Markedness bias in the reshaping of Malagasy paradigms

Jennifer Kuo

Cornell University

How do learners deal with conflicting patterns?

PRESENT	PAST
laugh	laughed
dance	danced
jump	jumped
heed	heeded

Rule: add /-d/ to form past tense

....

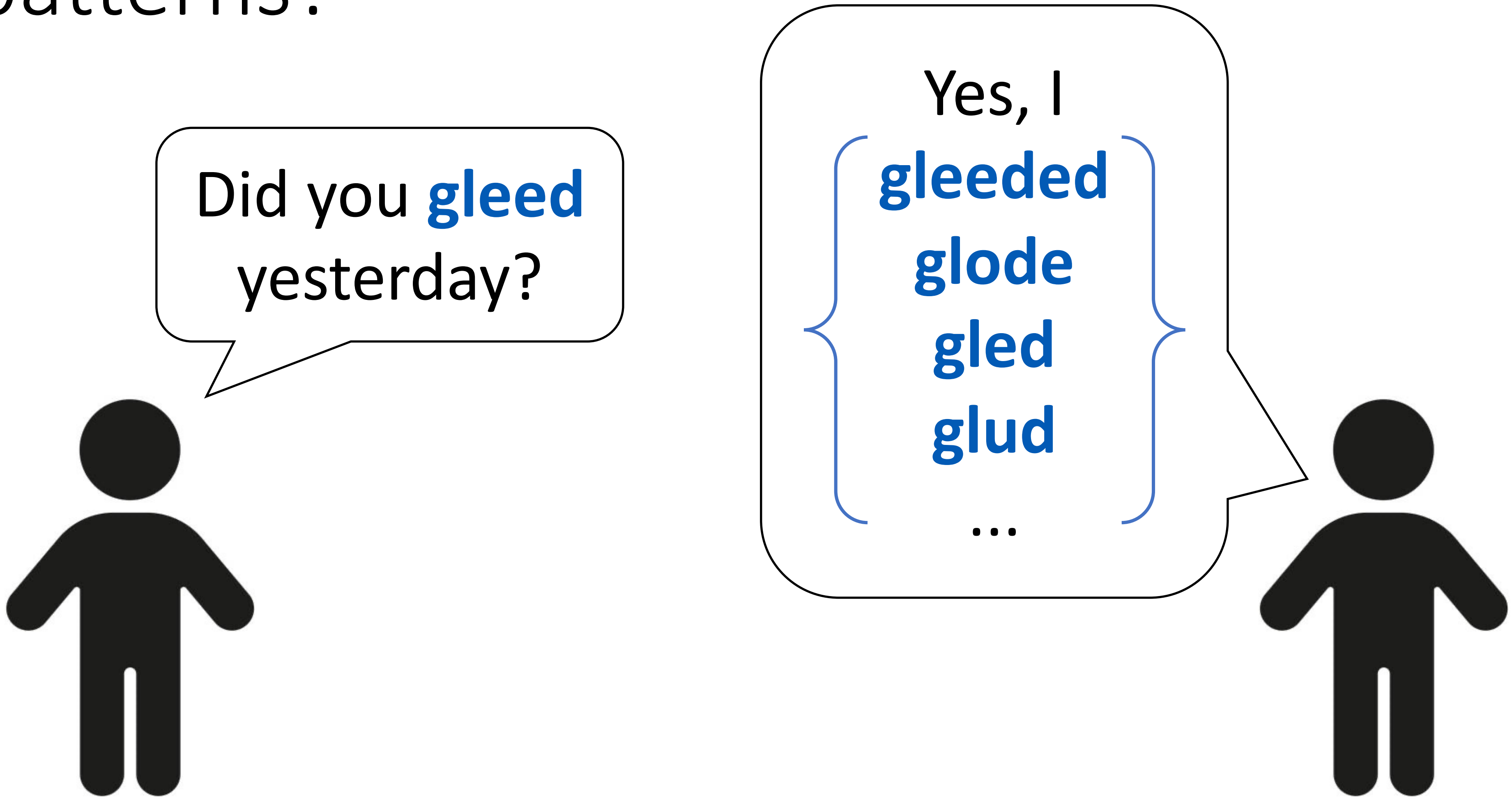
bleed [blid]	bled [blɛd]
read [rid]	read [rɛd]

Rule: in words ending in [id], change [i]→[ɛ] to form past tense.

...

~~heed~~	~~hed [hɛd]~~

How do learners deal with conflicting patterns?



Ambiguity can lead to **reanalysis**

Conflicting data patterns lead to **variance** that is informative.

OLD PATTERN	NEW PATTERN
go, went	go, goed

Ambiguity can lead to **reanalysis**

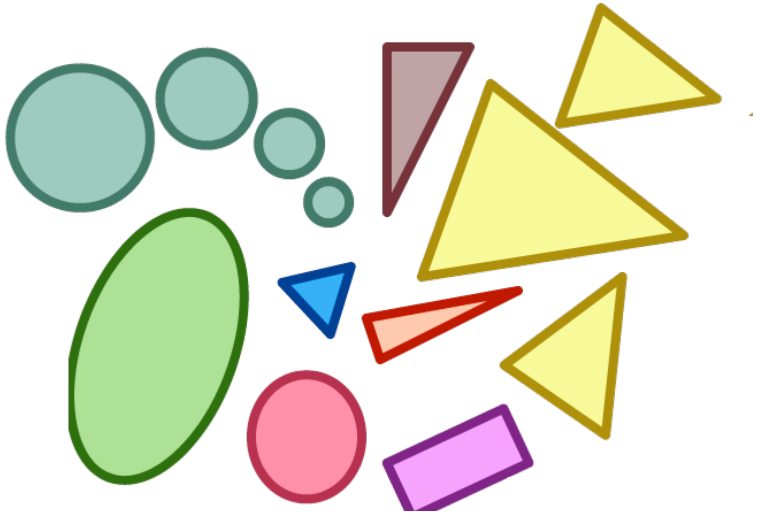
Conflicting data patterns lead to **variance** that is informative

