can capture the attested phonological patterns while not vastly over-predicting.

- Constraint based grammars (Optimality Theory\(^1\), Harmonic Grammar\(^2\), etc.) make strong typological predictions through **Factorial Typology**
- Recently, an abundance of work\(^3\) has investigated the hypothesis that learnability affects both categorical and soft typology.

---

\(^1\) Prince & Smolensky (1993/2004); McCarthy & Prince (1995)

\(^2\) Legendre *et al.* (1990); Pater (2016)

\(^3\) Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O’Hara (2018)
A major goal of phonological theory is to develop a model that can capture the attested phonological patterns while not vastly over-predicting.

- Constraint based grammars (Optimality Theory\(^1\), Harmonic Grammar\(^2\), etc.) make strong typological predictions through **Factorial Typology**
- Recently, an abundance of work\(^3\) has investigated the hypothesis that learnability affects both categorical and soft typology.

---

\(^1\)Prince & Smolensky (1993/2004); McCarthy & Prince (1995)
\(^2\)Legendre *et al.* (1990); Pater (2016)
\(^3\)Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O’Hara (2018)
Learnability Filter on Typology

Small asymmetries in learning across one generation can result in large changes to typology over time.\textsuperscript{4}

- The \textit{harder} a pattern is to learn, the more likely learners are to accidentally learn a different pattern.
- If one pattern is \textit{mislearned} more frequently than it is \textit{accidentally learned}, it will become less attested across many generations of learning.

\[ P \xrightarrow{\text{mislearning}} P' \]

\footnotesize{\textsuperscript{4} Bell (1971); Greenberg (1978); Kirby & Huford (2002)}
Typology and Stability

Learnability Filter on Typology

Small asymmetries in learning across one generation can result in large changes to typology over time.\(^4\)

- The *harder* a pattern is to learn, the more likely learners are to accidentally learn a different pattern.

- If one pattern is *mislearned* more frequently than it is *accidentally learned*, it will become less attested across many generations of learning.

\[^4\text{Bell (1971); Greenberg (1978); Kirby \\& Huford (2002)}\]
Small asymmetries in learning across one generation can result in large changes to typology over time.\(^4\)

- The *harder* a pattern is to learn, the more likely learners are to accidentally learn a different pattern.

- If one pattern is *mislearned* more frequently than it is *accidentally learned*, it will become less attested across many generations of learning.

\(^4\)Bell (1971); Greenberg (1978); Kirby & Huford (2002)
Typology and Stability

Stability Predictions

Typologically rare patterns are more likely to be mislearned than accidentally learned.

- This suggests that rare patterns are likely to be unstable.
- In O’Hara (2018), I look at initial vs. final asymmetries in stop place of articulation.
- I performed a survey of 77 languages with [k p t] in initial position.
- Finnish is the only language I could find with only [t] in final position.
- Must languages that exhibit rare patterns be unstable?
Typologically rare patterns are more likely to be mislearned than accidentally learned.

- This suggests that rare patterns are likely to be unstable.
- In O’Hara (2018), I look at initial vs. final asymmetries in stop place of articulation.
- I performed a survey of 77 languages with [k p t] in initial position.
- Finnish is the only language I could find with only [t] in final position.
- Must languages that exhibit rare patterns be unstable?

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Finals</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>T-Final</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PT-Final</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>All-Finals</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vp</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vk</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>43</td>
</tr>
</tbody>
</table>
Stability Predictions

Typologically rare patterns are more likely to be mislearned than accidentally learned.

- This suggests that rare patterns are likely to be unstable.
- In O’Hara (2018), I look at initial vs. final asymmetries in stop place of articulation.
- I performed a survey of 77 languages with [k p t] in initial position.
- Finnish is the only language I could find with only [t] in final position.
- Must languages that exhibit rare patterns be unstable?

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Finals</td>
<td>tV pV Vk</td>
<td>x x x</td>
</tr>
<tr>
<td>T-Final</td>
<td>tV pV Vk</td>
<td>Vt x x</td>
</tr>
<tr>
<td>PT-Final</td>
<td>tV pV Vk</td>
<td>Vt Vp x</td>
</tr>
<tr>
<td>All-Finals</td>
<td>tV pV Vk</td>
<td>Vt Vp Vk</td>
</tr>
</tbody>
</table>
Stability Predictions

Typologically rare patterns are more likely to be mislearned than accidentally learned.

