Soft Typology of Coda Place of Articulation Distributions Requires Synchronic Constraints

Charlie O'Hara
University of Southern California

Workshop on Emergence of Universals
February 19, 2018

Soft Typology

Crucial goal of linguistic theory is to explain generalizations observed across typology.

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

Emergent Biases

Why would a pattern be less likely than another pattern?

- **Inductive Bias:** It is more difficult to learn, even with pristine training data.
- **Channel Bias:** It is more difficult to learn, due to asymmetries in mistransmission of training data.

Structure vs. Substance

Biases can be divided based on their effects:

- Structural Bias—A bias for featurally simple patterns
  i.e. A pattern based on one feature is preferred to one based on two.
- Substantive Bias—Bias based on the substance of the features rather than abstract complexity.
  i.e. One pattern based on only one feature is preferred to another based only on some other feature.

Structure vs. Substance

Biases can be divided based on their effects:

- Structural Bias—A bias for featurally simple patterns
  i.e. A pattern based on one feature is preferred to one based on two.
- Substantive Bias—Bias based on the substance of the features rather than abstract complexity.
  i.e. One pattern based on only one feature is preferred to another based only on some other feature.

SIMPLE				**COMPLEX**		
***Coda**				***[Coda**+Dorsal]		
kV	pV	tV	>	kV	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**

Structure vs. Substance

Biases can be divided based on their effects:

- Structural Bias—A bias for featurally simple patterns
  i.e. A pattern based on one feature is preferred to one based on two.
- Substantive Bias—Bias based on the substance of the features rather than abstract complexity.
  i.e. One pattern based on only one feature is preferred to another based only on some other feature.

SIMPLE				**SIMPLE**		
***Coda**				*Dorsal		
kV	pV	tV	>	~~kV~~	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**

Structurally Biased Phonology

How do these biases interact?

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

	Structure	Substance
Inductive	✓	✗
Channel	?	✓

- Structural inductive bias:
  - Appears similarly in other domains of pattern-learning (Moreton & Pertsova, 2014)
  - Is well documented in artificial grammar learning experiments (Moreton & Pater, 2012)
- No Substantive Inductive Bias:
  - Less evidence that substantive bias emerges in artificial grammar learning (Moreton & Pater, 2012; Glewwe, 2017)
  - Much of substantive bias can be explained through channel bias, without requiring any innate biases (Blevins, 2004)

Substantive Bias in place of articulation inventories

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

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

SIMPLE				**COMPLEX**				**SIMPLE**		
***Coda**				***[Coda**+Dorsal]				*Dorsal		
kV	pV	tV	>	kV	pV	tV	>	~~kV~~	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	**Vp**	**Vt**

Substantive Bias in place of articulation inventories

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

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

SIMPLE				**COMPLEX**				**SIMPLE**		
***Coda**				***[Coda**+Dorsal]				*Dorsal		
kV	pV	tV	>	kV	pV	tV	>	~~kV~~	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	**Vp**	**Vt**

Substantive Bias in place of articulation inventories

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

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

SIMPLE				**COMPLEX**				**SIMPLE**		
***Coda**				***[Coda**+Dorsal]				*Dorsal		
kV	pV	tV	>	kV	pV	tV	>	~~kV~~	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	**Vp**	**Vt**

Substantive Bias in place of articulation inventories

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

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

SIMPLE				**COMPLEX**				**SIMPLE**		
***Coda**				***[Coda**+Dorsal]*				***Dorsal**		
kV	pV	tV	>	kV	pV	tV	>	~~kV~~	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	**Vp**	**Vt**

Explaining lack of No-Dorsals

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

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

Do we need substantive inductive bias?

	Structure	Substance
Inductive	✓	✗
Channel	?	✓

Explaining lack of No-Dorsals

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

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

Do we need substantive inductive bias?

	Structure	Substance
Inductive	✓	✓?
Channel	?	✓

Explaining lack of No-Dorsals

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

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

Do we need substantive inductive bias?

	Structure	Substance
Inductive	✓	✗
Channel	?	✓

Explaining lack of No-Dorsals

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

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

Do we need substantive inductive bias?

	Structure	Substance
Inductive	✓	✓?
Channel	?	✓

Typological Survey

Development of the Word Edge Consonant Database (WECD)

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

[1] That allow maximally three supralaryngeal places of articulation for stops

Typological Survey

Development of the Word Edge Consonant Database (WECD)

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

[1] That allow maximally three supralaryngeal places of articulation for stops

Markedness Hierarchies

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

- Onset vs. coda [2]
  CV $\succ$ VC
- Place of articulation[3]
  Coronal $\succ_{\frac{10}{13}}$ Labial $\succ_{\frac{12}{14}}$ Dorsal

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

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

Results of Typological Survey

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

	Onset			Coda				
No Codas	tV	pV	Vk	✗	✗	✗	27	
1 Coda	tV	pV	Vk	**Vt**	✗	✗	2	
2 Codas	tV	pV	Vk	**Vt**	**Vp**	✗	5	
3 Codas	tV	pV	Vk	**Vt**	**Vp**	**Vk**	43	

No Coda Pattern

	Onset			Coda			
No Codas	tV	pV	kV	✗	✗	✗	27
1 Coda	tV	pV	kV	**Vt**	✗	✗	2
2 Codas	tV	pV	kV	**Vt**	**Vp**	✗	5
3 Codas	tV	pV	kV	**Vt**	**Vp**	**Vk**	43

- Example: Italian

['**t**asto] *button* ['**p**asto] *meal* ['**k**asto] *chaste*

*[kasa**t**] *[kasa**p**] *[kasa**k**]

Only 1 Coda Pattern

	Onset			Coda				
No Codas	tV	pV	kV	✗	✗	✗	27	
1 Coda	tV	pV	kV	**Vt**	✗	✗	2	
2 Codas	tV	pV	kV	**Vt**	**Vp**	✗	5	
3 Codas	tV	pV	kV	**Vt**	**Vp**	**Vk**	43	

- Example: Finnish[4]

[**t**elata] *to paint* [**p**elata] *to play* [**k**elata] *to wind*
[keo**t**] *anthills* *[keo**p**] *[keo**k**]

[4] Nimboran (Anceaux, 1965) has only [p] word finally.