OLD PATTERN		NEW PATTERN
go, went		go, goed
help, halp		**help, helped** (c1300)
dive, dived	→	**dive, dove** (c1800)

Reanalysis: Innovative variants are adopted and passed down.

Phonological learning: competing views

- Frequency-matching (“statistical learning”)
  - Domain-general (Gallistel 1990; Saffran et al. 1999; Newport et al. 2004; Turk et al. 2015)
  - Experiments (e.g. Ernestus & Baayen, 2003; Albright & Hayes, 2003)
  - Acquisition (e.g. Maye, Werker, & Gerken 2002; Romberg & Saffran, 2010)



Turk et al. 2015

Gallistel 1990

Phonological learning: competing views

- Frequency-matching (“statistical learning”)
  - Domain-general (Gallistel 1990; Saffran et al. 1999; Newport et al. 2004; Turk et al. 2015)
  - Experiments (e.g. Ernestus & Baayen 2003; Albright & Hayes 2003)
  - Acquisition (e.g. Maye, Werker, & Gerken 2003; Romberg & Saffran 2010)

- Linguistically-motivated biases towards:
  - **simpler patterns** (complexity bias; Moreton & Pater 2012a)
  - **smaller changes** (perceptual similarity bias; Steriade 2001; Wilson 2006; White 2017)
  - patterns that are **easier to produce/perceive** (markedness bias; Jarosz 2006)

Factors driving reanalysis

- Existing models are frequency-matching

gleed → gleeded		**Rule:** add "ed" to form past tense N=1146/1234 (93%)
	gled	**Rule:** if a word ends in [id], change [i]→[ɛ] N=6/7 (86%)
	glode	**Rule:** if a word ends in [iC], change [i]→[o] N=6/184 (3.3%)

In this case (and many), frequency-matching makes the right predictions!

Generalizations from Albright & Hayes (2003), using data from CELEX database (Baayen et al. 1995)

Factors driving reanalysis

- Problem: Malagasy reanalysis is **not** entirely predictable from statistical learning.
- Proposal: reanalysis is sensitive to **markedness bias**
  - “Marked”= cross-linguistically dispreferred because of being harder to produce/perceive

Markedness effects in reanalysis

What type of markedness effects are possible in reanalysis?

“Universal”

All possible markedness effects decided by UG

“Active”

Markedness effects already active in the language as **stem phonotactics**

Markedness effects in reanalysis

What type of markedness effects are possible?

“Universal”

All possible markedness effects decided by UG

“Active”

Markedness effects already active in the language as **stem phonotactics**

Markedness effects in reanalysis

- Restricting bias to “active” markedness predicts a strong connection between phonotactics and morphophonology

  English ex: *[ʃsak] in roots
  *[dɪʃ-s] ‘dish + PL’ in morphophonology

- Support from:
  - Acquisition: phonotactics before alternation learning (Jarosz 2006; Tesar & Prince 2007)
  - Experiments (Pater & Tessier 2005; Chong 2021)

How does phonological learning work?

Frequency-matching

+

Markedness bias

Tools: historical change (reanalysis) + quantitative modeling

Goals of the talk

1. Show that reanalysis in Malagasy can be explained as **frequency matching** + **markedness bias**
   - Where markedness is restricted to effects already active in phonotactics.
2. Outline a model for incorporating markedness effects into reanalysis.
3. Demonstrate how quantitative models can be used to test theories about language learning.

Case study: Malagasy consonant alternations

1	2	3	4
Intro	Malagasy reanalysis	Model	Results

Case study: Malagasy final consonants

- Malagasy: language spoken in Madagascar
- Malayo-Polynesian
- Dialect: Official Malagasy, based on variant spoken in/around the capital city Antananarivo.

Weak stems (Albro 2005; Keenan and Polinsky 2017)

- always end in **‘ʈʂa’**, **‘ka’,** or **‘na’**
- When suffixed, the consonant in the weak syllable (ʈʂ/k/n) may alternate with another consonant

type	**alternant**	**unsuffixed**	**suffixed (+ana)**	
na	n	a^{n}ɖʐávi**na**	a^{n}ɖʐaví**n**-ana	‘to bear leaves’
	m	aná^{n}ɖʐa**na**	a^{n}ɖʐá**m**-ana	‘to try’
ka	h	a^{n}gáta**k**a	a^{n}gatá**h**-ana	‘to ask for’
	f	anáha**k**a	anahá**f**-ana	‘to scatter’
ʈʂa	r	iána**ʈʂ**a	ianá**r**-ana	‘to learn’
	t	aná^{n}ɖʐa**ʈʂ**a	ana^{n}ɖʐá**t**-ana	‘to promote’
	f	a^{n}ɖʐáku**ʈʂ**a	a^{n}ɖʐakú**f**ana	‘to cover’

How did weak stems happen?

