# Learnability captures soft typology of coda stop inventories

Charlie O'Hara
University of Southern California

Salt Lake City, LSA 2018
January 6, 2018

# Introduction

Constraint based grammars (Optimality Theory[2], Harmonic Grammar[3], etc.) are intended to make strong predictions about typology.

- **Factorial Typology**: All possible languages are predicted by some ranking/*weighting* of a universal set of constraints.
  - Restricts the search space of possible languages
- Only directly captures **categorical** generalizations

Not all typological generalizations are categorical.

[2] Prince & Smolensky (1993/2004); McCarthy & Prince (1995)

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

# Introduction

Constraint based grammars (Optimality Theory[2], Harmonic Grammar[3], etc.) are intended to make strong predictions about typology.

- **Factorial Typology**: All possible languages are predicted by some ranking/*weighting* of a universal set of constraints.
  - Restricts the search space of possible languages
- Only directly captures **categorical** generalizations

Not all typological generalizations are categorical.

[2] Prince & Smolensky (1993/2004); McCarthy & Prince (1995)

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

# Categorical vs. Soft generalizations

- **Categorical generalization**: Some logically possible pattern is never attested cross-linguistically
  - Ex: If a language has complex onsets, it allows simple onsets.
- **Soft generalization**: A pattern is relatively more frequently attested cross-linguistically than another.
  - Ex: If a language has [b] it **usually** has [g] as well. [4]

[4] (Pater, 2012)

# Learning Bias

Asymmetries in learnability can cause soft typological generalizations[5] [6]

- Hard to learn patterns are more likely to be mislearned; and therefore change across many generations.
- Learning Algorithm + Grammar + ??? → Learning Bias
- Learning simulations can uncover biases hard to identify by analysis alone.

[5] Greenberg (1978); Staubs (2014); Pater & Moreton (2012)

[6] Though c.f. R-Volume (Riggle, 2014; Anttila, 1997; Coetzee, 2002) for another way to capture soft typology

# Learning Bias

Asymmetries in learnability can cause soft typological generalizations[5] [6]

- Hard to learn patterns are more likely to be mislearned; and therefore change across many generations.

What makes a pattern hard to learn?

- Learning Algorithm + Grammar + ??? → Learning Bias
- Learning simulations can uncover biases hard to identify by analysis alone.

[5] Greenberg (1978); Staubs (2014); Pater & Moreton (2012)

[6] Though c.f. R-Volume (Riggle, 2014; Anttila, 1997; Coetzee, 2002) for another way to capture soft typology

# Learning Bias

Asymmetries in learnability can cause soft typological generalizations[5] [6]

- Hard to learn patterns are more likely to be mislearned; and therefore change across many generations.

What makes a pattern hard to learn?

- Learning Algorithm + Grammar + ??? → Learning Bias
- Learning simulations can uncover biases hard to identify by analysis alone.

[5] Greenberg (1978); Staubs (2014); Pater & Moreton (2012)

[6] Though c.f. R-Volume (Riggle, 2014; Anttila, 1997; Coetzee, 2002) for another way to capture soft typology

# Learning Bias

Asymmetries in learnability can cause soft typological generalizations[5] [6]

- Hard to learn patterns are more likely to be mislearned; and therefore change across many generations.

What makes a pattern hard to learn?

- Learning Algorithm + Grammar + ??? → Learning Bias
- Learning simulations can uncover biases hard to identify by analysis alone.

[5] Greenberg (1978); Staubs (2014); Pater & Moreton (2012)

[6] Though c.f. R-Volume (Riggle, 2014; Anttila, 1997; Coetzee, 2002) for another way to capture soft typology

# Simplicity Bias

The Perceptron algorithm[14] with a MaxEnt grammar predicts featurally simple patterns are more easily learned—and better attested[15].

- Simplicity for our purposes can be defined by number of features necessary to define a pattern.
    - Forms in a language *usually* follow a uniform pattern rather than be exceptions.
    - Phonological inventories are usually feature economic. [16]

[14] Rosenblatt (1958); Boersma & Pater (2016)

[15] Pater & Moreton (2012); Pater (2012); Culbertson *et al.* (2013)

[16] (de Groot, 1931; Hockett, 1955; Martinet, 1968; Clements, 2003)

# Simplicity Bias

The Perceptron algorithm[14] with a MaxEnt grammar predicts featurally simple patterns are more easily learned—and better attested[15].

- Simplicity for our purposes can be defined by number of features necessary to define a pattern.
  - Forms in a language *usually* follow a uniform pattern rather than be exceptions.
  - Phonological inventories are usually feature economic. [16]
- The pattern that bans all **coda** stops is simpler than the pattern that bans just dorsal **coda** stops.

| **Coda** | | | **Coda+dorsal** | | |
|---|---|---|---|---|---|
| kV | pV | tV | kV | pV | tV |
| ~~**Vk**~~ | ~~**Vp**~~ | ~~**Vt**~~ | ~~**Vk**~~ | **Vp** | **Vt** |

[14] Rosenblatt (1958); Boersma & Pater (2016)

[15] Pater & Moreton (2012); Pater (2012); Culbertson *et al.* (2013)

[16] (de Groot, 1931; Hockett, 1955; Martinet, 1968; Clements, 2003)

# Interacting Dimensions

The majority of previous work showing simplicity bias are focused on a simple unidimensional systems.

- [**Vk**] is linked to [**Vp**].
- But how is presence of [**Vk**] linked to [kV]?

# Interacting Dimensions

The majority of previous work showing simplicity bias are focused on a simple unidimensional systems.

- [**Vk**] is linked to [**Vp**].
- But how is presence of [**Vk**] linked to [kV]?

| | **Coda** | | | | dorsal | |
|---|---|---|---|---|---|---|
| kV | pV | tV | | ~~kV~~ | pV | tV |
| ~~**Vk**~~ | ~~**Vp**~~ | ~~**Vt**~~ | | ~~**Vk**~~ | **Vp** | **Vt** |

- Both of these patterns are featurally simple

# Interacting Dimensions

The majority of previous work showing simplicity bias are focused on a simple unidimensional systems.