- This suggests that rare patterns are likely to be unstable.
- In O’Hara (2018), I look at initial vs. final asymmetries in stop place of articulation.
- I performed a survey of 77 languages with [k p t] in initial position.
- Finnish is the only language I could find with only [t] in final position.

Must languages that exhibit rare patterns be unstable?

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Finals</td>
<td>tV pV Vk</td>
<td>X X X</td>
</tr>
<tr>
<td>T-Final</td>
<td>tV pV Vk</td>
<td>Vt X X</td>
</tr>
<tr>
<td>PT-Final</td>
<td>tV pV Vk</td>
<td>Vt Vp X</td>
</tr>
<tr>
<td>All-Finals</td>
<td>tV pV Vk</td>
<td>Vt Vp Vk</td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td></td>
</tr>
</tbody>
</table>
Stability Predictions

Typologically rare patterns are more likely to be mislearned than accidentally learned.

- This suggests that rare patterns are likely to be unstable.
- In O’Hara (2018), I look at initial vs. final asymmetries in stop place of articulation.
- I performed a survey of 77 languages with [k p t] in initial position.
- Finnish is the only language I could find with only [t] in final position.
- Must languages that exhibit rare patterns be unstable?

<table>
<thead>
<tr>
<th></th>
<th>Initial</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Finals</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T-Final</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT-Final</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All-Finals</td>
<td>tV</td>
<td>pV</td>
</tr>
<tr>
<td></td>
<td>Vk</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Finnish Problem

Finnish exhibits a rare pattern, but is it unstable?

- Finnish has stably shown this [t]-final pattern since at least Agricola (1542 (2014)).
- O’Hara (2018) shows that a [t]-final stage is on the pathway of learnability-conditioned final consonant loss.
- How can Finnish be stable, but the vast majority of the time, if a language shows the Finnish pattern, it is unstable?
**The Finnish Problem**

Finnish exhibits a rare pattern, but is it unstable?

- Finnish has stably shown this [t]-final pattern since at least Agricola (1542 (2014)).
- O’Hara (2018) shows that a [t]-final stage is on the pathway of learnability-conditioned final consonant loss.
- How can Finnish be stable, but the vast majority of the time, if a language shows the Finnish pattern, it is unstable?
The Finnish Problem

Finnish exhibits a rare pattern, but is it unstable?

- Finnish has stably shown this [t]-final pattern since at least Agricola (1542 (2014)).
- O’Hara (2018) shows that a [t]-final stage is on the pathway of learnability-conditioned final consonant loss.

- How can Finnish be stable, but the vast majority of the time, if a language shows the Finnish pattern, it is unstable?
The Finnish Problem

Finnish exhibits a rare pattern, but is it unstable?

- Finnish has stably shown this [t]-final pattern since at least Agricola (1542 (2014)).
- O’Hara (2018) shows that a [t]-final stage is on the pathway of learnability-conditioned final consonant loss.
- How can Finnish be stable, but the vast majority of the time, if a language shows the Finnish pattern, it is unstable?
The licit phonotactic forms of a language are just one of the ways in which languages can differ.

- Learning is not just affected by whether or not a form exists in the target data.
- But also how common that form is in the target data.
- Previous work has identified some ways in which the lexicon can interact with learning to shape typology, and affect language change.
  - Staubs (2014); Stanton (2016) show that the low frequency of long words is responsible for underattestation of certain stress patterns.
  - Wedel et al. (2013) show that the functional load of a contrast affects the likelihood of loss of a contrast: i.e. the more minimal pairs the less common merger is. (Though with no minimal pairs, phoneme frequency may increase the chance of merger.)
The licit phonotactic forms of a language are just one of the ways in which languages can differ.

- Learning is not just affected by whether or not a form exists in the target data
  - But also how common that form is in the target data
- Previous work has identified some ways in which the lexicon can interact with learning to shape typology, and affect language change.
  - Staubs (2014); Stanton (2016) show that the low frequency of long words is responsible for underattestation of certain stress patterns
  - Wedel et al. (2013) show that the functional load of a contrast affects the likelihood of loss of a contrast: i.e. the more minimal pairs the less common merger is. (Though with no minimal pairs, phoneme frequency may increase the chance of merger)
Lexical Factors Condition Stability

The licit phonotactic forms of a language are just one of the ways in which languages can differ.