2 Codas Pattern

	Onset			Coda			
No Codas	tV	pV	kV	✗	✗	✗	27
1 Coda	tV	pV	kV	**Vt**	✗	✗	2
2 Codas	tV	pV	kV	**Vt**	**Vp**	✗	5
3 Codas	tV	pV	kV	**Vt**	**Vp**	**Vk**	43

- Example: Movima (Haude, 2006)[5]

['**t**anna] *I cut* [**p**ɛnna] *my landing place* [**k**anan] *your food*

[tʃuː'ha**t**] *palm tree* [kuː'ɗu**p**] *flea* *[kuː'ɗu**k**]

[5] Korowai (Enk & Vries, 1997) and Navaho Haile (1926 (1974)) show [k p] and [k t] respectively.

3 Codas Pattern

	Onset			Coda				
No Codas	tV	pV	kV	✗	✗	✗	27	
1 Coda	tV	pV	kV	**Vt**	✗	✗	2	
2 Codas	tV	pV	kV	**Vt**	**Vp**	✗	5	
3 Codas	tV	pV	kV	**Vt**	**Vp**	**Vk**	43	

- Example: English

[**t**ap] *top* [**p**ap] *pop* [**k**ap] *cop*
[pa**t**] *pot* [pa**p**] *pop* [pa**k**] *pock*

Structural Bias

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

	Onset			Coda				
No Codas	tV	pV	kV	✗	✗	✗	27	
1 Coda	tV	pV	kV	**Vt**	✗	✗	2	
2 Coda	tV	pV	kV	**Vt**	**Vp**	✗	5	
3 Codas	tV	pV	kV	**Vt**	**Vp**	**Vk**	43	

Simplicity bias predicts these generalizations.

***Coda**			>	***[Coda**+Dorsal]			>	***[Coda**+[Dor∨Lab]]		
kV	pV	tV		kV	pV	tV		kV	pV	tV
~~**Vk**~~	~~**Vp**~~	~~**Vt**~~		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	~~**Vp**~~	**Vt**

Anti-Structural Bias

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

*[**Coda**+Dorsal]			<	*Dorsal			>	*[Dorsal∨**Coda**]		
kV	pV	tV		~~kV~~	pV	tV		~~kV~~	pV	tV
~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	~~**Vp**~~	~~**Vt**~~

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

Anti-Structural Bias

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

*[**Coda**+Dorsal]				*Dorsal				*[Dorsal∨**Coda**]		
kV	pV	tV	>	~~kV~~	pV	tV	<	~~kV~~	pV	tV
~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	**Vp**	**Vt**		~~**Vk**~~	~~**Vp**~~	~~**Vt**~~

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

Initial	Final	Language
p t ʔ	∅	Tahitian, Vanimo, Wutung, Xavante
p t	∅	Nouri
k p ʔ	∅	Hawaiian, Yellowknife Chipewyan, Colloquial Samoan
k t ʔ	ʔ	Ayulta Mixtec

Typological Generalization

Bias against losing a place of articulation

Results of Typological Survey

Structural	kV **Vk**	pV **Vp**	tV **Vt**	Simple	43	
	kV ~~**Vk**~~	pV **Vp**	tV **Vt**	Complex	5	
	kV ~~**Vk**~~	pV ~~**Vp**~~	tV ~~**Vt**~~	Simple	27	
Anti-Structural	kV ~~**Vk**~~	pV **Vp**	tV **Vt**	Complex	5	
	~~kV~~ ~~**Vk**~~	pV **Vp**	tV **Vt**	Simple	0	
	~~kV~~ ~~**Vk**~~	pV ~~**Vp**~~	tV ~~**Vt**~~	Complex	9	

Modeling

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

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

	Structure	Substance
Inductive	✓	✓?
Channel	✗	✓

Modeling

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

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

	Structure	Substance
Inductive	✓	✓
Channel	✗	✗

Modeling

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

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

	Structure	Substance
Inductive	✓	✗
Channel	✗	✗

Modeling

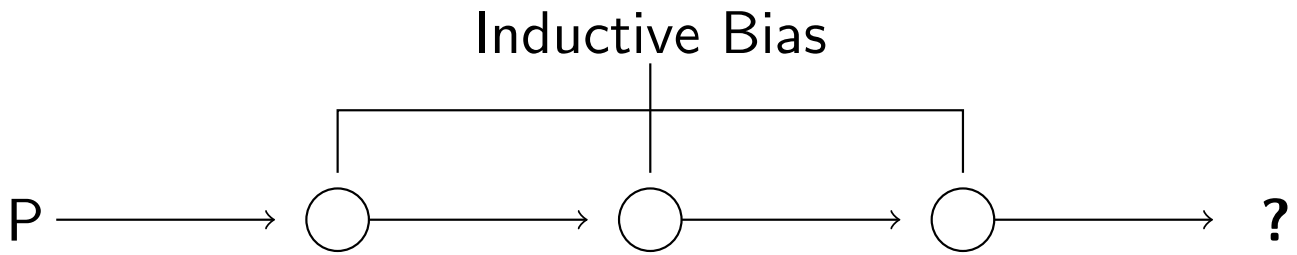
To see whether the substantive bias can be explained through inductive or channel biases three simulations were run.