- [**Vk**] is linked to [**Vp**].
- But how is presence of [**Vk**] linked to [kV]?

| | Coda | |
|---|---|---|
| kV | pV | tV |
| ~~**Vk**~~ | ~~**Vp**~~ | ~~**Vt**~~ |

| | dorsal | |
|---|---|---|
| ~~kV~~ | pV | tV |
| ~~**Vk**~~ | **Vp** | **Vt** |

- Both of these patterns are featurally simple

**But only the first pattern is common!**

# Claim

**Simplicity is not sufficient**—Not all soft generalizations in the observed typology are simplicity based.

**Learning Bias can do more**—Simplicity is not the only factor that can make a pattern easy to learn

- *Common* and *Easy to Learn* patterns are **Simple** and **Markedness Consistent**

# Markedness Hierarchies

Two markedness hierarchies—defined by well-studied typological implications thought to be categorical (and therefore encoded in the grammar.)

- Onset vs. coda [17]
  CV $\succ$ VC
- Place of articulation[18]
  Coronal $\succ$ Labial $\succ$ Dorsal

[17](Jakobson & Halle, 1956; Kingston, 1985; Goldsmith, 1990)

[18](de Lacy, 2006; Kean, 1975; Lombardi, 2001)

# Typological Survey

Development of the Word Edge Consonant Database (WECD)

- 173 Languages
- Languages with no consonants (of any sort) in word-final position were not included.
- Focus on 73 languages with just [k p t] initially.[19]

[19] that allow maximally three supralaryngeal places of articulation for stops

# Results of Typological Survey

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

| | Onset | | | Coda | | | |
|---|---|---|---|---|---|---|---|
| No Codas | tV | pV | Vk | ✗ | ✗ | ✗ | 31 |
| T-Coda | tV | pV | Vk | **Vt** | ✗ | ✗ | 1 |
| PT-Coda | tV | pV | Vk | **Vt** | **Vp** | ✗ | 6 |
| All-Codas | tV | pV | Vk | **Vt** | **Vp** | **Vk** | 35 |

# No Coda Pattern

| | Onset | | | Coda | | | |
|---|---|---|---|---|---|---|---|
| No Codas | tV | pV | kV | ✗ | ✗ | ✗ | 31 |
| T-Coda | tV | pV | kV | **Vt** | ✗ | ✗ | 1 |
| PT-Coda | tV | pV | kV | **Vt** | **Vp** | ✗ | 6 |
| All-Codas | tV | pV | kV | **Vt** | **Vp** | **Vk** | 35 |

- Example: Italian

[ˈtasto] *button* [ˈpasto] *meal* [ˈkasto] *chaste*
*[kasat] *[kasap] *[kasak]

# Only T-Coda Pattern

| | Onset | | | Coda | | | | |
|---|---|---|---|---|---|---|---|---|
| No Codas | tV | pV | kV | ✗ | ✗ | ✗ | 31 | |
| T-Coda | tV | pV | kV | **Vt** | ✗ | ✗ | 1 | |
| PT-Coda | tV | pV | kV | **Vt** | **Vp** | ✗ | 6 | |
| All-Codas | tV | pV | kV | **Vt** | **Vp** | **Vk** | 35 | |

- Example: Finnish

[telata] *to paint* [pelata] *to play* [kelata] *to wind*
[keot] *anthills* *[keop] *[keok]

# PT-Coda Pattern

| | Onset | | | Coda | | | | |
|---|---|---|---|---|---|---|---|---|
| No Codas | tV | pV | kV | ✗ | ✗ | ✗ | 31 | |
| T-Coda | tV | pV | kV | **Vt** | ✗ | ✗ | 1 | |
| PT-Coda | tV | pV | kV | **Vt** | **Vp** | ✗ | 6 | |
| All-Codas | tV | pV | kV | **Vt** | **Vp** | **Vk** | 35 | |

- Example: Movima (Haude, 2006)

[ˈtanna] *I cut* [pɛnna] *my landing place* [**k**anan] *your food*

[tʃuːˈhat] *palm tree* [kuːˈɗup] *flea* *[kuːˈɗuk]

# All-Codas Pattern

| | Onset | | | Coda | | | |
|---|---|---|---|---|---|---|---|
| No Codas | tV | pV | kV | ✗ | ✗ | ✗ | 31 |
| T-Coda | tV | pV | kV | **Vt** | ✗ | ✗ | 1 |
| PT-Coda | tV | pV | kV | **Vt** | **Vp** | ✗ | 6 |
| All-Codas | tV | pV | kV | **Vt** | **Vp** | **Vk** | 35 |

- Example: English

[tap] *top* [pap] *pop* [kap] *cop*
[pat] *pot* [pap] *pop* [pak] *pock*

## Soft Generalizations

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

| | Onset | | | Coda | | | | |
|---|---|---|---|---|---|---|---|---|
| No Codas | tV | pV | kV | ✗ | ✗ | ✗ | 31 | |
| T-Coda | tV | pV | kV | **Vt** | ✗ | ✗ | 1 | |
| PT-Coda | tV | pV | kV | **Vt** | **Vp** | ✗ | 6 | |
| All-Codas | tV | pV | kV | **Vt** | **Vp** | **Vk** | 35 | |

- [**Vt**] ***usually*** implies [**Vp**] ($\frac{41}{42}$)
- [**Vp**] *typically* implies [**Vk**] ($\frac{35}{41}$)

Simplicity bias predicts these generalizations.

| **Coda** | | |
|---|---|---|
| kV | pV | tV |
| ~~**Vk**~~ | ~~**Vp**~~ | ~~**Vt**~~ |

| **Coda**+Dorsal | | |
|---|---|---|
| kV | pV | tV |
| ~~**Vk**~~ | **Vp** | **Vt** |

| **Coda**+Dor,Lab | | |
|---|---|---|
| kV | pV | tV |
| ~~**Vk**~~ | ~~**Vp**~~ | **Vt** |

# Simplicity on the other dimension

Simplicity also predicts that the pattern with no dorsals will be well attested.

| | Dorsal | |
|---|---|---|
| ~~kV~~ | pV | tV |
| **~~Vk~~** | **Vp** | **Vt** |

- No languages that lack dorsal stops allow coda stops

# Simplicity on the other dimension

Simplicity also predicts that the pattern with no dorsals will be well attested.

| Dorsal | | |
|---|---|---|
| ~~kV~~ | pV | tV |
| ~~**Vk**~~ | **Vp** | **Vt** |