Generalizations taken from Dahl (1951, 1988), Mahdi (1988), Adelaar (2012)

Note: Data is simplified for ease of presentation, and does not accurately reflect all historical changes

Proto-Malayo-Polynesian ↓ Malagasy

UNSUFFIXED	SUFFIXED	PROCESS
avut	avut-an	
avu**ʈʂ**	avut-an	t,r→ʈʂ at the end of words
avuʈʂ**a**	avut-an**a**	insert vowel after final C

~600-700AD

Changes induced by migration to Madagascar.

Reanalysis in weak stems

1	2	3	4
Intro	**Malagasy reanalysis**	Model	Results

Reanalysis in weak stems

- Ambiguity in the unsuffixed form → reanalyses

vuhiʈʂa → vuhi**t**-ana?

vuhiʈʂa → vuhi**r**-ana?

Reanalysis: change to a sound pattern over generations of speakers

- Possible reanalyses for [vuhiʈʂa]:

DIRECTION	SUFFIXED (+ana)
t → r	vuhit-ana → vuhir-ana
r → t	vuhir-ana → vuhit-ana

Reanalysis in weak stems

- As a preview, reanalysis appears to have largely happened in the following directions:

Type	Direction	Predicted by statistical learning?
ka	f→h	Yes
na	m→n	Yes
ţşa	**t → r**	**No**

→Note: I will largely focus on reanalysis in ţşa-final words.

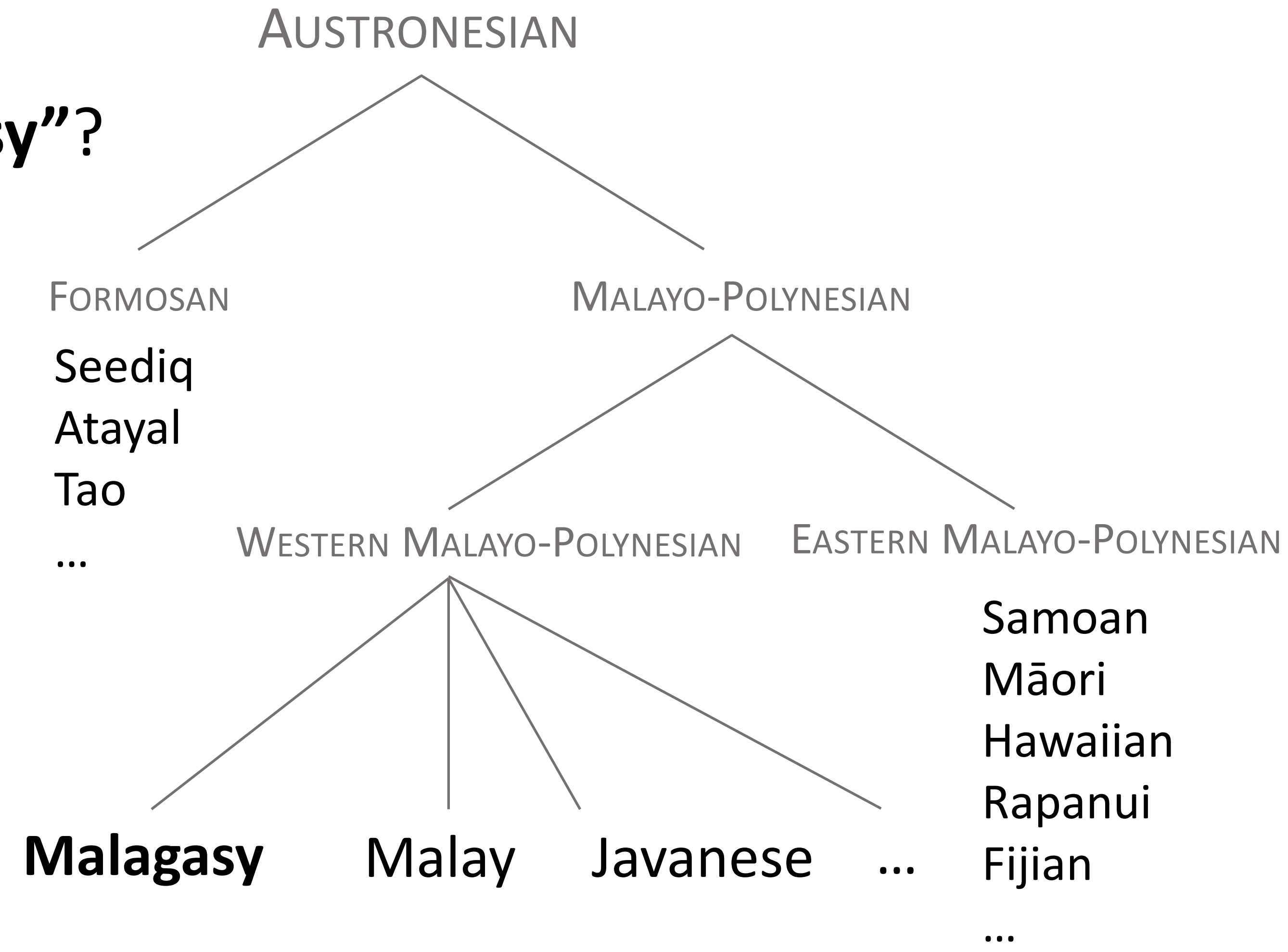
Methodology and data

- Compare data from “old” Malagasy (pre-reanalysis) to “new” Malagasy (post-reanalysis)

Methodology and data

- What is **“old Malagasy”**?