- Learning is not just affected by whether or not a form exists in the target data
  - But also **how** common that form is in the target data
- Previous work has identified some ways in which the lexicon can interact with learning to shape typology, and affect language change.
  - Staubs (2014); Stanton (2016) show that the low frequency of long words is responsible for underattestation of certain stress patterns
  - Wedel et al. (2013) show that the functional load of a contrast affects the likelihood of loss of a contrast: i.e. the more minimal pairs the less common merger is. (Though with no minimal pairs, phoneme frequency may increase the chance of merger)
Lexical Factors Condition Stability

The licit phonotactic forms of a language are just one of the ways in which languages can differ.

- Learning is not just affected by whether or not a form exists in the target data
  - But also **how** common that form is in the target data
- Previous work has identified some ways in which the lexicon can interact with learning to shape typology, and affect language change.
  - Staubs (2014); Stanton (2016) show that the low frequency of long words is responsible for underattestation of certain stress patterns
  - Wedel *et al.* (2013) show that the functional load of a contrast affects the likelihood of loss of a contrast: i.e. the more minimal pairs the less common merger is. (Though with no minimal pairs, phoneme frequency may increase the chance of merger)
The licit phonotactic forms of a language are just one of the ways in which languages can differ.

- Learning is not just affected by whether or not a form exists in the target data
  - But also **how** common that form is in the target data
- Previous work has identified some ways in which the lexicon can interact with learning to shape typology, and affect language change.
  - Staubs (2014); Stanton (2016) show that the low frequency of long words is responsible for underattestation of certain stress patterns
  - Wedel *et al.* (2013) show that the functional load of a contrast affects the likelihood of loss of a contrast: i.e. the more minimal pairs the less common merger is. (Though with no minimal pairs, phoneme frequency may increase the chance of merger)
Lexical Factors Condition Stability

The licit phonotactic forms of a language are just one of the ways in which languages can differ.

- Learning is not just affected by whether or not a form exists in the target data
  - But also **how** common that form is in the target data
- Previous work has identified some ways in which the lexicon can interact with learning to shape typology, and affect language change.
  - Staubs (2014); Stanton (2016) show that the low frequency of long words is responsible for underattestation of certain stress patterns
  - Wedel *et al.* (2013) show that the functional load of a contrast affects the likelihood of loss of a contrast: i.e. the more minimal pairs the less common merger is. (Though with no minimal pairs, phoneme frequency may increase the chance of merger)
Proposal: Lexical Frequencies Condition Stability

CLAIMS

- Finnish is stable due to its lexical frequency
- Language families that have shown different patterns of change have different lexical frequencies.
- The [t]-final pattern is rare because the lexical frequencies that predict the [t]-final pattern are rare.
Generational Learning Model\textsuperscript{5}

- Simulated learners using MaxEnt\textsuperscript{6} grammars
- Learners are initialized with Markedness constraints high, faith low\textsuperscript{7}
- Using the Truncated Perceptron algorithm\textsuperscript{8} train a learning agent off of some limited number of forms\textsuperscript{9} from a teacher

\[ \text{Pattern} \rightarrow \bigcirc \]

\textsuperscript{5}Following Staubs (2014); Hughto (2018)
\textsuperscript{6}Goldwater & Johnson (2003), Hayes (this morning)
\textsuperscript{7}Gnanadesikan (2004); Tesar & Smolensky (2000); Jesney & Tessier (2011)
\textsuperscript{8}Rosenblatt (1958); Magri (2015)
\textsuperscript{9}Kirby & Huford (2002)
Generational Learning Model⁵

- Simulated learners using MaxEnt⁶ grammars
- Learners are initialized with Markedness constraints high, faith low⁷
- Using the Truncated Perceptron algorithm⁸ train a learning agent off of some limited number of forms⁹ from a teacher

Pattern → ○ → ○ → ○ → ?

---

⁵Following Staubs (2014); Hughto (2018)
⁶Goldwater & Johnson (2003), Hayes (this morning)
⁷Gnanadesikan (2004); Tesar & Smolensky (2000); Jesney & Tessier (2011)
⁸Rosenblatt (1958); Magri (2015)
⁹Kirby & Huford (2002)
Update Rule

Error-Driven Perceptron Algorithm \(^{10}\)

- On each iteration, teacher selects an input at random, and produces an output.
- The learner produces an output as well.
- If the learner and teacher differ, raise the weights on the constraints the learner violated, and lower the weights on the constraints the teacher violated.

