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)

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)

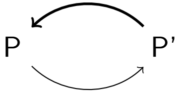
[2] Legendre *et al.* (1990); Pater (2016)

[3] Boersma (2003); Pater & Moreton (2012); Staubs (2014); Hughto (2018); O'Hara (2018)

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.

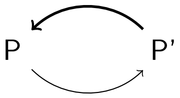


[4] Bell (1971); Greenberg (1978); Kirby & Huford (2002)

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] Bell (1971); Greenberg (1978); Kirby & Huford (2002)

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.

$$\begin{tikzcd} \mathbf{P} \arrow[r, bend right] & \mathrm{P'} \arrow[l, bend right] \end{tikzcd}$$

[4] Bell (1971); Greenberg (1978); Kirby & Huford (2002)

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?

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?

	Initial			Final			
No Finals	tV	pV	Vk	✗	✗	✗	27
T-Final	tV	pV	Vk	**Vt**	✗	✗	1
PT-Final	tV	pV	Vk	**Vt**	**Vp**	✗	3
All-Finals	tV	pV	Vk	**Vt**	**Vp**	**Vk**	43

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?

	Initial			Final			
No Finals	tV	pV	Vk	✗	✗	✗	27
T-Final	tV	pV	Vk	**Vt**	✗	✗	1
PT-Final	tV	pV	Vk	**Vt**	**Vp**	✗	3
All-Finals	tV	pV	Vk	**Vt**	**Vp**	**Vk**	43

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?

	Initial			Final				
No Finals	tV	pV	Vk	✗	✗	✗	27	
T-Final	tV	pV	Vk	Vt	✗	✗	1	
PT-Final	tV	pV	Vk	Vt	Vp	✗	3	
All-Finals	tV	pV	Vk	Vt	Vp	Vk	43	

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?

	Initial			Final			
No Finals	tV	pV	Vk	✗	✗	✗	27
T-Final	tV	pV	Vk	**Vt**	✗	✗	1
PT-Final	tV	pV	Vk	**Vt**	**Vp**	✗	3
All-Finals	tV	pV	Vk	**Vt**	**Vp**	**Vk**	43

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

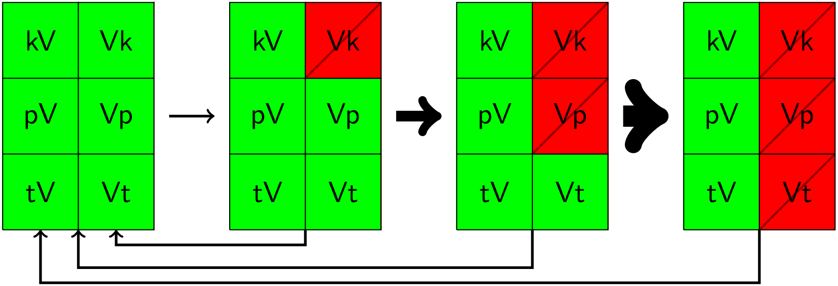
Finnish exhibits a rare pattern, but is it unstable?

- 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?

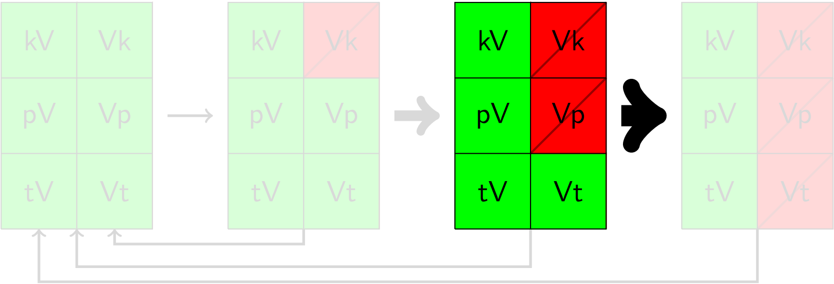
- Finnish has stably shown this [t]-final pattern since at least Agricola (1542 (2014)).
- O'Hara (2018) shows that a [t]-final stage is on the pathway of learnability-conditioned final consonant loss.
- How can Finnish be stable, but the vast majority of the time, if a language shows the Finnish pattern,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



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)

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)

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)

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 Model

Generational Learning Model[5]

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

Pattern $\longrightarrow$ ◯

[5] Following Staubs (2014); Hughto (2018)

[6] Goldwater & Johnson (2003), Hayes (this morning)