- No languages that lack dorsal stops allow coda stops

| Language | Family | Initial | Final | |
|---|---|---|---|---|
| Xavante | Macro-Ge | p t ʔ | ∅ | (Estevam, 2011) |
| Tahitian | Austronesian | p t ʔ | ∅ | (Tryon, 1970) |
| Wutung | Skou | p t ʔ | ∅ | (Marmion, 2010) |
| Vanimo | Skou | p t ʔ | ∅ | (Clifton, 1995) |
| Nouri | Skou | p t | ∅ | (Donohue, 2010) |

Typological Generalization: [**Vt**] implies [kV]

# Results of Typological Survey

- Soft Generalizations:
  - [**Vt**] ***usually*** implies [**Vp**]
  - [**Vp**] *typically* implies [**Vk**]
- Categorical Generalization:
  - [**Vt**] implies [kV]

# Stability Model

To model how learnability would shape soft typology, the **stability** of each pattern across many generations is tested.[20]

- Patterns are rarely *perfectly* transmitted from teacher to learner
- Some patterns are more resilient to small errors—whereas others fall apart quickly across generations.

Generational stability is modeled by iterating a learning model.[21]

- First generation is trained from categorical data from the pattern in question.
- Data is cut off after a limited number (3200) of forms.
- The learner then produces forms from the grammar they have learned to train the next learner.
- And so on for 20 generations.

[20] c.f. Hughto (2018): stability $\neq$ interactive model i.e. Pater (2012); Hughto & Pater (2017)

[21] (Staubs, 2014; Dowman *et al.* , 2006)

# Learning Model

Each learner uses a MaxEnt grammar.[22]

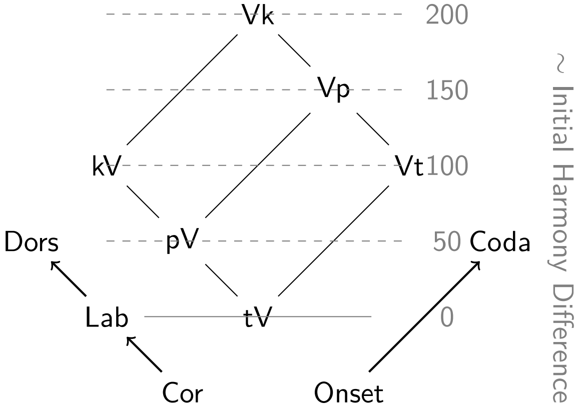
- Constraints Used:
  - Place of Articulation scale: *k, *kp, *kpt
  - Onset vs. coda: ONSET, NOCODA
  - Faithfulness: MAX
- Initial State[23]
  - Markedness Constraints at 50
  - Faithfulness (MAX) at 1

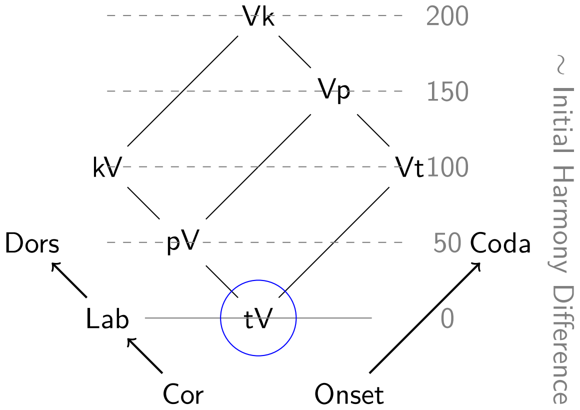
[22] (Goldwater & Johnson, 2003)

[23] (As in Jesney & Tessier 2011)

This graph represents the initial state of a learner in these learning simulations.

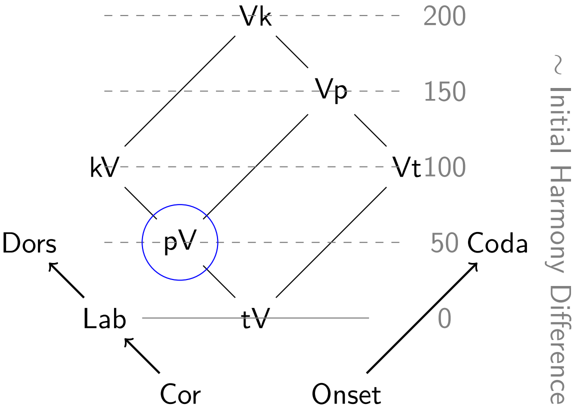
- Positive values on the y-axis represent forms where the consonant deletes more than it surfaces.





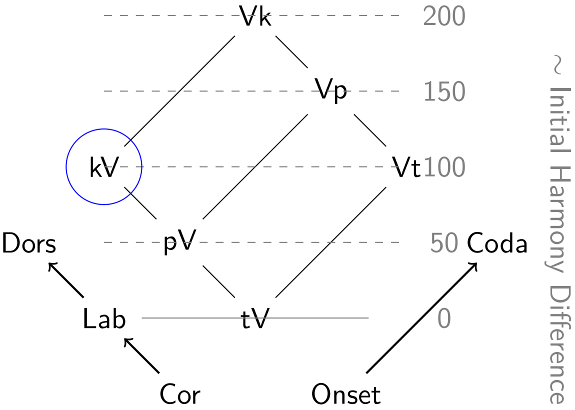

$-51 - (-50) = -1$

| | 50 | 50 | 50 | 50 | 50 | 1 | |
|---|---|---|---|---|---|---|---|
| tV | *K | *KP | *KPT | ONSET | NOCODA | MAX | HARM |
| ☞ a. tV | | | -1 | | | | -50 |
| b. V | | | | -1 | | -1 | -51 |



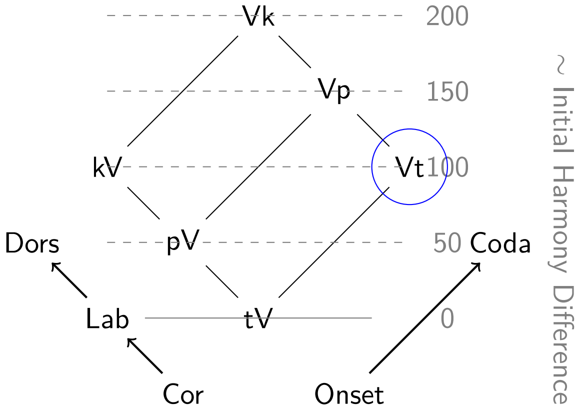

$-51 - (-100) = 49$

| pV | 50 *K | 50 *KP | 50 *KPT | 50 ONSET | 50 NOCODA | 1 MAX | HARM |
|---|---|---|---|---|---|---|---|
| ☞ a. pV | | -1 | -1 | | | | -100 |
| b. V | | | | -1 | | -1 | -51 |