Example

- Learner:
Generational Model

Update Rule

Error-Driven Perceptron Algorithm ¹⁰

- On each iteration, teacher selects an input at random, and produces an output.
- The learner produces an output as well.
- If the learner and teacher differ, raise the weights on the constraints the learner violated, and lower the weights on the constraints the teacher violated.

Example

- Learner: /tV/-[V]

¹⁰Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)
**Update Rule**

**Error-Driven Perceptron Algorithm**

- On each iteration, teacher selects an input at random, and produces an output.
- The learner produces an output as well.
- If the learner and teacher differ, raise the weights on the constraints the learner violated, and lower the weights on the constraints the teacher violated.

**Example**

- **Teacher**: /\text{tV}-[\text{V}] /\text{pV}-[\text{pV}] /\text{kV}-[\text{kV}] /\text{Vt}-[\text{Vt}] /\text{Vp}-[\text{Vp}] /\text{Vk}-[\text{Vk}]
- **Learner**: /\text{tV}-[\text{V}]

<table>
<thead>
<tr>
<th>T</th>
<th>a</th>
<th>tV</th>
<th>50</th>
<th>50</th>
<th>50</th>
<th>50</th>
<th>1</th>
<th>*K</th>
<th>*KP</th>
<th>*KPT</th>
<th>Onset</th>
<th>NoCoda</th>
<th>Max</th>
<th>HARM</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>(T)</td>
<td>a.</td>
<td>tV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td>-50</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>(L)</td>
<td>b.</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1</td>
<td></td>
<td></td>
<td></td>
<td>-51</td>
<td>.27</td>
<td></td>
</tr>
</tbody>
</table>

---

10 Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)
**Update Rule**

**Error-Driven Perceptron Algorithm**

- On each iteration, teacher selects an input at random, and produces an output.
- The learner produces an output as well.
- If the learner and teacher differ, raise the weights on the constraints the learner violated, and lower the weights on the constraints the teacher violated.

**Example**

- **Teacher:** /tV/-[tV] /pV/-[pV] /kV/-[kV] /Vt/-[Vt] /Vp/-[Vp] /Vk/-[Vk]
- **Learner:** /tV/-[V]

<table>
<thead>
<tr>
<th></th>
<th>50</th>
<th>50↓</th>
<th>50↑</th>
<th>50</th>
<th>1↑</th>
</tr>
</thead>
<tbody>
<tr>
<td>tV</td>
<td>*K</td>
<td>*KP</td>
<td>*KPT</td>
<td>Onset</td>
<td>NoCoda</td>
</tr>
<tr>
<td>(T) a. tV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(L) b. V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10 Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)
Error-Driven Perceptron Algorithm

- On each iteration, teacher selects an input at random, and produces an output.
- The learner produces an output as well.
- If the learner and teacher differ, raise the weights on the constraints the learner violated, and lower the weights on the constraints the teacher violated.

Example


Learner: /tV/-[V]

<table>
<thead>
<tr>
<th>tV</th>
<th>*K</th>
<th>*KP</th>
<th>*KPT</th>
<th>Onset</th>
<th>NoCoda</th>
<th>Max</th>
<th>Harm</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>50</td>
<td>49↓</td>
<td>51↑</td>
<td>50</td>
<td>2↑</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(T) a. tV
(T) b. V

Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)
Learning Bias

The Perceptron is a stochastic algorithm.

- Noise emerges in the learning process both from the selection of input forms, and output forms.
- This noise results in mistransmission across generations, which can compound over many generations.

- Patterns/languages differ in the expected speed of learning
- Faster learned patterns will have less noise than slower learned ones.
The Perceptron is a stochastic algorithm.

- Noise emerges in the learning process both from the selection of input forms, and output forms.
- This noise results in mistransmission across generations, which can compound over many generations.
- Patterns/languages differ in the expected speed of learning
  - Faster learned patterns will have less noise than slower learned ones.
Learning Bias

The Perceptron is a stochastic algorithm.