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

	Structure	Substance
Inductive	✓	✗
Channel	✗	✓

Generational Transmission

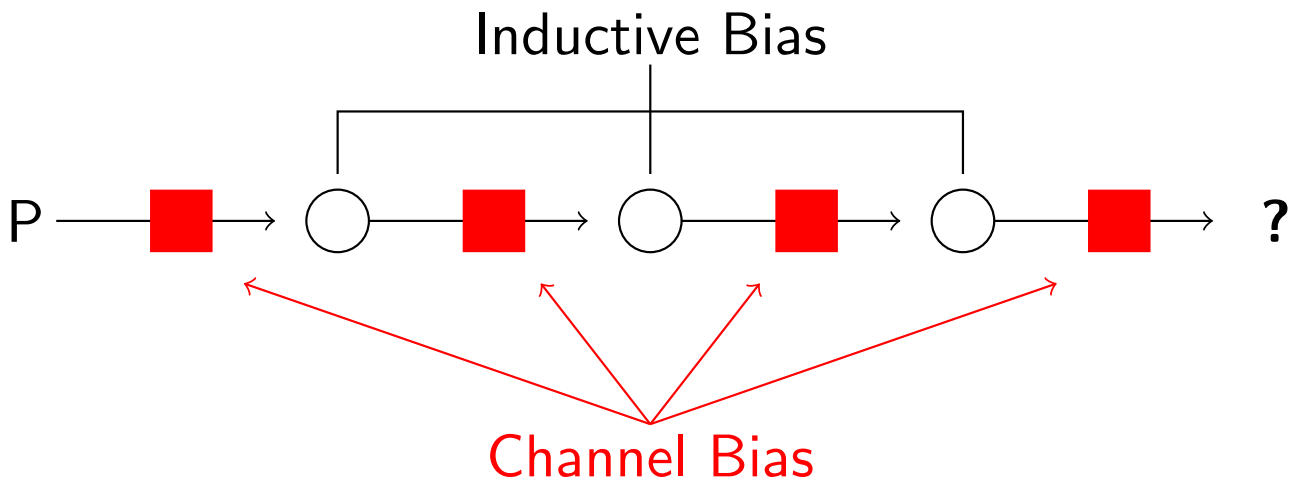
- Rate of attestation is related to stability across generations.[6]
- Generational Stability Model:
  - A stable run is one which is closer to P than any other pattern after x generations of y iterations.



[6] Staubs (2014); Kirby (2017); Hughto (2018)

Generational Transmission

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



[6] Staubs (2014); Kirby (2017); Hughto (2018)

Learners

I model learners using MaxEnt grammars[7] with the Perceptron algorithm[8].

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

[7] Goldwater & Johnson (2003)

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

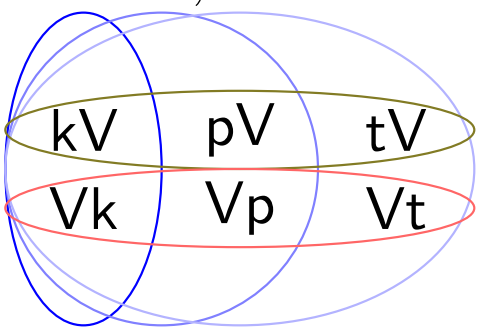
[9] Jesney & Tessier (2011)

Substantive Bias

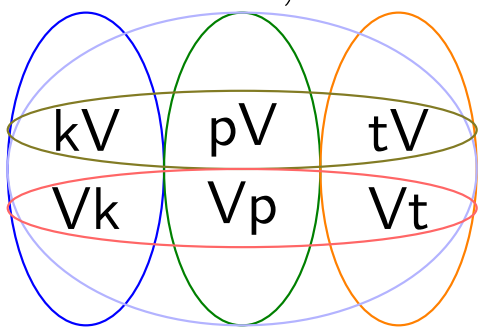
Substantive Bias can enter the system in two ways:

- Through constraint set:

Substantively biased set:
*K, *KP, *KPT, NoCoda, Onset, Max



Non-Substantively biased set:
*K, *P, *T, *KPT, NoCoda, NoOnset, Max



- Through asymmetries in misperception:

Substantively biased channel:
Mutate the data presented to a learner.

Un-biased channel:
Let learners receive pristine data.

Substantively biased constraints

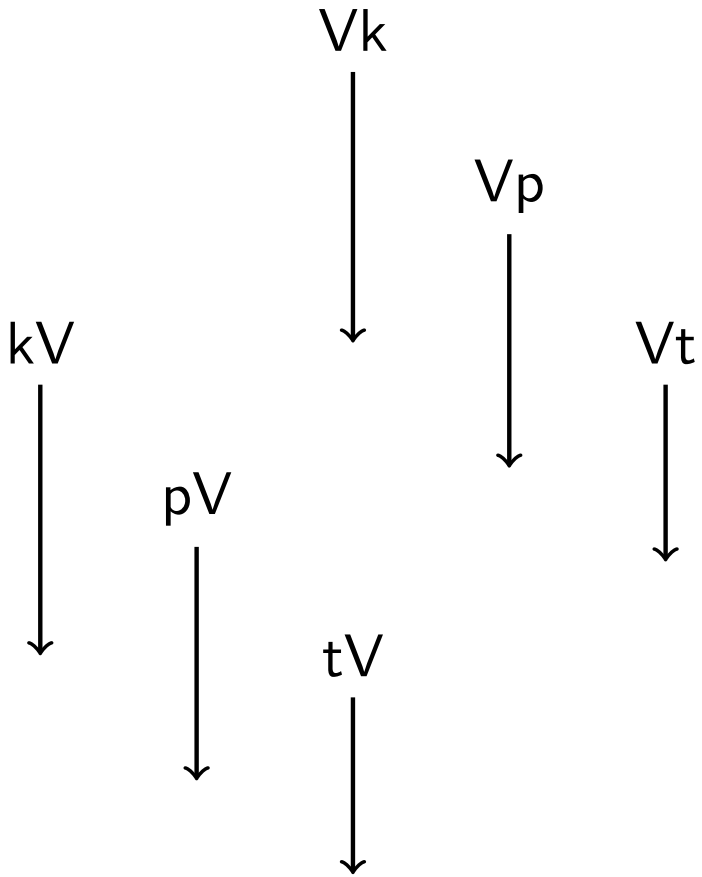
- With substantively biased constraints, both the structural, and anti-structural bias are captured.[10]