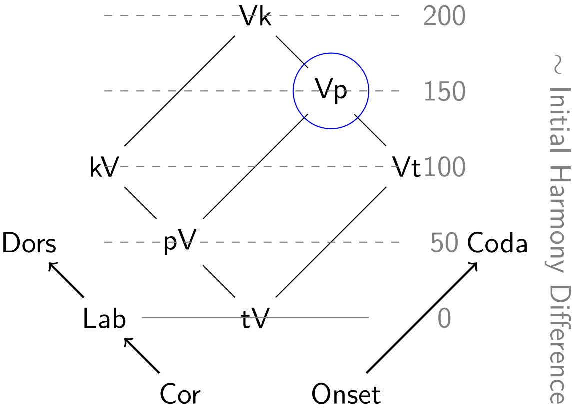

$-51 - (-150) = 99$

| | 50 | 50 | 50 | 50 | 50 | 1 | |
|---|---|---|---|---|---|---|---|
| kV | *K | *KP | *KPT | ONSET | NOCODA | MAX | HARM |
| ☞ a. kV | -1 | -1 | -1 | | | | -150 |
| b. V | | | | -1 | | -1 | -51 |



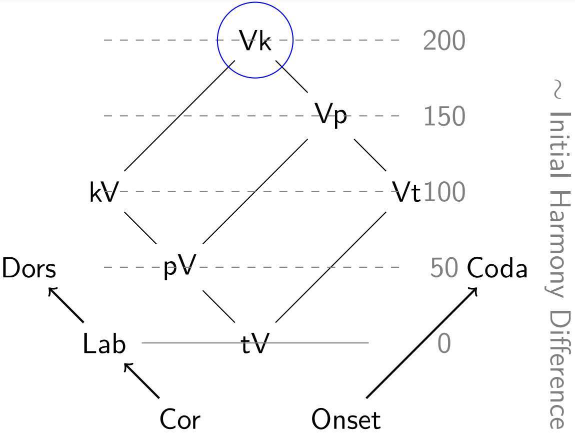

$-1-(-100)=99$

| | 50 | 50 | 50 | 50 | 50 | 1 | |
|---|---|---|---|---|---|---|---|
| Vt | *K | *KP | *KPT | ONSET | NOCODA | MAX | HARM |
| ☞ a. Vt | | | -1 | | -1 | | -100 |
| b. V | | | | | | -1 | -1 |



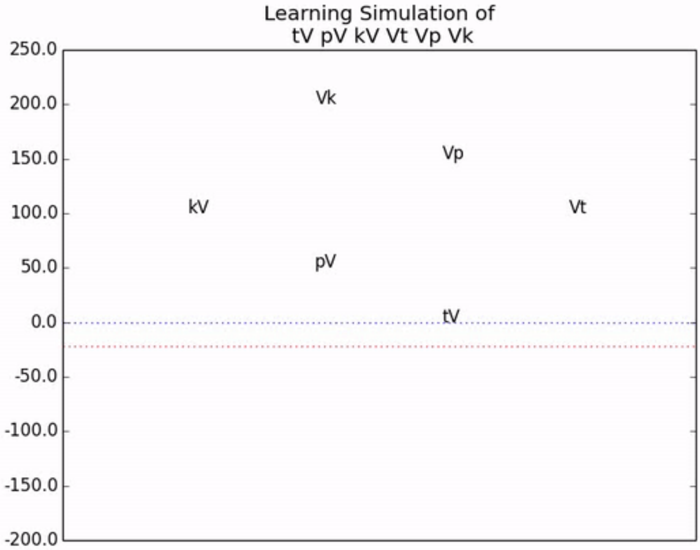

$-1-(-150)=149$

| | 50 | 50 | 50 | 50 | 50 | 1 | |
|---|---|---|---|---|---|---|---|
| Vp | *K | *KP | *KPT | ONSET | NOCODA | MAX | HARM |
| ☞ a. Vp | | -1 | -1 | | -1 | | -150 |
| b. V | | | | | | -1 | -1 |




$-1-(-200)=199$

| | 50 | 50 | 50 | 50 | 50 | 1 | |
|---|---|---|---|---|---|---|---|
| Vk | *K | *KP | *KPT | ONSET | NOCODA | MAX | HARM |
| ☞ a. Vk | -1 | -1 | -1 | | -1 | | -200 |
| b. V | | | | | | -1 | -1 |


Learning Simulation of
tV pV kV Vt Vp Vk
250.0
200.0
150.0
100.0
50.0
0.0
-50.0
-100.0
-150.0
-200.0
Vk
Vp
kV
Vt
pV
tV

# Update Rule

Error-Driven Perceptron Algorithm (Rosenblatt, 1958; Boersma & Pater, 2016)

- 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]
- 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 |

# Update Rule

Error-Driven Perceptron Algorithm (Rosenblatt, 1958; Boersma & Pater, 2016)

- 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]
- 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 |

# Update Rule

Error-Driven Perceptron Algorithm (Rosenblatt, 1958; Boersma & Pater, 2016)

- 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]
- 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 |

# Generational Stability Results

Stability Rates formulated by running simulation 50 times using Soft Typology Tool (O'Hara, 2017)[24] for each pattern.

- A run is considered to be *stable* if: $\max(G_{20}(x)) = G_0(x)$

[24] http://dornsife.usc.edu/ohara/stt/

# Generational Stability Results

Stability Rates formulated by running simulation 50 times using Soft Typology Tool (O'Hara, 2017)[24] for each pattern.