[7] Gnanadesikan (2004); Tesar & Smolensky (2000); Jesney & Tessier (2011)

[8] Rosenblatt (1958); Magri (2015)

[9] Kirby & Huford (2002)

Generational Model

Generational Learning Model[5]

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

Pattern ⟶ ○ ⟶ ○ ⟶ ○ ⟶ **?**

[5] Following Staubs (2014); Hughto (2018)

[6] Goldwater & Johnson (2003), Hayes (this morning)

[7] Gnanadesikan (2004); Tesar & Smolensky (2000); Jesney & Tessier (2011)

[8] Rosenblatt (1958); Magri (2015)

[9] Kirby & 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

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

[10] Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)

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

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

[10] Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)

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

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

	50	50	50	50	50	1		
tV	*K	*KP	*KPT	ONSET	NOCODA	MAX	HARM	PROB
(T) a. tV			-1				-50	.73
(L) b. V				-1		-1	-51	.27

[10] Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)

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

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

	50	50	50↓	50↑	50	1↑		
tV	*K	*KP	*KPT	ONSET	NOCODA	MAX	HARM	PROB
(T) a. tV			-1↓				-50↑	.73
(L) b. V				-1↑		-1↑	-51↓	.27

[10] Rosenblatt (1958); Boersma & Pater (2016); Magri (2015)

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

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

	50	50	49↓	51↑	50	2↑		
tV	*K	*KP	*KPT	ONSET	NOCODA	MAX	HARM	PROB
(T) a. tV			-1↓				-49↑	.98
(L) b. V				-1↑		-1↑	-53↓	.02

[10] 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.

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.

Learning Bias

The Perceptron is a stochastic algorithm.

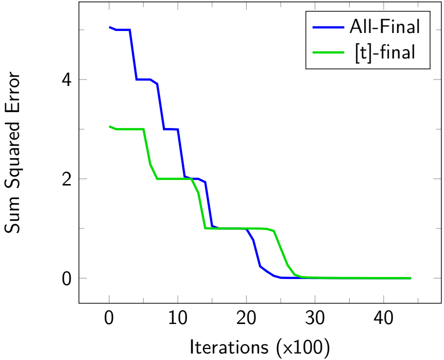
- Noise emerges in the learning process both from the selection of input forms, and output forms.
- This noise results in mistransmission across generations, which can compound over many generations.
- Patterns/languages differ in the expected speed of learning
- Faster learned patterns will have less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

Uniform Baseline

Consider a uniform frequency across the forms

tV	pV	kV	Vt	Vp	Vk
.167	.167	.167	.167	.167	.167

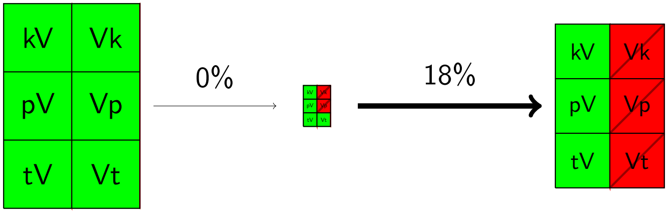
- The All-Final pattern is learned faster than the [t]-final pattern



Uniform Dynamics

[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



Finnish

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

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.107	.096	.142	.115	.0004	.0031
Normalized	.23	.21	.31	.25	.00	.01

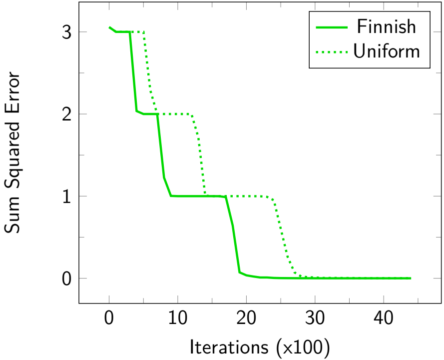
- Finnish frequencies were determined using corpora of 44040 words.[11]

[11] Goldsmith & Riggle (2012)

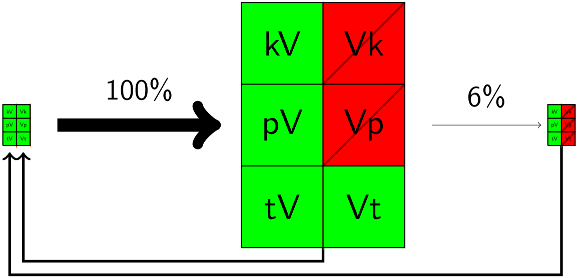
Finnish Simulations

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.