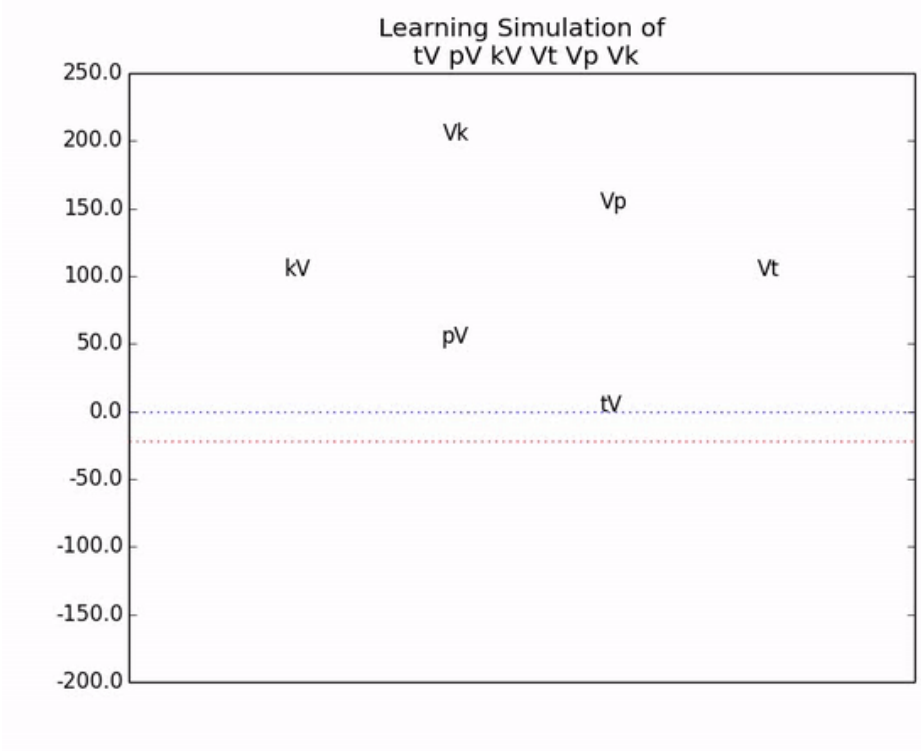
Structural	kV **Vk** pV **Vp** tV **Vt**	Simple	$\frac{42}{50}$	=	84%	
	kV ~~**Vk**~~ pV **Vp** tV **Vt**	Complex	$\frac{24}{50}$	=	48%	
	kV ~~**Vk**~~ pV ~~**Vp**~~ tV ~~**Vt**~~	Simple	$\frac{46}{50}$	=	92%	
Anti-Structural	kV ~~**Vk**~~ pV **Vp** tV **Vt**	Complex	$\frac{24}{50}$	=	48%	
	~~kV~~ ~~**Vk**~~ pV **Vp** tV **Vt**	Simple	$\frac{0}{50}$	=	0%	
	~~kV~~ ~~**Vk**~~ pV ~~**Vp**~~ tV ~~**Vt**~~	Complex	$\frac{46}{50}$	=	92%	

[10] Simulations run for 25 generations of 3200 iterations at learning rate .05

Speed vs. Distance

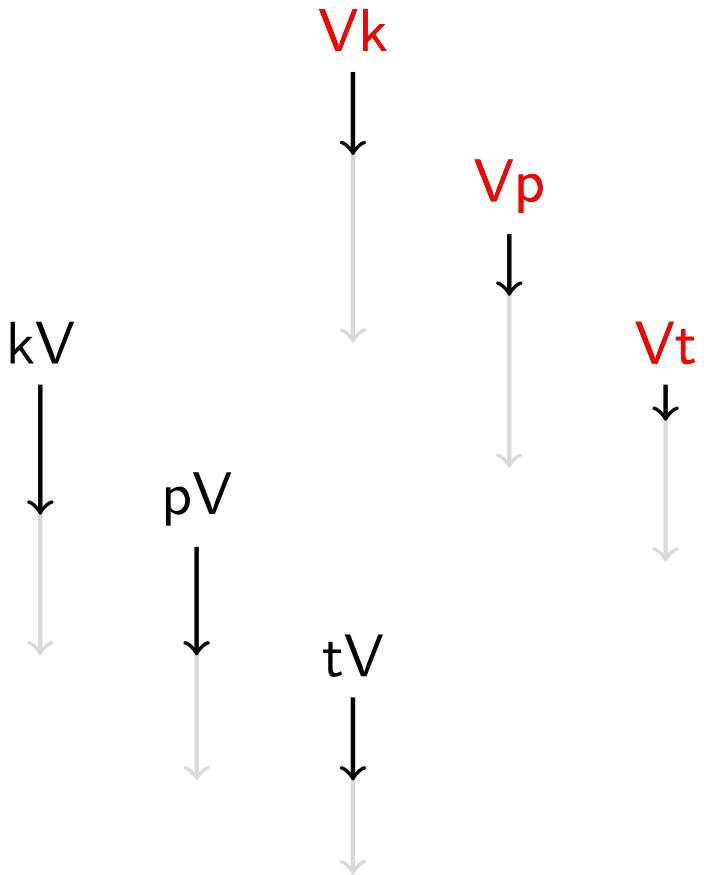
- Starting distance = Initial grammatical bias
- Speed = expected rate of change of all violated constraints