- A run is considered to be *stable* if: $\max(G_{20}(x))=G_0(x)$

| Pattern | Stability (%) | |
|---|---|---|
| All Codas | 84% | |
| PT-Codas | 48% | |
| T-Codas | 0% | |
| No Codas | 92% | |
| No Dorsals | 0% | |

[24] http://dornsife.usc.edu/ohara/stt/

# Simulations capture observed soft generalizations

- Soft Generalizations:
  - ✓ [**Vt**] ***usually*** implies [**Vp**]
  - ✓ [**Vp**] *typically* implies [***Vk***]
- Categorical Generalization:
  - ✓ [**Vt**] implies [kV]

| Pattern | Stability (%)/Observed (lgs) | |
|---|---|---|
| All Codas | 84%<br>35 lgs | |
| PT-Codas | 48%<br>6 lgs | |
| T-Codas | 0%<br>1 lg | |
| No Codas | 92%<br>31 lgs | |
| No Dorsals | 0%<br>0 lgs | |

## Simulation Details

- Most patterns *decayed*—marked forms were lost over time.
  - T-Coda runs quickly lost [**Vt**].
  - Unstable PT-Coda runs lost [**Vp**] eventually, and then quickly lost [**Vt**].
- Not the case for the No Dorsals runs.
  - In this case [kV] was learned accidentally while trying to learn [**Vp**]
- All patterns moved towards No-Codas.

# Simulation Details

- Most patterns *decayed*—marked forms were lost over time.
  - T-Coda runs quickly lost [**Vt**].
  - Unstable PT-Coda runs lost [**Vp**] eventually, and then quickly lost [**Vt**].
- Not the case for the No Dorsals runs.
  - In this case [kV] was learned accidentally while trying to learn [**Vp**]
- All patterns moved towards No-Codas.

# Simulation Details

- Most patterns *decayed*—marked forms were lost over time.
  - T-Coda runs quickly lost [**Vt**].
  - Unstable PT-Coda runs lost [**Vp**] eventually, and then quickly lost [**Vt**].
- Not the case for the No Dorsals runs.
  - In this case [kV] was learned accidentally while trying to learn [**Vp**]
- All patterns moved towards No-Codas.

# Simulation Details

- Most patterns *decayed*—marked forms were lost over time.
  - T-Coda runs quickly lost [**Vt**].
  - Unstable PT-Coda runs lost [**Vp**] eventually, and then quickly lost [**Vt**].
- Not the case for the No Dorsals runs.
  - In this case [kV] was learned accidentally while trying to learn [**Vp**]
- All patterns moved towards No-Codas.

# Simulation Details

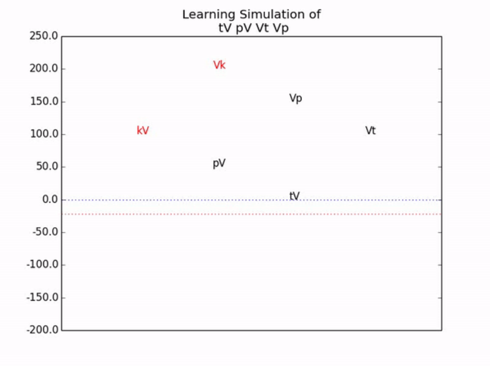
- Most patterns *decayed*—marked forms were lost over time.
  - T-Coda runs quickly lost [**Vt**].
  - Unstable PT-Coda runs lost [**Vp**] eventually, and then quickly lost [**Vt**].
- Not the case for the No Dorsals runs.
  - In this case [kV] was learned accidentally while trying to learn [**Vp**]
- All patterns moved towards No-Codas.

# Simulation Details

- Most patterns *decayed*—marked forms were lost over time.
  - T-Coda runs quickly lost [**Vt**].
  - Unstable PT-Coda runs lost [**Vp**] eventually, and then quickly lost [**Vt**].
- Not the case for the No Dorsals runs.
  - In this case [kV] was learned accidentally while trying to learn [**Vp**]
- All patterns moved towards No-Codas.

# Why is No-Dorsals so unstable?

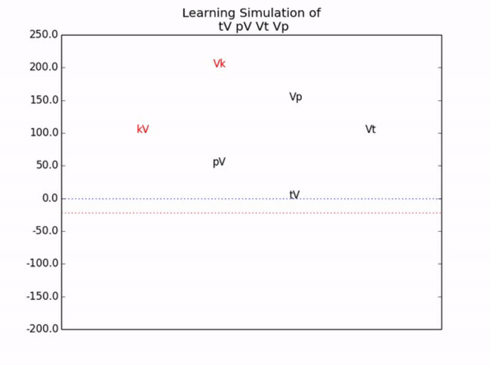
- Why is [kV] learned when training on the No-Dorsals pattern?




- [kV] is initially less marked than [**Vp**].
- MAX has less far to move to make [kV] licit in the language, than [**Vp**].
- Even though [**Vp**] moves faster,

# Why is No-Dorsals so unstable?

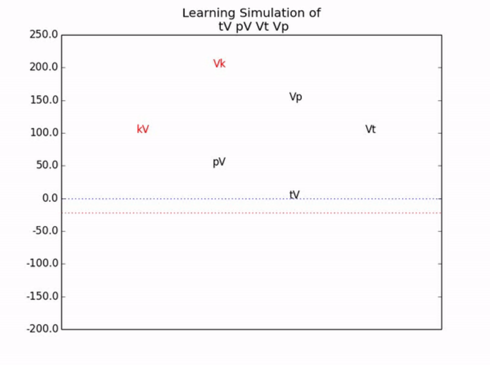
- Why is [kV] learned when training on the No-Dorsals pattern?




- [kV] is initially less marked than [**Vp**].
- MAX has less far to move to make [kV] licit in the language, than [**Vp**].
- Even though [**Vp**] moves faster,

# Why is No-Dorsals so unstable?

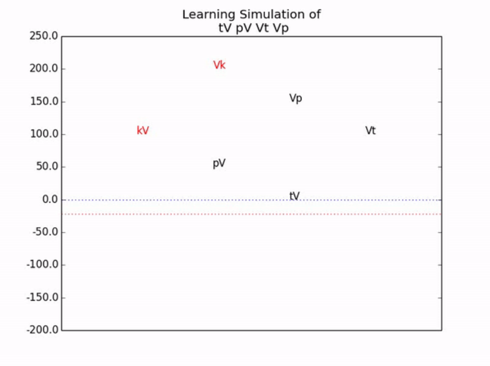
- Why is [kV] learned when training on the No-Dorsals pattern?




- [kV] is initially less marked than [**Vp**].
- MAX has less far to move to make [kV] licit in the language, than [**Vp**].
- Even though [**Vp**] moves faster,

# Why is No-Dorsals so unstable?

- Why is [kV] learned when training on the No-Dorsals pattern?