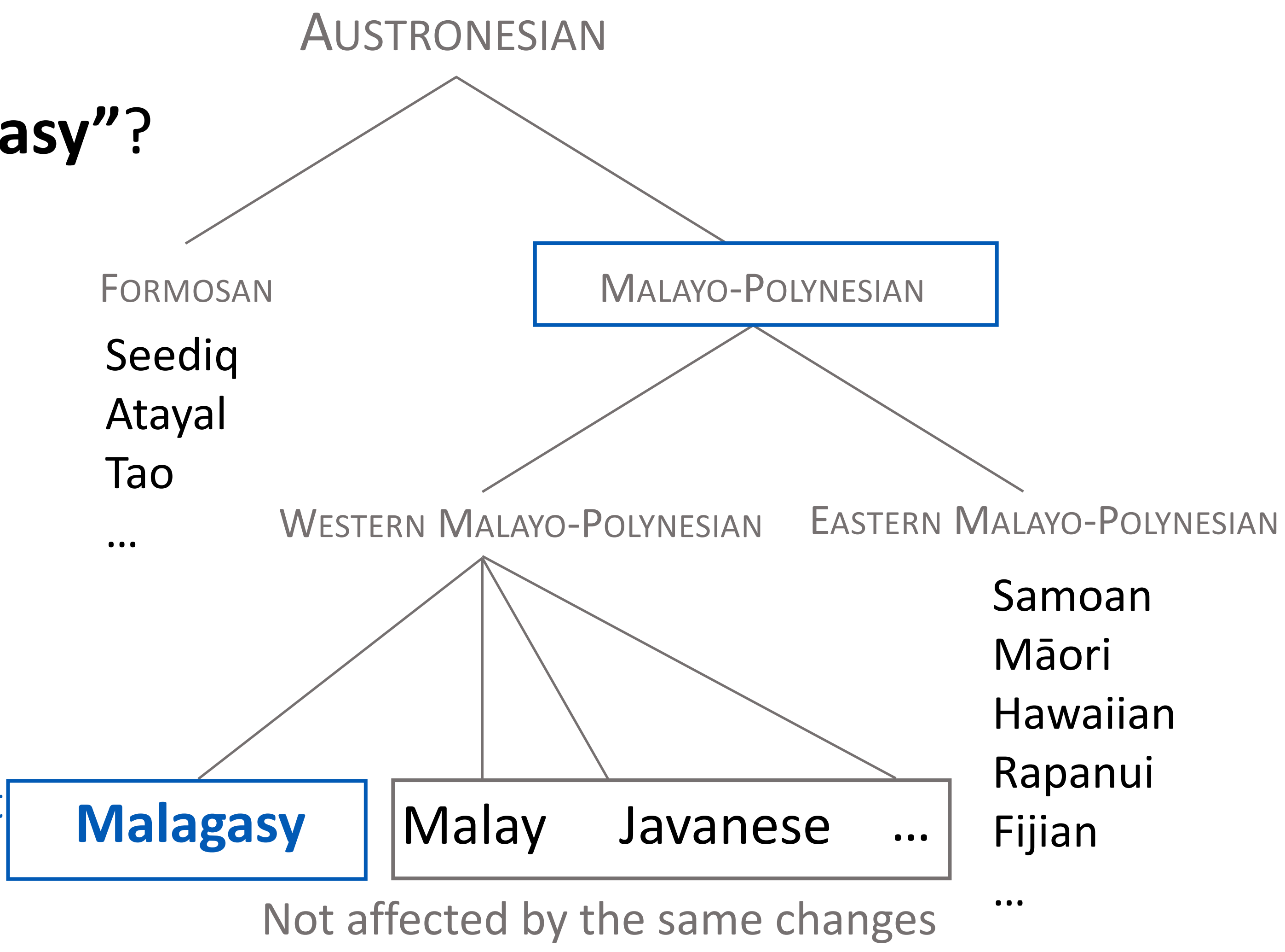
  Dahl (1951, 1988)
  Blust (1984)
  Mahdi (1988)
  Adelaar (2013)
  etc.



Methodology and data

- What is **“old Malagasy”**?

  Proto-Malayo-Polynesian **(PMP)**



Reanalyses
(due to development of weak stems)

Methodology and data

PMP (before reanalysis)	**Malagasy** (after reanalysis)
approx. 7th century AD	approx. 1880–present
n=215	n=1893
• Austronesian Comparative Dictionary (Blust & Trussel 2010) • World Loanword Database (Adelaar 2009)	• Malagasy Dictionary and Encyclopedia of Madagascar (MDEM; de La Beaujardière 2004) • 108,000 words/phrases, filtered with help of a script. • Native speaker consultant

Expected direction of reanalysis in ʈʂa words

Tables: ʈʂa-stem alternants in **PMP** (before reanalysis)

(a) all words

Consonant	n	%
(ʈʂ~) r	17	26.6%
(ʈʂ~) **t**	**47**	**73.4%**

- Assuming statistical learning, we predict:
  - Reanalysis of r→t

Expected direction of reanalysis in ʈʂa words

Tables: ʈʂa-stem alternants in **PMP** (before reanalysis)

(a) all words

Consonant	n	%
(ʈʂ~) r	17	26.6%
(ʈʂ~) **t**	**47**	**73.4%**

(b) words with a preceding [r]