- Noise emerges in the learning process both from the selection of input forms, and output forms.
- This noise results in mistransmission across generations, which can compound over many generations.
- Patterns/languages differ in the expected speed of learning
- Faster learned patterns will have less noise than slower learned ones.
Consider a uniform frequency across the forms:

<table>
<thead>
<tr>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>.167</td>
<td>.167</td>
<td>.167</td>
<td>.167</td>
<td>.167</td>
<td>.167</td>
</tr>
</tbody>
</table>

The All-Final pattern is learned faster than the [t]-final pattern.
[t]-Final pattern ends up being underattested with these dynamics.

- Change rates are percentage of 50 runs of 40 generations of 4600 iterations at .05 learning rate.

<table>
<thead>
<tr>
<th></th>
<th>kV</th>
<th>Vp</th>
<th>tV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>kV</th>
<th>Vp</th>
<th>tV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vk</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18%
Finnish

Finnish has much more final [t] than the uniform baseline (nearly 25% of syllables with ANY voiceless stops) have final t.

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.107</td>
<td>.096</td>
<td>.142</td>
<td>.115</td>
<td>.0004</td>
<td>.0031</td>
</tr>
<tr>
<td>Normalized</td>
<td>.23</td>
<td>.21</td>
<td>.31</td>
<td>.25</td>
<td>.00</td>
<td>.01</td>
</tr>
</tbody>
</table>

- Finnish frequencies were determined using corpora of 44040 words.\textsuperscript{11}

\textsuperscript{11} Goldsmith & Riggle (2012)
Finnish Simulations

Because /Vt/ is common, the [t]-final pattern is learned much faster than the uniform baseline.

- [t]-final is unlikely to be mislearned, but likely to be accidentally learned.
Finnish Stability

Claim 1
The [t]-final pattern is likely with Finnish frequencies.
Potential New Issues

- It is likely that Finnish stably shows the [t]-final pattern, but how likely was it for Finnish to appear?
- The unmarkedness of coronals makes high frequency of final [t] unsurprising. Why don’t other languages with a lot of [Vt] show Finnish’s [t]-final pattern?
- If [t]-final can be stable, when would a language lose all final stops?
- Three case studies will be used to investigate these issues.
It is likely that Finnish stably shows the [t]-final pattern, but how likely was it for Finnish to appear?

The unmarkedness of coronals makes high frequency of final [t] unsurprising. Why don’t other languages with a lot of [Vt] show Finnish’s [t]-final pattern?

If [t]-final can be stable, when would a language lose all final stops?

Three case studies will be used to investigate these issues.
Potential New Issues

- It is likely that Finnish stably shows the [t]-final pattern, but how likely was it for Finnish to appear?

- The unmarkedness of coronals makes high frequency of final [t] unsurprising. Why don’t other languages with a lot of [Vt] show Finnish’s [t]-final pattern?

- If [t]-final can be stable, when would a language lose all final stops?

- Three case studies will be used to investigate these issues.
It is likely that Finnish stably shows the [t]-final pattern, but how likely was it for Finnish to appear?

The unmarkedness of coronals makes high frequency of final [t] unsurprising. Why don’t other languages with a lot of [Vt] show Finnish’s [t]-final pattern?

If [t]-final can be stable, when would a language lose *all* final stops?

Three case studies will be used to investigate these issues.
Potential New Issues

- It is likely that Finnish stably shows the [t]-final pattern, but how likely was it for Finnish to appear?
- The unmarkedness of coronals makes high frequency of final [t] unsurprising. Why don’t other languages with a lot of [Vt] show Finnish’s [t]-final pattern?
- If [t]-final can be stable, when would a language lose all final stops?
- Three case studies will be used to investigate these issues.
Finno-Ugric: Estonian

- Estonian is closely related to Finnish and still allows final [k].
- Serve as rough estimate of Proto-Finnish.
- In order to better base this on acquisition, we use available child directed speech corpora \(^{12}\), with 15,472 unique words.

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.0899</td>
<td>.11</td>
<td>.174</td>
<td>.0843</td>
<td>.005</td>
<td>.0187</td>
</tr>
<tr>
<td>Normalized</td>
<td>.19</td>
<td>.23</td>
<td>.36</td>
<td>.17</td>
<td>.01</td>
<td>.04</td>
</tr>
</tbody>
</table>

- Estonian has more final [Vk] and less [Vt] than Finnish.

\(^{12}\) Argus (1998); Kohler (2004); Kutt (2018), a.o.
Finno-Ugric: Estonian

- Estonian is closely related to Finnish and still allows final [k].
- Serve as rough estimate of Proto-Finnish.
- In order to better base this on acquisition, we use available child directed speech corpora\(^{12}\), with 15,472 unique words.