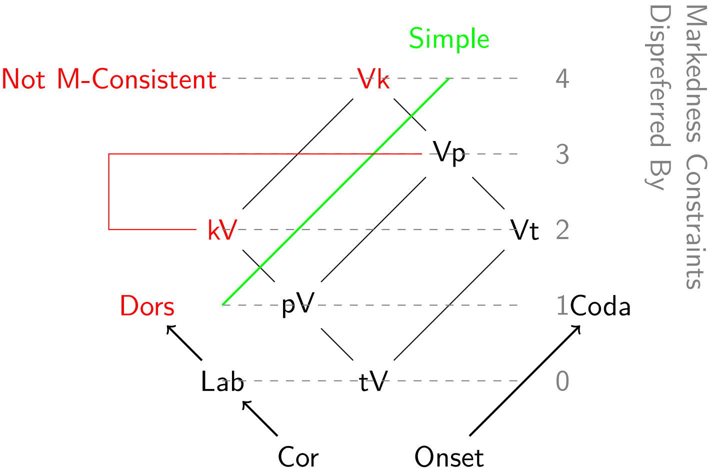

- [kV] is initially less marked than [**Vp**].
- MAX has less far to move to make [kV] licit in the language, than [**Vp**].
- Even though [**Vp**] moves faster, it doesn't catch up in time.

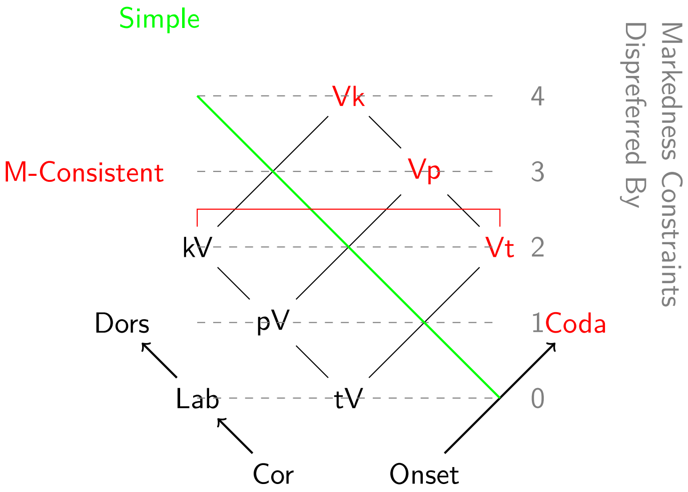
# Markedness Consistency

Learning Bias towards Markedness Consistent patterns.

- **Markedness Consistent**- A pattern $P$ is markedness consistent iff for $x, y$; $x$ is dispreferred by more net markedness constraints than $y$ is, and $x$ is part of a pattern, $y$ is as well.
    - i.e. The most marked form allowed is not more marked than the least marked banned form.
- Learners prefer the patterns that are both *simple* and *markedness consistent*.

Common patterns are both *simple* and *markedness consistent*.




Simple
M-Consistent
Vk
Vp
kV
Vt
pV
Dors
Lab
tV
Cor
Onset
Coda
4
3
2
1
0
Markedness Constraints
Dispreferred By

# Conclusion

**Simplicity is not sufficient**—Simplicity alone cannot explain the generalization against the No Dorsals pattern.
**Learning Bias can do more**—The No Dorsals pattern is less stable than the more common patterns, due to interactions of the learning algorithm and the grammar, which cannot be collapsed into simplicity.

- **Common** and **Easy to Learn** patterns are **Simple** and **Markedness Consistent**

# Implications

- With hard coded markedness constraints (a substantive bias), the typology can be captured.
- Can a bias like this Markedness Consistency bias arise if substantive bias is not present in the phonological grammar, but is due to channel bias?
  - Stay tuned, but looks like no
- How does stability extend to other domains?
  - Soft Typ Tool (O'Hara, 2017).

# Works Cited I


ANTTILA, ARTO. 1997. Deriving variation from Grammar. *Pages 35–68 of:* HINSKENS, FRANS, VAN HOUT, ROELAND, & WETZELS, LEO (eds), *Variation, Change and Phonological Theory*. Philadelphia, PA: John Benjamins.

BANE, MAX, & RIGGLE, JASON. 2008. Three Correlates of the Typological Frequency of Quantity-Insensitive Stress Systems. *In: Proceedings of the tenth meeting of the ACL Special Interest Group on Computational Morphology and Phonology*.

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.

CLARK, ROSS. 1976. *Aspects of Proto-Polynesian syntax*. Te Reo Monograph.

CLEMENTS, G. N. 2003. Feature economy in sound systems. *Phonology*, **20**, 287–333.

CLIFTON, JOHN M. 1995. *Vanimo Organised Phonology Data*. Tech. rept. SIL.

COETZEE, ANDRIES. 2002. *Between-language frequency effects in phonological theory*. ms. University of Massachusetts Amherst.

CULBERTSON, JENNIFER, SMOLENSKY, PAUL, & WILSON, COLIN. 2013. Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in cognitive science*, **5**(3), 392–424.

DE GROOT, A. W. 1931. Phonologie und Phonetik als Funktionswissenschaften. *Travaux du Cercle Linguistique de Prague*, **4**, 116–147.

DE LACY, PAUL. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.

DONOHUE, MARK. 2010. Skou. *In: Working Papers of the Automated Similarity Judgement Program*.

# Works Cited II

Dowman, Mike, Kirby, Simon, & Griffiths, Thomas L. 2006. Innateness and Culture in the evolution of language. *In:* Cangelosi, A, Smith, A., & Smith, K. (eds), *The Evolution of Language: Proceedings of the 6th international conference on the Evolution of Language*.

Elbert, Samuel H., & Pukui, Mary Kawena. 1979. *Hawaiian grammar*. Honolulu: University of Hawaii Press.

Estevam, Adriana Machado. 2011. *Morphosyntaxe Du Xavante*. Ph.D. thesis, Universite Paris.Diderot (Paris 7).

Goldsmith, John A. 1990. *Autosegmental and Metrical Phonology*. Blackwell.

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.

Haude, Katharina. 2006. *A grammar of Movima*. Ph.D. thesis, Radboud Universiteit Nijmegen.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore, Maryland: Waverly Press.

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

Hughto, Coral, & Pater, Joe. 2017. *Emergence of strict domination effects with weighted constraints*. Talk given at the CLS Workshop on Dynamic Modeling in Phonetics and Phonology.