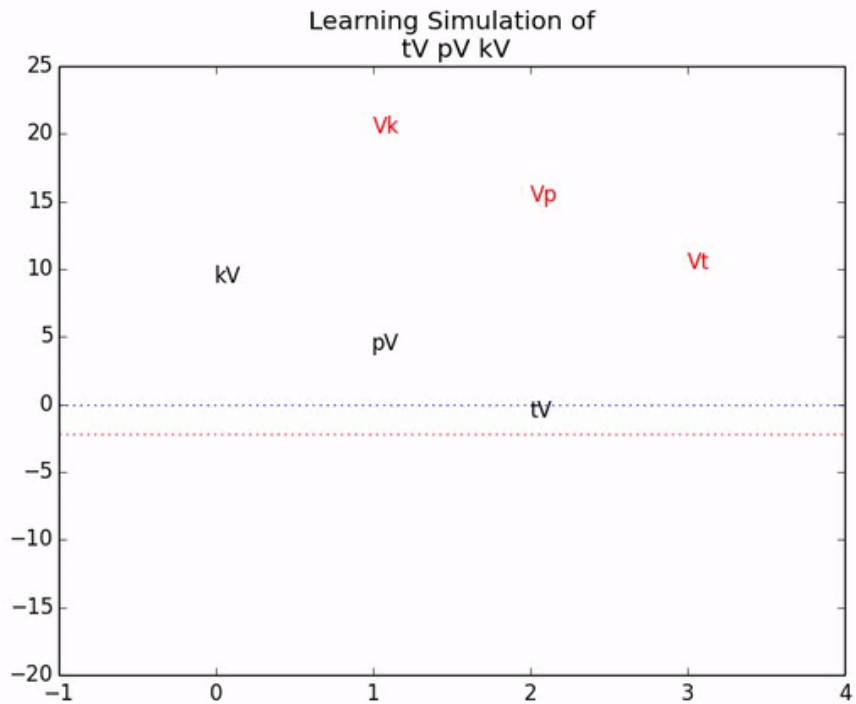




Speed vs. Distance

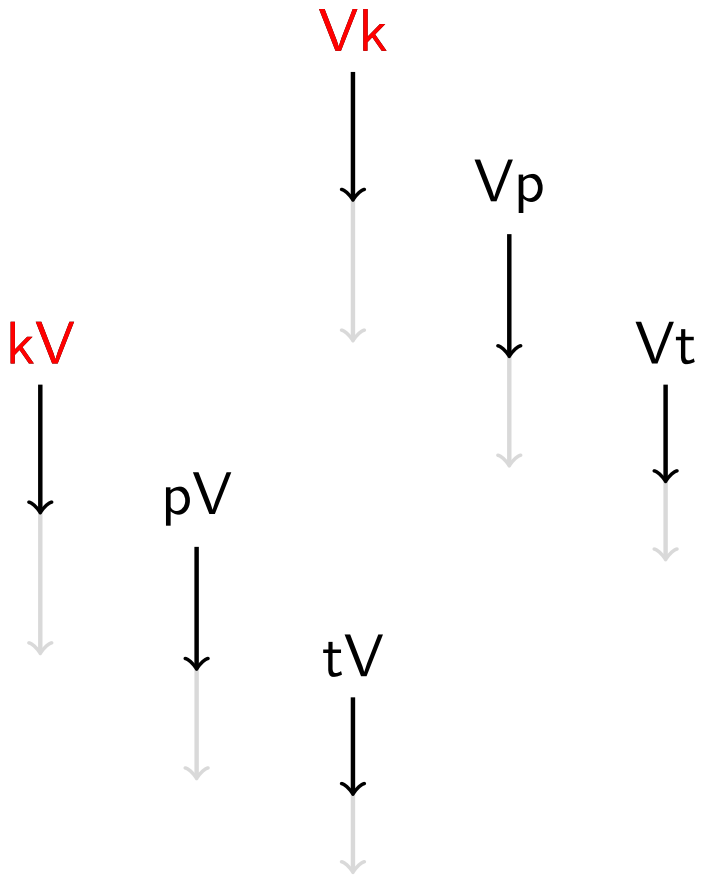
- Starting distance = Initial grammatical bias
- Speed = expected rate of change of all violated constraints

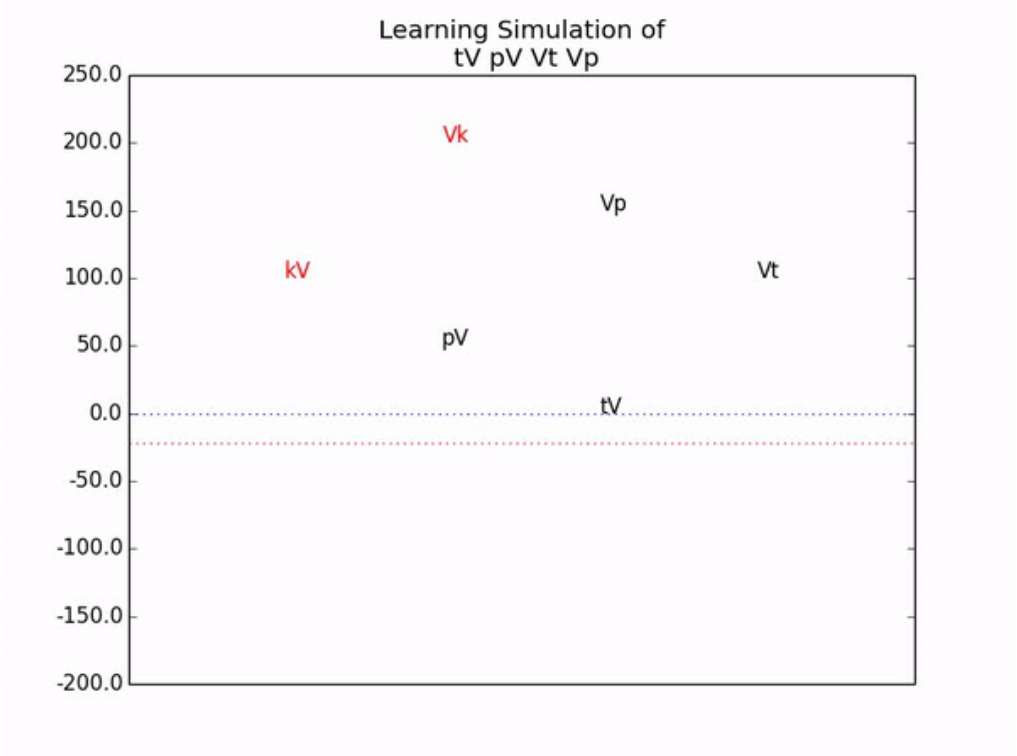




Speed vs. Distance

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





No Substantive Bias

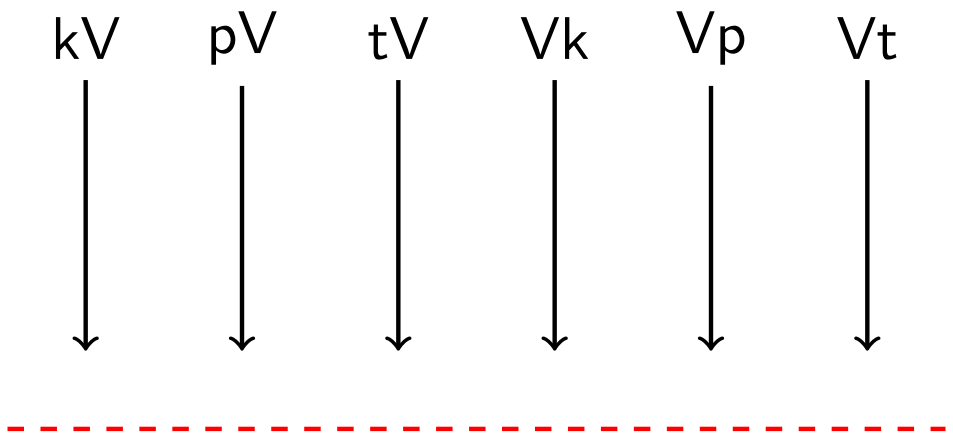
- Neutral constraint set and no misperception[11]
- Stability results confirm structural bias towards No-Dorsals.