Consonant	n	%	
(ʈʂ~) r	0	0	puriʈʂa~purir-ana
(ʈʂ~) **t**	**8**	**100%**	puriʈʂa~purit-ana

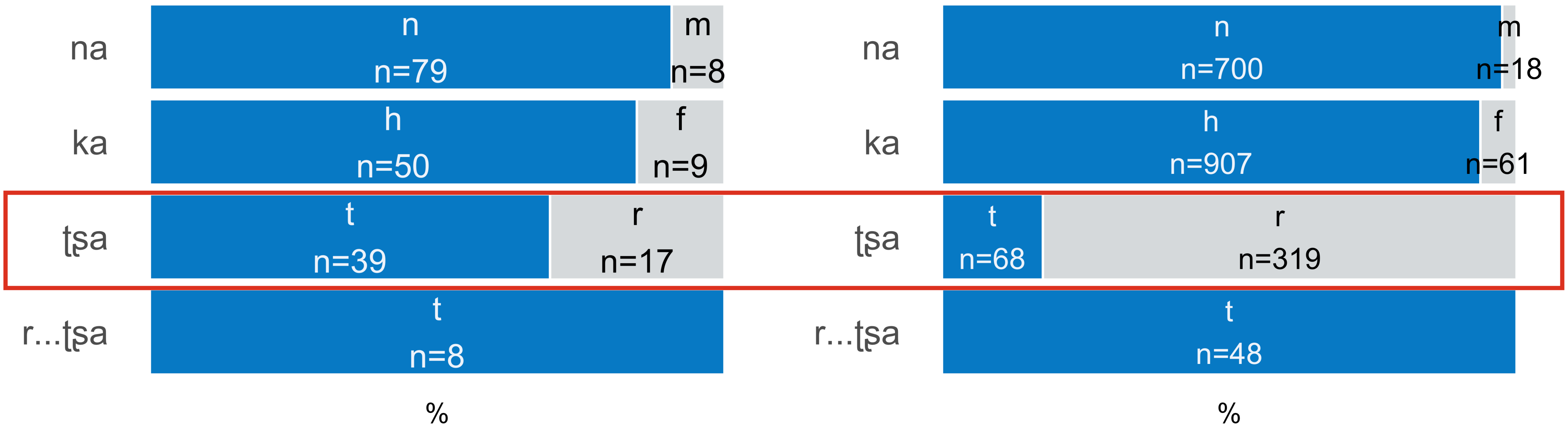
- Assuming statistical learning, we predict:
  - Reanalysis of r→t
  - **r-dissimilation:** alternant should **not** be [r] if the word already has an [r]

Observed directions of reanalysis

- Indirect evidence of reanalysis: comparing PMP vs. Malagasy

Distribution of alternants:

(a) PMP (before reanalysis) (b) Malagasy (after reanalysis)



Observed directions of reanalysis

- Direct evidence: words that have undergone reanalysis

PMP→Mlg	count	Example
r → r	18 (95%)	velaʈʂa ~ velar-ana →velar-ana `to spread out'
r → t	1 (5%)	sa^{n}ɖʐaʈʂa ~ sa^{n}ɖʐar-ana →sa^{n}ɖʐat-ana `to rise up'
t → t	23 (43%)	uruʈʂa ~ urut-ana→ urut-ana `to massage'
t → r	30 (57%)	akaʈʂa ~ akat-ana → akar-ana `to raise'

Observed directions of reanalysis

- Direct evidence: words that have undergone reanalysis

PMP→Mlg	**count**	**Example**
r → **r**	18 (95%)	velaʈʂa ~ vela**r**-ana →vela**r**-ana \`to spread out’
r → **t**	1 (5%)	saⁿɖʐaʈʂa ~ saⁿɖʐa**r**-ana →saⁿɖʐa**t**-ana \`to rise up’
t → **t**	23 (43%)	uruʈʂa ~ uru**t**-ana→ uru**t**-ana \`to massage’
t → **r**	30 (57%)	akaʈʂa ~ aka**t**-ana → aka**r**-ana \`to raise’

- Overwhelmingly, reanalysis is in the direction **t** → **r**

Observed directions of reanalysis

- Direct evidence: words that have undergone reanalysis

old→new	count	% preceding r
r → r	18 (95%)	0%
r → t	1 (5%)	100%
t → t	23 (43%)	61%
t → r	30 (57%)	0%

- Overwhelmingly, reanalysis is in the direction **t → r**
  - Except when the word already has a preceding [r]