Hughto, Coral, Staubs, Robert, & Pater, Joe. 2014. *Typological consequences of agent interaction*. Presented at the Northeast Computational Phonology Circle (NECPhon 8).

Jakobson, Roman, & Halle, Morris. 1956. *Fundamentals of language*. The Hague: Mouton.

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

# Works Cited III

Kean, Mary-Louise. 1975 (June). *The Theory of Markedness in Generative Grammar*. Ph.D. thesis, Massachusetts Institute of Technology.

Kingston, John. 1985. *The Phonetics and Phonology of the Timing of Oral and Glottal Events*. Ph.D. thesis, University of California, Berkeley.

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*.

Li, Fang-Kuei. 1946. Chipewyan. *Pages 398–423 of:* Hoijer, Harry, Bloomfield, Leonard, & Haas, Mary R. (eds), *Linguistic Structures of Native America*, vol. 6. New York: Johnson Reprint Corporation.

Lombardi, Linda. 2001. Why place and voice are different: Constraint-specific alternations in Optimality Theory. *Pages 13–45 of:* Lombardi, Linda (ed), *Segmental phonology in Optimality Theory: Constraints and Representations*. Cambridge: Cambridge University Press.

Marmion, Douglas Edric. 2010. *Topics in the Phonology and Morphology of Wutung*. Ph.D. thesis, The Australian National University.

Martinet, André. 1968. Phonetics and linguistic evolution. *Pages 464–487 of:* Malmberg, Bertil (ed), *Manual of phonetics*.

McCarthy, John J. 1988. Feature Geometry and Dependency: A Review. *Phonetica*, **43**(45), 84–108.

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

McLeod, Ruth, & Mitchell, Valerie. 2003. *Aspectos da língua xavante*. SIL International.

O'Hara, Charlie. *Word-Edge Consonant Database (ADD LINK)*.

O'Hara, Charlie. 2017. *Soft Typology Tool v. 0.1*. Software package http://dornsife.usc.edu/ohara/stt/.

# Works Cited IV

PANKRATZ, LEO, & PIKE, EUNICE V. 1967. Phonology and morphotonemics of Ayulta Mixtec. *International Journal of American Linguistics*, **33**(287-99).

PATER, JOE. 2012. Serial Harmonic Grammar and Berber syllabification. *Pages 43–72 of:* BOROWSKY, TONI, KAWAHARA, SHIGETO, SHINYA, TAKAHITO, & SUGAHARA, MARIKO (eds), *Prosody Matters: Essays in Honor of Elisabeth O. Selkirk*. Equinox Press.

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.

RIGGLE, JASON. 2014. The Sampling Model of Variation. *In: Proceedings of the 46th annual meeting of the Chicago Linguistics Society*.

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

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

TRYON, D. T. 1970. *Conversational Tahitian: An Introduction to the Tahitian Language of French Polynesia*. Australian National University Press.

# Works Cited I


ANTTILA, ARTO. 1997. Deriving variation from Grammar. *Pages 35–68 of:* HINSKENS, FRANS, VAN HOUT, ROELAND, & WETZELS, LEO (eds), *Variation, Change and Phonological Theory*. Philadelphia, PA: John Benjamins.

BANE, MAX, & RIGGLE, JASON. 2008. Three Correlates of the Typological Frequency of Quantity-Insensitive Stress Systems. *In: Proceedings of the tenth meeting of the ACL Special Interest Group on Computational Morphology and Phonology*.

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.

CLARK, ROSS. 1976. *Aspects of Proto-Polynesian syntax*. Te Reo Monograph.

CLEMENTS, G. N. 2003. Feature economy in sound systems. *Phonology*, **20**, 287–333.

CLIFTON, JOHN M. 1995. *Vanimo Organised Phonology Data*. Tech. rept. SIL.

COETZEE, ANDRIES. 2002. *Between-language frequency effects in phonological theory*. ms. University of Massachusetts Amherst.

CULBERTSON, JENNIFER, SMOLENSKY, PAUL, & WILSON, COLIN. 2013. Cognitive biases, linguistic universals, and constraint-based grammar learning. *Topics in cognitive science*, **5**(3), 392–424.

DE GROOT, A. W. 1931. Phonologie und Phonetik als Funktionswissenschaften. *Travaux du Cercle Linguistique de Prague*, **4**, 116–147.

DE LACY, PAUL. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.

DONOHUE, MARK. 2010. Skou. *In: Working Papers of the Automated Similarity Judgement Program*.

# Works Cited II

Dowman, Mike, Kirby, Simon, & Griffiths, Thomas L. 2006. Innateness and Culture in the evolution of language. *In:* Cangelosi, A, Smith, A., & Smith, K. (eds), *The Evolution of Language: Proceedings of the 6th international conference on the Evolution of Language*.

Elbert, Samuel H., & Pukui, Mary Kawena. 1979. *Hawaiian grammar*. Honolulu: University of Hawaii Press.

Estevam, Adriana Machado. 2011. *Morphosyntaxe Du Xavante*. Ph.D. thesis, Universite Paris.Diderot (Paris 7).

Goldsmith, John A. 1990. *Autosegmental and Metrical Phonology*. Blackwell.

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.

Haude, Katharina. 2006. *A grammar of Movima*. Ph.D. thesis, Radboud Universiteit Nijmegen.

Hockett, Charles F. 1955. *A manual of phonology*. Baltimore, Maryland: Waverly Press.

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

Hughto, Coral, & Pater, Joe. 2017. *Emergence of strict domination effects with weighted constraints*. Talk given at the CLS Workshop on Dynamic Modeling in Phonetics and Phonology.

Hughto, Coral, Staubs, Robert, & Pater, Joe. 2014. *Typological consequences of agent interaction*. Presented at the Northeast Computational Phonology Circle (NECPhon 8).

Jakobson, Roman, & Halle, Morris. 1956. *Fundamentals of language*. The Hague: Mouton.

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

# Works Cited III

Kean, Mary-Louise. 1975 (June). *The Theory of Markedness in Generative Grammar*. Ph.D. thesis, Massachusetts Institute of Technology.

Kingston, John. 1985. *The Phonetics and Phonology of the Timing of Oral and Glottal Events*. Ph.D. thesis, University of California, Berkeley.

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*.