Structural	kV **Vk**	pV **Vp**	tV **Vt**	Simple	$\frac{20}{20}$ =	100%
	kV ~~**Vk**~~	pV **Vp**	tV **Vt**	Complex	$\frac{5}{20}$ =	25%
	kV ~~**Vk**~~	pV ~~**Vp**~~	tV ~~**Vt**~~	Simple	$\frac{18}{20}$ =	90%
Anti-Structural	kV ~~**Vk**~~	pV **Vp**	tV **Vt**	Complex	$\frac{5}{20}$ =	25%
	~~kV~~ ~~**Vk**~~	pV **Vp**	tV **Vt**	Simple	$\frac{13}{20}$ =	65%
	~~kV~~ ~~**Vk**~~	pV ~~**Vp**~~	tV ~~**Vt**~~	Complex	$\frac{10}{20}$ =	50%

[11] Simulations run for 20 generations of 2000 iterations at learning rate .05

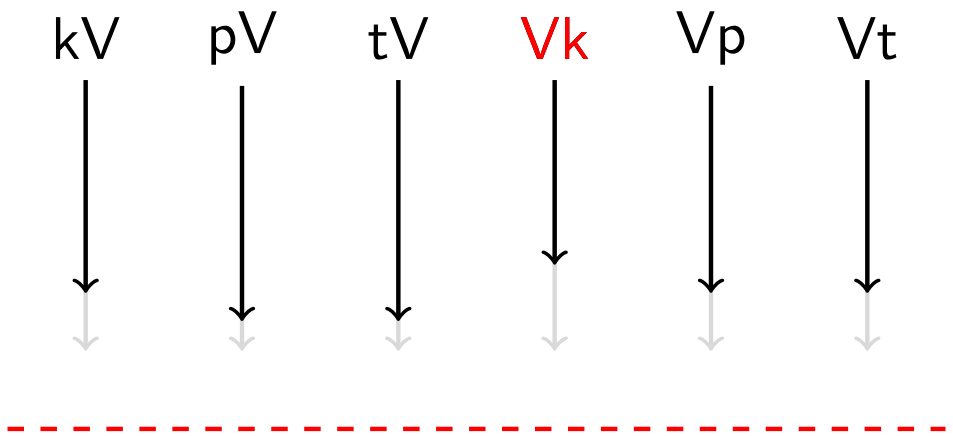
Substance-Free Grammar

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



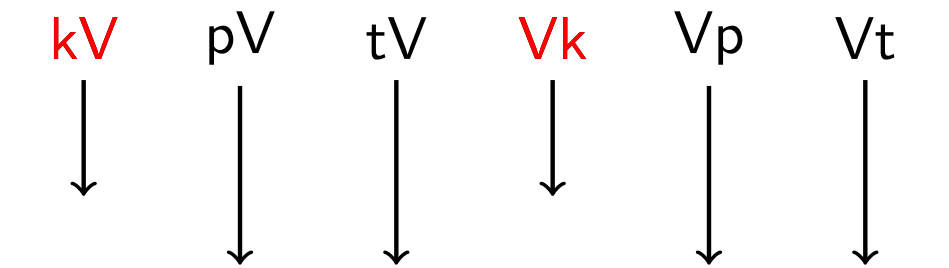
Substance-Free Grammar

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



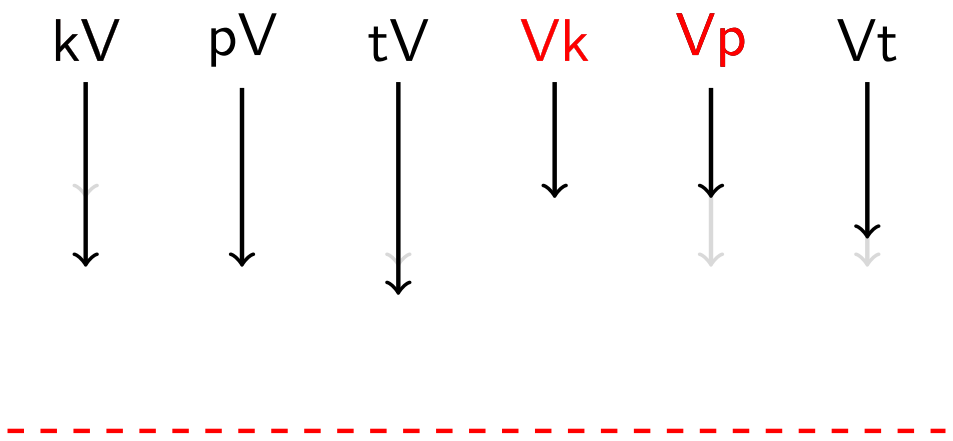
Substance-Free Grammar

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



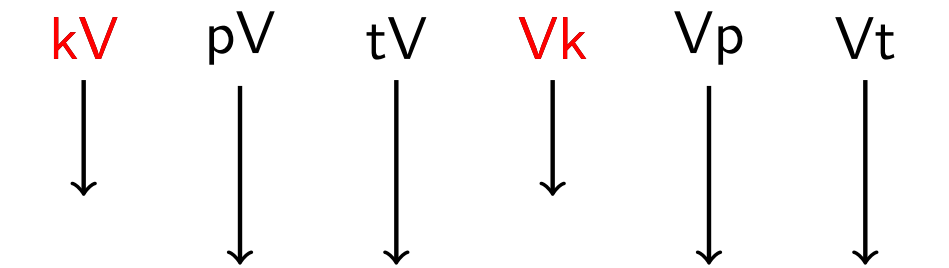
Substance-Free Grammar

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



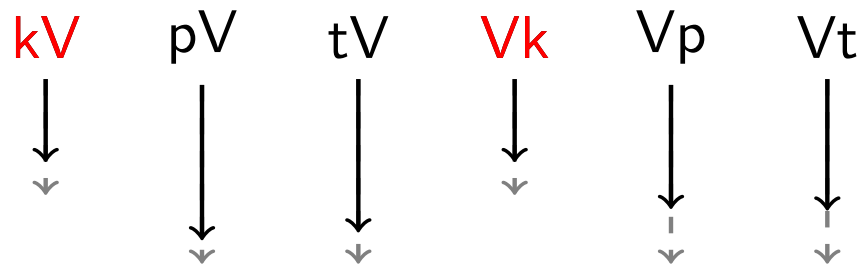
Substance-Free Grammar

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



Substance-Free Grammar

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



Substantively biased channel

- Between the teacher and learners, there is a probability of mistransmission for each sound.

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

Initial	p	t	k	∅