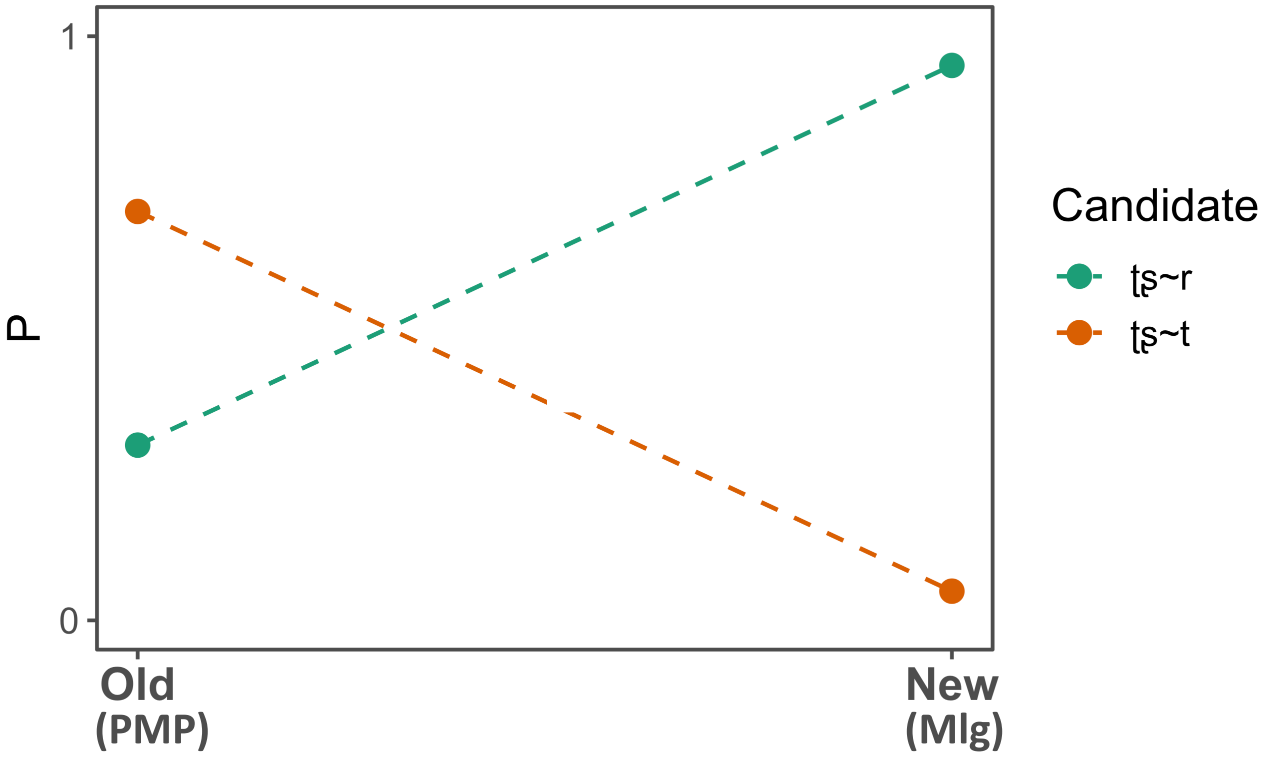
Summary of pattern

	PMP	**Malagasy**
ka words	prefer [h]	prefer [h]
na words	prefer [n]	prefer [n]
ţşa words	prefer [t] avoid r…r	**prefer [r]** avoid r…r

Graphing the pattern: ʈʂa words (no preceding [r])

input	output	PMP	Mlg
vukiʈʂa	vuki<u>r</u>ana	0.3	0.95
	vuki<u>t</u>ana	0.7	0.05

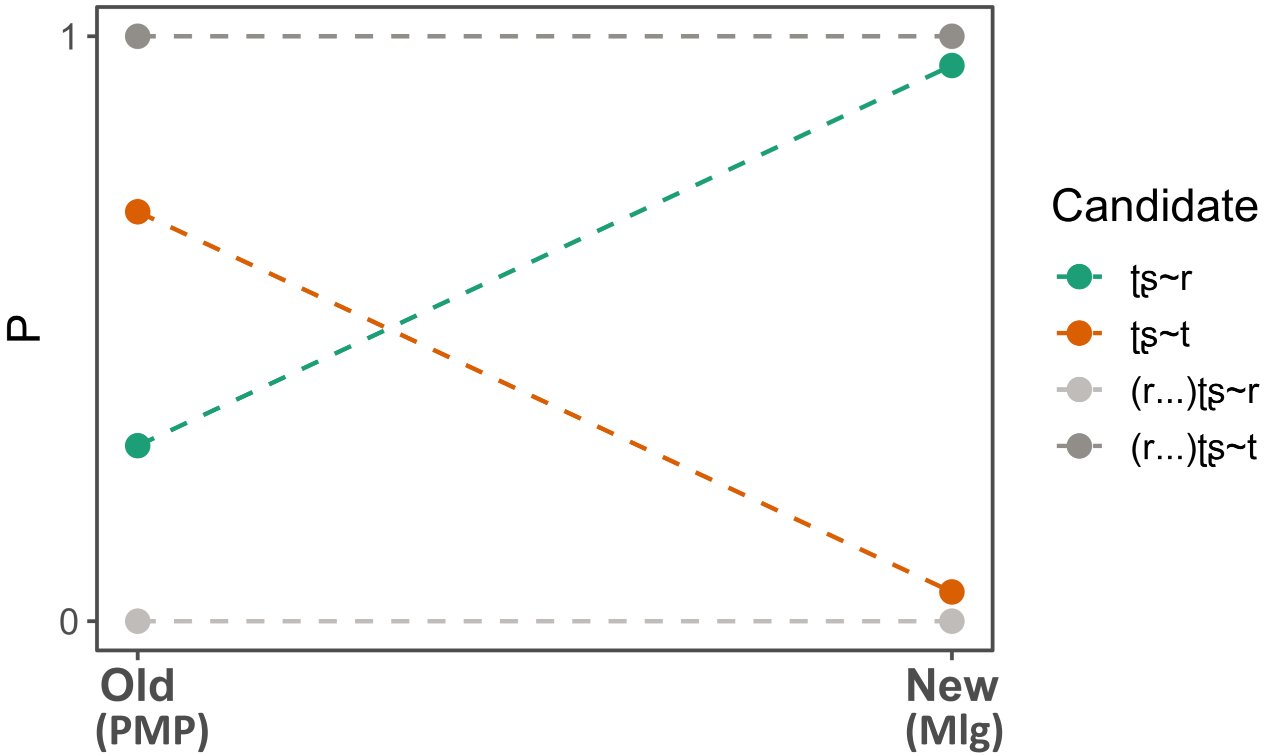
Figure: Proportion of consonant (t vs. r) that surfaces under suffixation:



Graphing the pattern: ʈʂa words (all)

input	output	PMP	Mlg
vukiʈʂa	vukirana	0.3	0.95
	vukitana	0.7	0.05
vuriʈʂa	vurirana	0	0
	vuritana	1	1

Figure: Proportion of consonant (t vs. r) that surfaces under suffixation:



Frequency-matching vs. markedness bias

- Reanalysis is not predictable from statistical distributions within the paradigm
- Proposal: Reanalysis is **markedness-reducing**

Stops between vowels are marked

- *VTV: disprefer (voiceless) **stops** between vowels (/p t k ʈʂ/)
  - Bad: atu, faike, papi, beʈʂuka…vuhit-ana
  - Good: aro, azi, lumu, tafi, etc… vuhir-ana
- harder to produce (Kirchner, 1998; Kaplan, 2010; Katz, 2016)
- cross-linguistic support
  - English ex: tapping! vo[t]e→ vo[ɾ]ing "vote/voting" (Hayes 2011, 143-144)

*VTV in phonotactics

- **Active markedness proposal**: markedness effects are restricted to those already active in the phonotactics.
- Is this true for Malagasy?

*VTV in phonotactics

- **Active markedness proposal**: markedness effects are restricted to those already active in the phonotactics.
- Is this true for Malagasy? **Yes**
  - A probabilistic phonotactic model trained on stems assigns *VTV significant weight (UCLA Phonotactic Learner; Hayes & Wilson 2008)
  - A series of sound changes removed most intervocalic stops for a period of Malagasy (Adelaar, 2012)

Modeling reanalysis

1	2	3	4
Intro	Malagasy reanalysis	**Model**	Results

Elements in a model of reanalysis

1. A probabilistic phonological grammar
2. Ability to incorporate learning biases
3. Simulate generations of change

Elements in a model of reanalysis

1. **A probabilistic phonological grammar**
2. Ability to incorporate learning biases
3. Simulate generations of change

1 Phonological grammar

- Maximum Entropy Harmonic Grammar (e.g., Smolensky 1986; Goldwater & Johnson, 2003)
  - Extension of Optimality Theory (Prince & Smolensky 1993/2004)

1 Phonological grammar

/dɪʃ-z/ 'dish-PL'	*ʃs	IDENT	DEP
	w=3	w=2	w=0.5
a. [dɪʃs]	1		
b. [dʌʃ]		1	
c. [dɪʃəz]			1

weighted constraints

1 Phonological grammar

/dɪʃ-z/ 'dish-PL'	*ʃs	IDENT	DEP	H
	w=3	w=2	w=0.5	
a. [dɪʃs]	1			$1\times3 + 0\times2 + 0\times0.5 = 3$
b. [dʌʃ]		1		$0\times3 + 1\times2 + 0\times0.5 = 2$
c. [dɪʃəz]			1	$0\times2 + 0\times2 + 1\times0.5 = 0.5$

Each candidate receives a penalty score that is the **weighted sum** of all its constraint violations.

1 Phonological grammar

/dɪʃ-z/ (x_i) 'dish-PL'	*ʃs w=3	IDENT w=2	DEP w=0.5	H	$P(y_j\|x_i)$
a. [dɪʃs] (y_1)	1			$\mathbf{1\times3} + 0\times2 + 0\times0.5 = 3$	0.06
b. [dʌʃ] (y_2)		1		$0\times3 + \mathbf{1\times2} + 0\times0.5 = 2$	0.17
c. [dɪʃəz] (y_3)			1	$0\times2 + 0\times2 + \mathbf{1\times0.5} = 0.5$	0.70

Convert to probabilities
probability of candidate y_j given input x_i

1 Phonological grammar

/dɪʃ-z/ (x_i) 'dish-PL'	*ʃs w=3	IDENT w=2	DEP w=0.5	H	$P(y_j \mid x_i)$
a. [dɪʃs] (y_1)	1			$1\times3 + 0\times2 + 0\times0.5 = 3$	0.06
b. [dʌʃ] (y_2)		1		$0\times3 + \mathbf{1\times2} + 0\times0.5 = 2$	0.17
c. [dɪʃəz] (y_3)			1	$0\times2 + 0\times2 + \mathbf{1\times0.5} = 0.5$	0.70

Takeaway: If a candidate output violates a highly weighted constraint, it will receive low probability.