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.

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.

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.

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.

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.0899	.11	.174	.0843	.005	.0187
Normalized	.19	.23	.36	.17	.01	.04

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

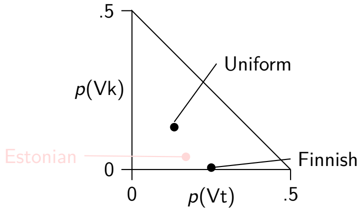
[12] Argus (1998); Kohler (2004); Kutt (2018), a.o.

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.

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.0899	.11	.174	.0843	.005	.0187
Normalized	.19	.23	.36	.17	.01	.04

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



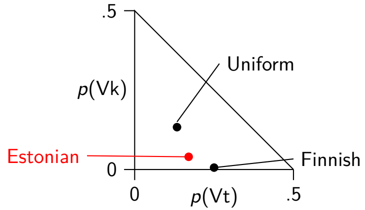
[12] Argus (1998); Kohler (2004); Kutt (2018), a.o.

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.

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.0899	.11	.174	.0843	.005	.0187
Normalized	.19	.23	.36	.17	.01	.04

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

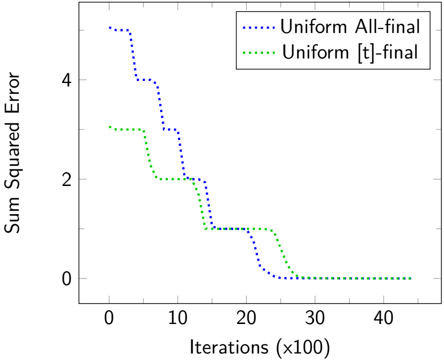


[12] Argus (1998); Kohler (2004); Kutt (2018), a.o.

Estonian Simulations

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

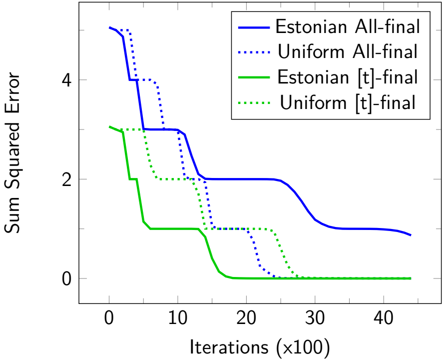
- [t]-final is unlikely to be mislearned, but likely to be accidentally learned.



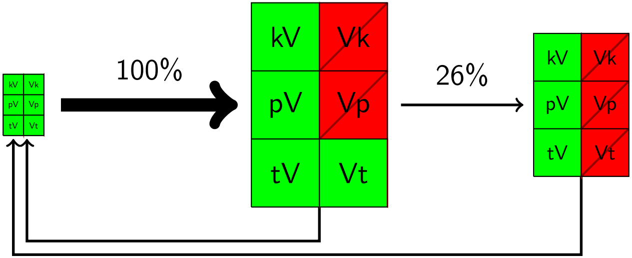
Estonian Simulations

[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.



Finno-Ugric Dynamics



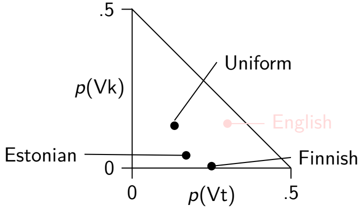
These dynamics predict that the [t]-final pattern is likely in the Finno-Ugric family.

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]

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.055	.060	.075	.111	.021	.052
Normalized	.15	.16	.20	.30	.06	.14



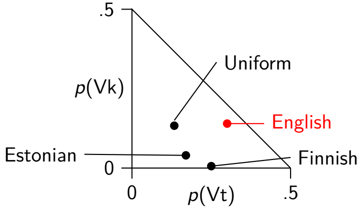
[13] Bernstein-Ratner (1987); Brent & Cartwright (1996)

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]

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.055	.060	.075	.111	.021	.052
Normalized	.15	.16	.20	.30	.06	.14

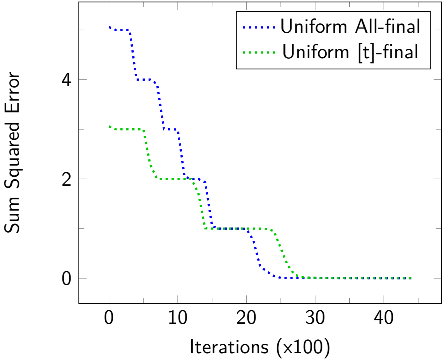


[13] Bernstein-Ratner (1987); Brent & Cartwright (1996)

English Simulations

Results of simulations run on English are shown below.