Li, Fang-Kuei. 1946. Chipewyan. *Pages 398–423 of:* Hoijer, Harry, Bloomfield, Leonard, & Haas, Mary R. (eds), *Linguistic Structures of Native America*, vol. 6. New York: Johnson Reprint Corporation.

Lombardi, Linda. 2001. Why place and voice are different: Constraint-specific alternations in Optimality Theory. *Pages 13–45 of:* Lombardi, Linda (ed), *Segmental phonology in Optimality Theory: Constraints and Representations*. Cambridge: Cambridge University Press.

Marmion, Douglas Edric. 2010. *Topics in the Phonology and Morphology of Wutung*. Ph.D. thesis, The Australian National University.

Martinet, André. 1968. Phonetics and linguistic evolution. *Pages 464–487 of:* Malmberg, Bertil (ed), *Manual of phonetics*.

McCarthy, John J. 1988. Feature Geometry and Dependency: A Review. *Phonetica*, **43**(45), 84–108.

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

McLeod, Ruth, & Mitchell, Valerie. 2003. *Aspectos da língua xavante*. SIL International.

O'Hara, Charlie. *Word-Edge Consonant Database (ADD LINK)*.

O'Hara, Charlie. 2017. *Soft Typology Tool v. 0.1*. Software package http://dornsife.usc.edu/ohara/stt/.

# Works Cited IV

Pankratz, Leo, & Pike, Eunice V. 1967. Phonology and morphotonemics of Ayulta Mixtec. *International Journal of American Linguistics*, **33**(287-99).

Pater, Joe. 2012. Serial Harmonic Grammar and Berber syllabification. *Pages 43–72 of:* Borowsky, Toni, Kawahara, Shigeto, Shinya, Takahito, & Sugahara, Mariko (eds), *Prosody Matters: Essays in Honor of Elisabeth O. Selkirk*. Equinox Press.

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.

Riggle, Jason. 2014. The Sampling Model of Variation. *In: Proceedings of the 46th annual meeting of the Chicago Linguistics Society*.

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

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

Tryon, D. T. 1970. *Conversational Tahitian: An Introduction to the Tahitian Language of French Polynesia*. Australian National University Press.

# Not Just No Dorsals

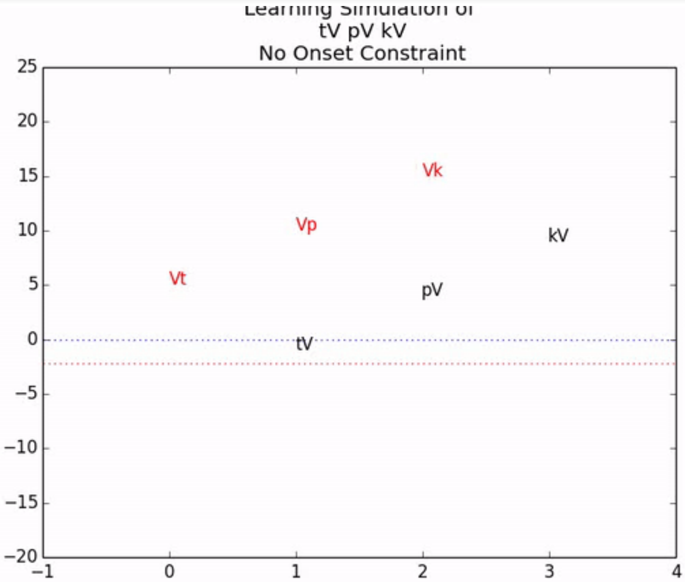
No language with less than three (supralaryngeal) stops word-initially, have any word-finally.

| Language | Family | Initial | Final | |
|---|---|---|---|---|
| Xavante | Macro-Ge | p t ʔ | ∅ | (McLeod & Mitchell, 2003; Estevam, 2011) |
| Tahitian | Austronesian | p t ʔ | ∅ | (Tryon, 1970) |
| Wutung | Skou | p t ʔ | ∅ | (Marmion, 2010) |
| Vanimo | Skou | p t ʔ | ∅ | (Clifton, 1995) |
| Nouri | Skou | p t | ∅ | (Donohue, 2010) |
| Hawaiian | Austronesian | k p ʔ | ∅ | (Elbert & Pukui, 1979) |
| Yellowknife Chipewyan | Na-Dene | k p ʔ | ∅ | (Li, 1946; de Lacy, 2006) |
| Colloquial Samoan | Austronesian | k p ʔ | ∅ | (Clark, 1976) |
| Ayutla Mixtec | Oto-Manguean | k t ʔ | ʔ | (Pankratz & Pike, 1967) |

# What happens if you get rid of ONSET

Why is [**Vp**] initially more marked than [kV]

- ONSET and NOCODA are separate constraints making the same distinction in this domain.
- By removing one of these constraints, a different picture arises.
- Without ONSET, the No-Codas pattern becomes unstable, and the No-Dorsals pattern becomes the attractor.
  - No Codas loses [kV], while gaining [**Vt**].


Learning Simulation of
tV pV kV
No Onset Constraint
Vk
Vp
kV
Vt
pV
tV
25
20
15
10
5
0
−5
−10
−15
−20
−1
0
1
2
3
4

# R-Volume Results

R-volume[25] has been used to model typological frequency.

- Calculated by sampling weights from uniform distribution 0-20
- Fails here to get the simplicity all-or-nothing bias.

| | Onset | | | Coda | | | Percentage |
|---|---|---|---|---|---|---|---|
| Zero | ✗ | ✗ | ✗ | ✗ | ✗ | ✗ | 15.9 |
| T | tV | ✗ | ✗ | ✗ | ✗ | ✗ | 30.8 |
| Tt | tV | ✗ | ✗ | Vt | ✗ | ✗ | 3.6 |
| TP | tV | pV | ✗ | ✗ | ✗ | ✗ | 22.9 |
| TPt | tV | pV | ✗ | Vt | ✗ | ✗ | 3.8 |
| No Dorsals | tV | pV | ✗ | Vt | Vp | ✗ | 1.5 |
| No Codas | tV | pV | kV | ✗ | ✗ | ✗ | 12.3 |
| T-Codas | tV | pV | kV | Vt | ✗ | ✗ | 5.6 |
| PT-Codas | tV | pV | kV | Vt | Vp | ✗ | 2.8 |
| All Codas | tV | pV | kV | Vt | Vp | Vk | .8 |

[25] (Anttila, 1997; Coetzee, 2002; Riggle, 2014)