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.0899</td>
<td>.11</td>
<td>.174</td>
<td>.0843</td>
<td>.005</td>
<td>.0187</td>
</tr>
<tr>
<td>Normalized</td>
<td>.19</td>
<td>.23</td>
<td>.36</td>
<td>.17</td>
<td>.01</td>
<td>.04</td>
</tr>
</tbody>
</table>

- Estonian has more final [Vk] and less [Vt] than Finnish.

---

\(^{12}\)Argus (1998); Kohler (2004); Kutt (2018), a.o.
Finno-Ugric: Estonian

- Estonian is closely related to Finnish and still allows final [k].
- Serve as rough estimate of Proto-Finnish.
- In order to better base this on acquisition, we use available child directed speech corpora, with 15,472 unique words.

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.0899</td>
<td>.11</td>
<td>.174</td>
<td>.0843</td>
<td>.005</td>
<td>.0187</td>
</tr>
<tr>
<td>Normalized</td>
<td>.19</td>
<td>.23</td>
<td>.36</td>
<td>.17</td>
<td>.01</td>
<td>.04</td>
</tr>
</tbody>
</table>

- Estonian has more final [Vk] and less [Vt] than Finnish.

---

12 Argus (1998); Kohler (2004); Kutt (2018), a.o.
Estonian Simulations

[t]-final is learned faster than baseline, but All-Final is not.

- [t]-final is unlikely to be mislearned, but likely to be accidentally learned.
Estonian Simulations

[t]-final is learned faster than baseline, but All-Final is not.

- [t]-final is unlikely to be mislearned, but likely to be accidentally learned.

![Graph showing the comparison of Estonian [t]-final, Uniform [t]-final, Estonian All-final, and Uniform All-final simulations. The graph compares the sum squared error over iterations. The Estonian [t]-final and Uniform [t]-final show a similar trend, whereas the Estonian All-final and Uniform All-final show a different trend.]
Finno-Ugric Dynamics

These dynamics predict that the [t]-final pattern is likely in the Finno-Ugric family.
West Germanic: English

English, like Finnish has many coronal-final suffixes.

- But no related languages show [t]-final
- Lexical frequencies of English are found using child directed speech (1321 unique words).\(^{13}\)

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.055</td>
<td>.060</td>
<td>.075</td>
<td>.111</td>
<td>.021</td>
<td>.052</td>
</tr>
<tr>
<td>Normalized</td>
<td>.15</td>
<td>.16</td>
<td>.20</td>
<td>.30</td>
<td>.06</td>
<td>.14</td>
</tr>
</tbody>
</table>

\(^{13}\) Bernstein-Ratner (1987); Brent & Cartwright (1996)
West Germanic: English

English, like Finnish has many coronal-final suffixes.

- But no related languages show [t]-final
- Lexical frequencies of English are found using child directed speech (1321 unique words).\(^{13}\)

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.055</td>
<td>.060</td>
<td>.075</td>
<td>.111</td>
<td>.021</td>
<td>.052</td>
</tr>
<tr>
<td>Normalized</td>
<td>.15</td>
<td>.16</td>
<td>.20</td>
<td>.30</td>
<td>.06</td>
<td>.14</td>
</tr>
</tbody>
</table>

Bernstein-Ratner (1987); Brent & Cartwright (1996)
English Simulations

Results of simulations run on English are shown below.

- English learns both simulations faster than the uniform baseline does.
English Simulations

Results of simulations run on English are shown below.

- English learns both simulations faster than the uniform baseline does.
Languages with this sort of profile are more likely to maintain All-Final than Finno-Ugric languages.
Oceanic: Proto-Gela

In the Austronesian family, loss of final consonants has independently occurred at least 14 times.\(^{14}\)

- Gela (Solomon Islands) has lost all final stops.
- No Oceanic languages exhibit the [t]-final pattern.
- Lexical frequencies of Proto-Gela are found using (720) proto-forms from the Comparative Austronesian Dictionary\(^ {15}\).