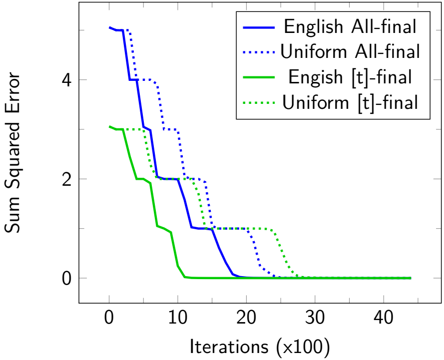
- English learns both simulations faster than the uniform baseline does.



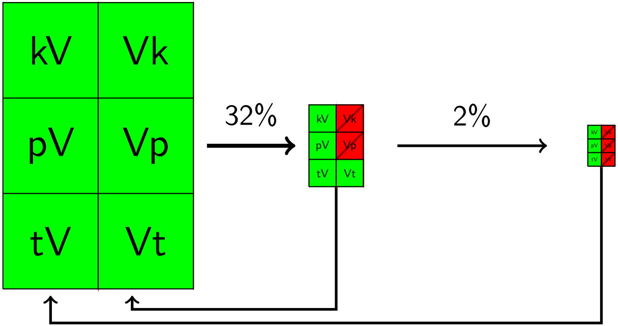
English Simulations

Results of simulations run on English are shown below.

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



English Dynamics



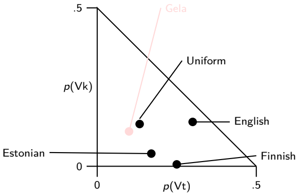
Languages with this sort of profile are more likely to maintain All-Final than Finno-Ugric languages.

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].

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.118	.216	.1398	.064	.022	.075
Normalized	.19	.34	.22	.10	.03	.12



[14] Blevins (2004)

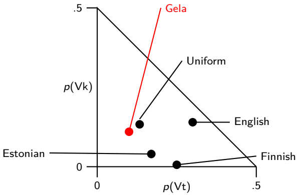
[15] Blust & Trussel (2010 (2018))

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 Dictionary[15].

Forms	tV	pV	kV	Vt	Vp	Vk
Frequency	.118	.216	.1398	.064	.022	.075
Normalized	.19	.34	.22	.10	.03	.12

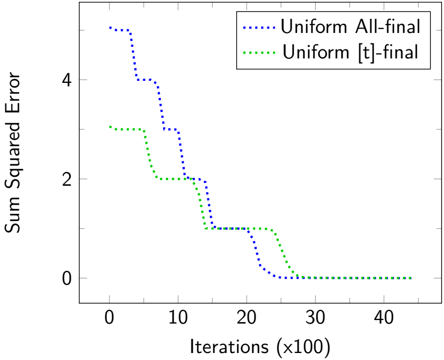


14 Blevins (2004)

15 Blust & Trussel (2010 (2018))

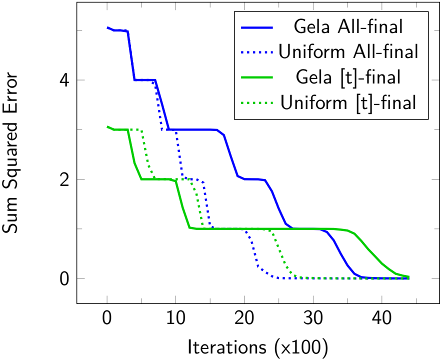
Gela Simulations

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

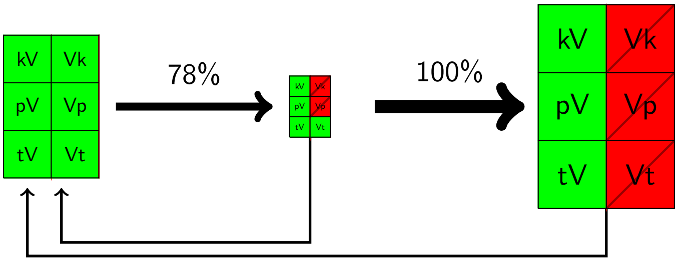


Gela Simulations

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



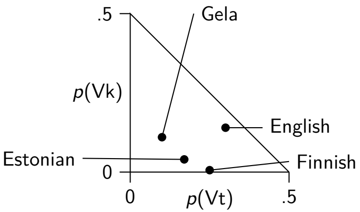
Oceanic Dynamics



It is predicted that Oceanic languages should show All-Final and No-Final patterns, but not [t]-Final.