p	.96	.0095	.015	.016
t	.039	.934	010	.016
k	.021	.015	.948	.016
Final	**p**	**t**	**k**	**∅**
p	.697	.099	.135	.070
t	.083	.766	.060	.091
k	.030	.042	.885	.043

Results: Substantively biased channel

- With substantively biased channel, neither the structural, nor anti-structural bias are captured.[12]

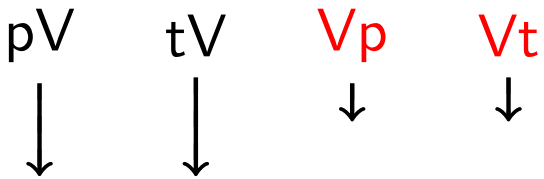
Structural	kV **Vk**	pV **Vp**	tV **Vt**	Simple	$\frac{5}{50}$ =	10%
	kV ~~**Vk**~~	pV **Vp**	tV **Vt**	Complex	$\frac{31}{50}$ =	62%
	kV ~~**Vk**~~	pV ~~**Vp**~~	tV ~~**Vt**~~	Simple	$\frac{37}{50}$ =	74%
Anti-Structural	kV ~~**Vk**~~	pV **Vp**	tV **Vt**	Complex	$\frac{31}{50}$ =	62%
	~~kV~~ ~~**Vk**~~	pV **Vp**	tV **Vt**	Simple	$\frac{31}{50}$ =	62%
	~~kV~~ ~~**Vk**~~	pV ~~**Vp**~~	tV ~~**Vt**~~	Complex	$\frac{0}{50}$ =	0%

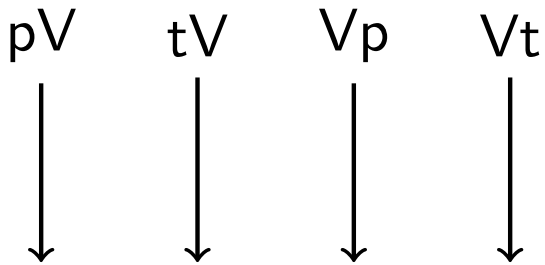
[12] Simulations run for 10 generations of 4000 iterations at learning rate .05

What error rates would be needed?

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

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

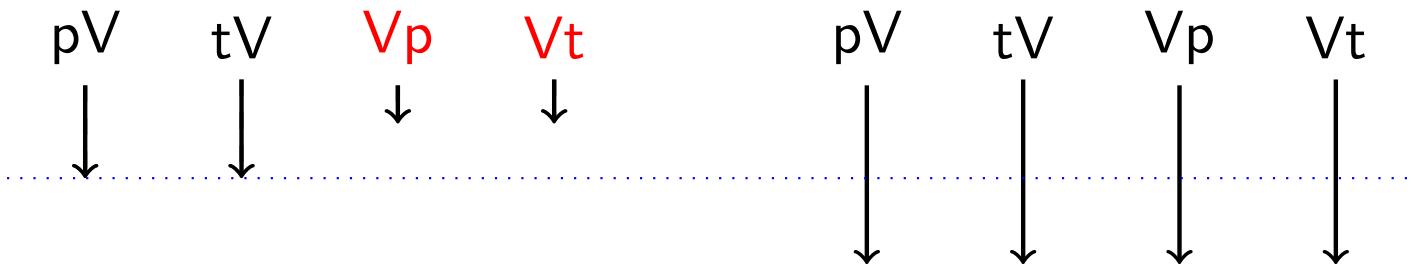




What error rates would be needed?

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

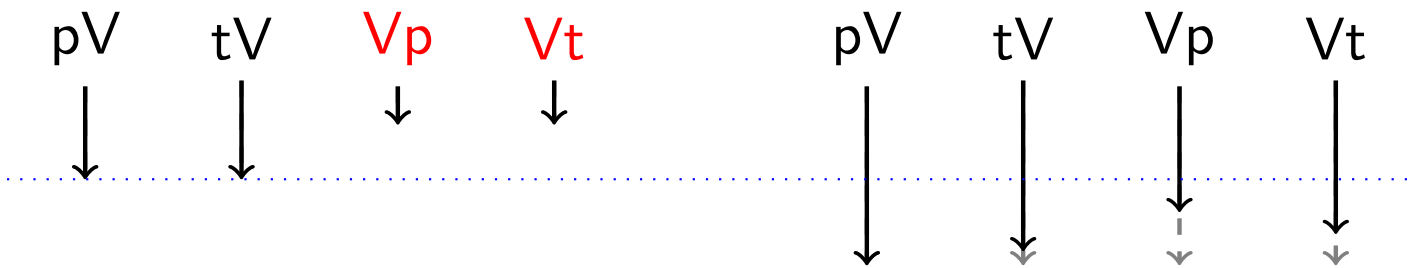
- Two challenges:
  - Capture simplicity bias for syllable-position.
  - Capture anti-structural bias for place of articulation.



What error rates would be needed?