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.118</td>
<td>.216</td>
<td>.1398</td>
<td>.064</td>
<td>.022</td>
<td>.075</td>
</tr>
<tr>
<td>Normalized</td>
<td>.19</td>
<td>.34</td>
<td>.22</td>
<td>.10</td>
<td>.03</td>
<td>.12</td>
</tr>
</tbody>
</table>

---

\(^{14}\) Blevins (2004)

\(^{15}\) Blust & Trussel (2010 (2018))
Oceanic: Proto-Gela

In the Austronesian family, loss of final consonants has independently occurred at least 14 times.¹⁴

- Gela (Solomon Islands) has lost all final stops.
- No Oceanic languages exhibit the [t]-final pattern.
- Lexical frequencies of Proto-Gela are found using (720) proto-forms from the Comparative Austronesian Dictionary¹⁵.

<table>
<thead>
<tr>
<th>Forms</th>
<th>tV</th>
<th>pV</th>
<th>kV</th>
<th>Vt</th>
<th>Vp</th>
<th>Vk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>.118</td>
<td>.216</td>
<td>.1398</td>
<td>.064</td>
<td>.022</td>
<td>.075</td>
</tr>
<tr>
<td>Normalized</td>
<td>.19</td>
<td>.34</td>
<td>.22</td>
<td>.10</td>
<td>.03</td>
<td>.12</td>
</tr>
</tbody>
</table>

---

¹⁴ Blevins (2004)
¹⁵ Blust & Trussel (2010 (2018))
Gela Simulations

- Gela performs worse than baseline on both patterns than uniform baseline
Gela Simulations

- Gela performs worse than baseline on both patterns than uniform baseline
Oceanic Dynamics

It is predicted that Oceanic languages should show All-Final and No-Final patterns, but not [t]-Final.
We’ve looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each:

- Above the blue line, languages maintain the All-Final pattern.
- In the bottom left green sector, languages are unstable in All-Final and [t]-Final, so may lose coda stops.
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns.
Interim Summary

We’ve looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each

- Above the blue line, languages maintain the All-Final pattern
- In the bottom left green sector, languages are unstable in All-Final and [t]-Final, so may lose coda stops
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns
Interim Summary

We’ve looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each

- Above the blue line, languages maintain the All-Final pattern
- In the bottom left green sector, languages are unstable in All-Final and [t]-Final, so may lose coda stops
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns
We’ve looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each:

- Above the blue line, languages maintain the All-Final pattern.
- In the bottom left green sector, languages are unstable in All-Final and [t]-Final, so may lose coda stops.
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns.
Interim Summary

We’ve looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each:

- Above the blue line, languages maintain the All-Final pattern.
- In the bottom left green sector, languages are unstable in All-Final and [t]-Final, so may lose coda stops.
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns.
Interim Summary

We’ve looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each.

- Above the blue line, languages maintain the All-Final pattern.
- In the bottom left green sector, languages are unstable in All-Final and [t]-Final, so may lose coda stops.
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns.

CLAIM 2

Language families that have shown different patterns of change have different lexical frequencies.
But why is \([t]\)-Final rare?

The \([t]\)-final pattern is a likely result for languages with frequencies similar to the Finno-Ugric languages.

Why do we not see it in other language families?

- The \([t]\)-final pattern is restricted to one small region of the lexical frequency space
- How big is this sector?
How big is [t]-final sector?

To see how many of the possible frequency profiles predict that [t]-final should be likely and stable, I ran simulations across many frequencies.

- For each of the 6 forms, I iterated with a step size of .1 probability, ranging from 0 to 1; while ensuring that the sum of all 6 forms was 1.
  - This resulted in 2002 frequency profiles
  - 5 runs of 2 generations with 360 iterations with a learning rate of .5.
Results

- The [t]-Final stable region is smaller than the other regions.
- This causes [t]-Final to be cross-linguistically rare, even when it can be stable.

CLAIM 3

The [t]-final pattern is rare because the lexical frequencies that predict the [t]-final pattern are rare.
Conclusion

Lexical Frequency greatly conditions the learnability of different patterns.

- Frequency is an important factor to consider when making typological generalizations based on learning.
- Some lexical frequencies can show stability patterns quite at odds with the rest of the frequency space.

Future Questions

- Languages are not likely uniformly distributed across the lexical frequency space, so volume as measured here may not be the best metric.
- Lexical Frequency changes as languages evolve. A model integrating both phonotactic and lexicon learning may make further different predictions about how languages are distributed across frequency space.
Works Cited I


Works Cited II


Works Cited III