Interim Summary

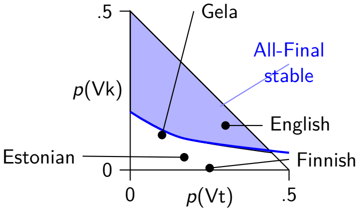
We've looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each



- 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

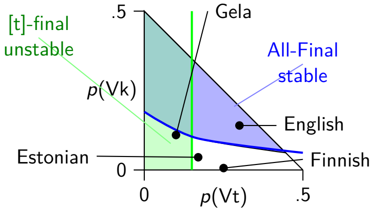
We've looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each



- 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

Interim Summary

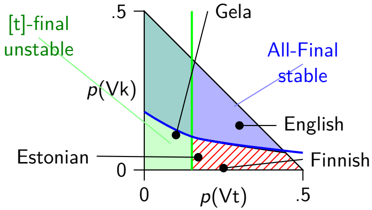
We've looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each



- 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

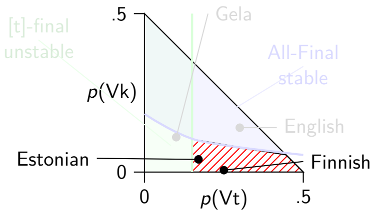
We've looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each



- 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

Interim Summary

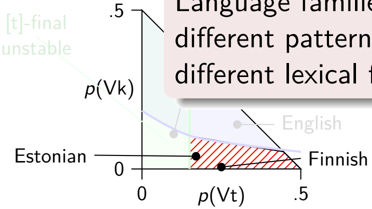
We've looked at three language families, and seen that the difference in frequencies predicts a different pattern of stability in each



- 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



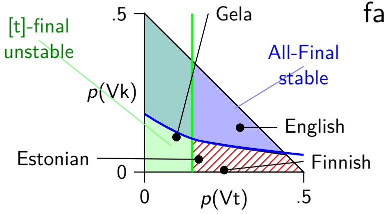
- Above the blue line, languages … l pattern
- … een sector, … e in All-Final … lose coda …
- In the red region, All-Final is sufficiently unstable, and [t]-Final is sufficiently stable to predict [t]-Final patterns

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.



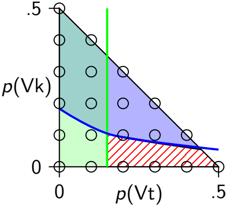
Why do we not see it in other language families?

- 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?

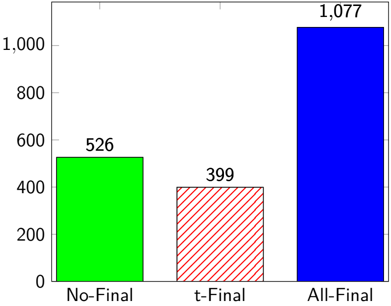
To see how many of the possible frequency profiles predict that [t]-final should be likely and stable, I ran simulations across many frequencies

- 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

Agricola, Mikael. 1542 (2014). *ABC-kirja ja Rukouskirjan alkuosa raamatullisten rukousten loppuun s. 1 - 344 riville 7. [tekstikorpus]*.

Argus, Reili. 1998. *CHILDES'i eesti andmepank ja selle suhtluskeskne analüüs (Hendrik, 1.6-2.6)*. Tallinn: Tallinna Pedagoogikaülikool.

Bell, A. 1971. Some patterns of the occurrence and formation of syllabic structure. *Pages 23–138 of: Working Papers on Language Universals*, vol. 6.

Bernstein-Ratner, N. 1987. The phonology of parent-child speech. *In:* Nelson, K., & van Kleeck, A. (eds), *Children's Language*, vol. 6. Hillsdale, NJ: Erlbaum.

Blevins, Juliette. 2004. The Mystery of Austronesian Final Consonant Deletion. *Oceanic Linguistics*, **43**(1), 208–213.