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

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

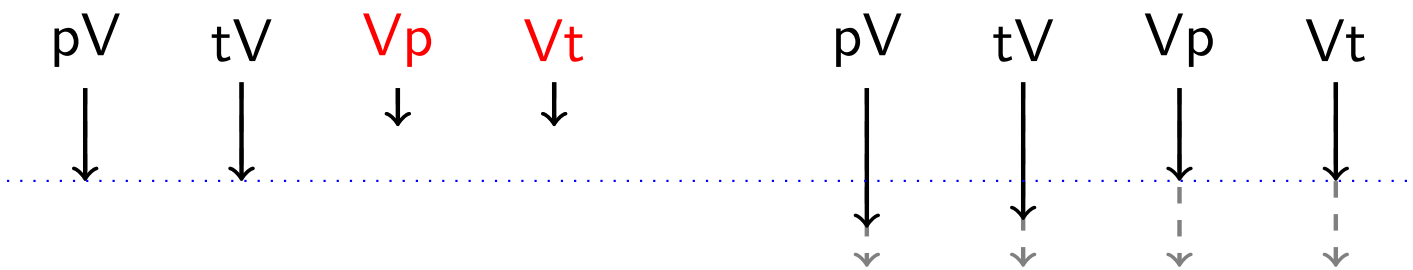


∼Actual Confusion Matrix Data

What error rates would be needed?

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

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

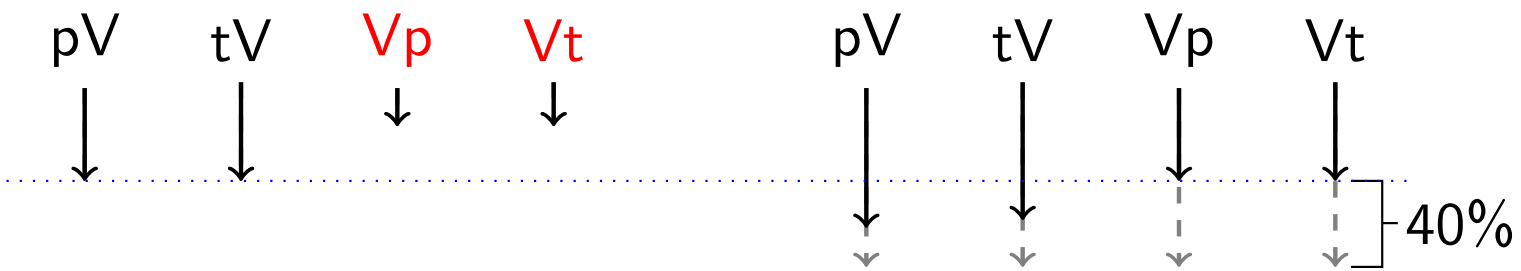


Needed Misperception Rates

What error rates would be needed?

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

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



Needed Misperception Rates

Problems with high rates of misperception

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

Conclusion

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

Conclusion

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

Thank You!

Special thanks go to:

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

Works Cited I

Anceaux, Johannes. 1965. *The Nimboran language; phonology and morphology*. M. Nijhoff.

Blevins, Juliette. 2004. *Evolutionary Phonology: the emergence of sound patterns*. Cambridge University Press.

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.

de Lacy, Paul. 2006. *Markedness: Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press.

Enk, Gerrit J. van, & Vries, Lourens de. 1997. *Korowai of Irian Jaya: Their Language in Its Cultural Context*. Oxford University Press.

Glewwe, Eleanor. 2017. *Substantive Bias in Phonotactic Learning: Positional Extension of Place and Voicing Contrasts*. Paper presented at AMP 2017, NYU.

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.

Haile, Berard. 1926 (1974). *A manual of Navaho grammar*. AMS Press.

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

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

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

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

Works Cited II

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

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.

Kirby, Simon. 2017. Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin and Review*, **24**, 118–137.

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.

Moreton, Elliott, & Pater, Joe. 2012. Structure and substance in artificial-phonology learning Part II, Substance. *Language and Linguistics Compass*, **6**(11), 702–718.

Moreton, Elliott, & Pertsova, Katya. 2014. Pastry phonotactics: Is phonological learning special? *Pages 1–12 of:* Huang, Hsin-Lun, Poole, Ethan, & Rysling, Amanda (eds), *Proceedings of NELS 43*, vol. II.

O'Hara, Charlie. 2017. Soft typology arises from learning bias even with markedness hierarchies. *In: Annual Meeting on Phonology*.

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

O'Hara, Charlie. submitted. *Emergent Learning Bias and the Underattestation of Simple Patterns*. Ms. University of Southern California.

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

Redford, Melissa A., & Diehl, Randy L. 1999. The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *The Journal of the Acoustical Society of America*, **106**, 1555–1565.

Works Cited III

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.

Languages

Word-Final	Number	Languages
[k p t]	43	Aklan, Apinajé, Arara, Athpare, Bajau, Balantak, Barok, Biak, Bobongko, Bullom So, Canela-Kraho, Cebuano, Chontal Mayan, Chrau, Dupaningan Agta, Goemai, Haitian Creole, Hiligaynon, Ik, Ilocano, Kakua, Kapampangan, Kuteb, Lango, Limbu, Maanyan, Mangap-mbula, Mising, Paamese, Sherbro, Sierra Popoluca, Sonora Yaqui, Sukur, Tagalog, Telefol, Tigak, Toba Batak, Tok Pisin, Tondano, Ts'amakko, Waskia, Wayana, West African Pidgin
[p t]	3	Bahasa Indonesia, Itzaj Mayan, Movima
[t]	1	Finnish
[Ø]	27	Adamawa Fulani, Apalai, Barain, Boumaa Fijian, Faranah-Maninka, Japanese Pidgin English, Jalonke, Jamsay, Kalapalo, Kilivila, Kokama-Kokamilla, Koromfe, Kubeo, Mako, Makary Kotoko, Matses, Miya, Nanga, Piraha, Rao, Sabanê, Sandawe, Tibetan, Tiriyo, Tommo So, Warekena, Western Sisaala
[k p]	1	Korowai
[k t]	1	Navaho
[p]	1	Nimboran
[k]	0	
Total	77	

Marked-Faithfulness

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