Blust, Robert, & Trussel, Stephen. 2010 (2018). *Austronesian Comparative Dictionary, web edition*.

Boersma, Paul. 2003. Review of Bruce Tesar and Paul Smolensky 2000, Learnability in Optimality Theory. *Phonology*, **20**, 436–446.

Boersma, Paul, & Pater, Joe. 2016. Convergence properties of a gradual learning algorithm for Harmonic Grammar. *In:* McCarthy, John J., & Pater, Joe (eds), *Harmonic Grammar and Harmonic Serialism*. Equinox.

Brent, M.R., & Cartwright, T.A. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, **61**, 93–125.

Gnanadesikan, Amalia. 2004. Markedness and Faithfulness in child phonology [ROA-67]. *Pages 73–108 of:* Kager, René, Pater, Joe, & Zonneveld, Wim (eds), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge: Cambridge University Press.

Goldsmith, John, & Riggle, Jason. 2012. Information theoretic approaches to phonological structure: The case of Finnish vowel harmony. *Natural Language & Linguistic Theory*, **30**, 859–896.

Works Cited II

Goldwater, Sharon, & Johnson, Mark. 2003. Learning OT constraint rankings using a Maximum Entropy model. *In: Proceedings of the Workshop on Variation within Optimality Theory*. Stockholm University.

Greenberg, Joseph H. 1978. Diachrony, synchrony, and language universals. *Pages 61–91 of:* Greenberg, Joseph H., Ferguson, C.A., & Moravcsik, E.A. (eds), *Universals of human language, volume 1 method and theory*. Stanford, CA: Stanford University Press.

Hughto, Coral. 2018. Investigating the Consequences of Iterated Learning in Phonological Typology. *In: Proceedings of the Society for Computation in Linguistics*, vol. 1.

Jesney, Karen, & Tessier, Anne-Michelle. 2011. Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language & Linguistic Theory*, **29**.

Kirby, Simon, & Huford, James. 2002. The emergence of linguistic structure: An overview of the iterated learning model. *Chap. 6, pages 121–148 of:* Cangelosi, A, & Parisi, D. (eds), *Simulating the Evolution of Language*. London: Springer Verlag.

Kohler, K. 2004. *Erwerb der fruhen Verbmorphologie im Estnischen*. Ph.D. thesis, University of Potsdam.

Kutt, Andra. 2018. Testi "The Multilingual Assessment Instrument for Narratives" kasutamine eesti laste jutustamisoskuse hindamiseks. *Eesti Rakenduslingvistika Ühingu aastaraamat*, **14**.

Legendre, Géraldine, Miyata, Yoshiro, & Smolensky, Paul. 1990. Harmonic Grammar - a formal multi-level conectionist theory of linguistic wellformedness: an application. *Pages 884–891 of:* Erlbaum, Lawrence (ed), *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*.

Magri, Giorgiog. 2015. How to keep the HG weights non-negative: the truncated Perceptron reweighting rule. *Journal of Language Modeling*, **3**(2), 345–375.

McCarthy, John J., & Prince, Alan. 1995. Faithfulness and reduplicative identity. *University of Massachusetts Occasional Papers*, **18**, 249–384.

O'Hara, Charlie. 2018. *Learnability Captures Soft Typology of Coda Stop Inventories,*. Presented at LSA 2018.

Works Cited III

Pater, Joe. 2016. Universal Grammar with Weighted Constraints. *Pages 1–46 of:* McCarthy, John J., & Pater, Joe (eds), *Harmonic Grammar and Harmonic Serialism*. London: Equinox.

Pater, Joe, & Moreton, Elliott. 2012. Structurally biased phonology: complexity in language learning and typology. *The EFL Journal*, **3**(2), 1–44.

Prince, Alan, & Smolensky, Paul. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Rosenblatt, F. 1958. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, **65**, 386–408.

Stanton, Juliet. 2016. Learnability shapes typology: the case of the midpoint pathology. *Language*, **92**(4), 753–791.

Staubs, Robert. 2014. *Computational modeling of learning biases in stress typology*. Ph.D. thesis, University of Massachusetts Amherst, Amherst.

Tesar, Bruce, & Smolensky, Paul. 2000. *Learnability in Optimality Theory*. MIT Press1.

Wedel, Andrew, Kaplan, Abby, & Jakckson, Scott. 2013. High functional load inhibits phonological contrst loss: A corpus study. *Cognition*, **128**(2), 179–186.