Convert to probabilities
probability of candidate *y* given input x_i

1 Phonological grammar

/dɪʃ-z/ (x_i) 'dish-PL'	*ʃs w=?	IDENT w=?	DEP w=?	H	$P(y\|x_i)$
a. [dɪʃs] (y_1)	1			-	$p(y_1\|x_i)$
b. [dʌʃ] (y_2)		1		-	$p(y_2\|x_i)$
c. [dɪʃəz] (y_3)			1	-	$p(y_3\|x_i)$

$$\log(p(y_1|x_i)p(y_2|x_i)\ldots p(y_n|x_i)) = \sum_{n=1}^{N} \log(P(y_n|x_i))$$

How are weights learned? by **maximizing objective function** using gradient-based optimization (Goldwater & Johnson, 2003; Lafferty et al., 2001; McCallum, 2003)

1 Phonological grammar

/dɪʃ-z/ (x_i) 'dish-PL'	*ʃs	IDENT	DEP	H	$P(y\|x_i)$
	w=?	w=?	w=?		
a. [dɪʃs] (y_1)	1			-	$p(y_1\|x_i)$
b. [dʌʃ] (y_2)		1		-	$p(y_2\|x_i)$
c. [dɪʃəz] (y_3)			1	-	$p(y_3\|x_i)$

$$\log(p(y_1|x_i)p(y_2|x_i)\ldots p(y_n|x_i))$$
$$= \sum_{n=1}^{N} \log(P(y_n|x_i))$$

The resulting model is **frequency-matching**.

1 Phonological grammar

Now let’s apply this to Malagasy!

1 Phonological grammar: constraints

***VTV**	**Assess violation for voiceless stops/affricates between vowels (/p, t, k, ʈʂ/)** *[+syllabic][-continuant,-voice][+syllabic]
*ʈʂ]V, *k]V, *n]V	Assess violations for every C]V, where C is at a *morpheme boundary* (Pater 2007; Chong 2020) • within stems, prevocalic ʈʂ, k, and n are allowed • (e.g. beʈʂuka ‘to swell up’)
*r…r	Assess violation for sequences of r…r
*IDENT-IO[F]	The specification for [F] in an input segment must be preserved in its output correspondent (McCarthy & Prince 1995)

A Malagasy example (simplified)

pakuʈʂ+ana	*ʈʂ]V	***VTV**	**IDENT [VC]**		
	w=10	**w=4**	**w=3**	**H**	$P(y\|x_i)$
a. paku**ʈʂ**-ana	1			10	0
b. paku**t**-ana		1		4	0.3
c. paku**r**-ana			1	3	**0.7**

If w(*VTV) > w(IDENT[voice]), the grammar will prefer [pakur-ana]

If w(IDENT[voice]) > w(*VTV), the grammar will prefer [paku**t**-ana]

A Malagasy example (simplified)

pakuʈʂ+ana	*ʈʂ]V w=10	***VTV** **w=3**	**IDENT[vc]** **w=4**	**H**	$P(y\|x_i)$
a. paku**ʈʂ**-ana	1			10	0
b. paku**t**-ana		1		3	**0.7**
c. paku**r**-ana			1	4	0.3

If w(*VTV) > w(IDENT[voice]), the grammar will prefer [pakur-ana]

If w(IDENT[voice]) > w(*VTV), the grammar will prefer [pakut-ana]

Elements in a model of reanalysis

1. A probabilistic phonological grammar
2. **Ability to incorporate learning biases**
3. Simulate generations of change

2 Learning biases

To implement a bias, we can give the model a **Gaussian prior** (Chen & Rosenfield 1999; Wilson 2006; White 2013)

- Functionally equivalent to L2 regularization

Each constraint weight *w* is associated with a Gaussian distribution, defined in terms of a **mean** μ and a **standard deviation** σ.

$$\frac{(w_m - \mu_m)^2}{2\sigma^2}$$

2 Learning biases

/dɪʃ-s/ (x_i)	*ʃs	IDENT	DEP	
	w_1	w_2	w_3	$\boldsymbol{P(y\|x_i)}$
a. [dɪʃs] (y_1)	1			$p(y_1\|x_i)$
b. [dʌʃ] (y_2)		1		$p(y_2\|x_i)$
c. [dɪʃəs] (y_3)			1	$p(y_3\|x_i)$

Old objective function

L = $\sum_{n=1}^{N} \log\left(P(y_n|x_i)\right)$

2 Learning biases

/dɪʃ-s/ (x_i)	*ʃs	IDENT	DEP	
	w$_1$	w$_2$	w$_3$	
	$\boldsymbol{\mu_1}$ $\frac{(w_1-\mu_1)^2}{2\sigma^2}$	$\boldsymbol{\mu_2}$ $\frac{(w_2-\mu_2)^2}{2\sigma^2}$	$\boldsymbol{\mu_3}$ $\frac{(w_3-\mu_3)^2}{2\sigma^2}$	$\boldsymbol{P(y\vert x_i)}$
a. [dɪʃs] (y_1)	1			$p(y_1\vert x_i)$
b. [dʌʃ] (y_2)		1		$p(y_2\vert x_i)$
c. [dɪʃəs] (y_3)			1	$p(y_3\vert x_i)$

Implementing a
Gaussian prior

New objective function

L = $\sum_{n=1}^{N} \log\big(P(y_n|x_i)\big) - \boxed{\sum_{m=1}^{M} \frac{(w_m-\mu_m)^2}{2\sigma^2}}$

The bigger this value, the bigger the penalty.

2 Learning biases

- Bias can be injected into the model by varying μ for each constraint.
  - high μ = high preferred weight
  - low μ = low preferred weight

$$\frac{(\boldsymbol{w_m} - \boldsymbol{\mu_m})^{\mathbf{2}}}{2\sigma^2}$$

- σ set to 1.0 for all constraints

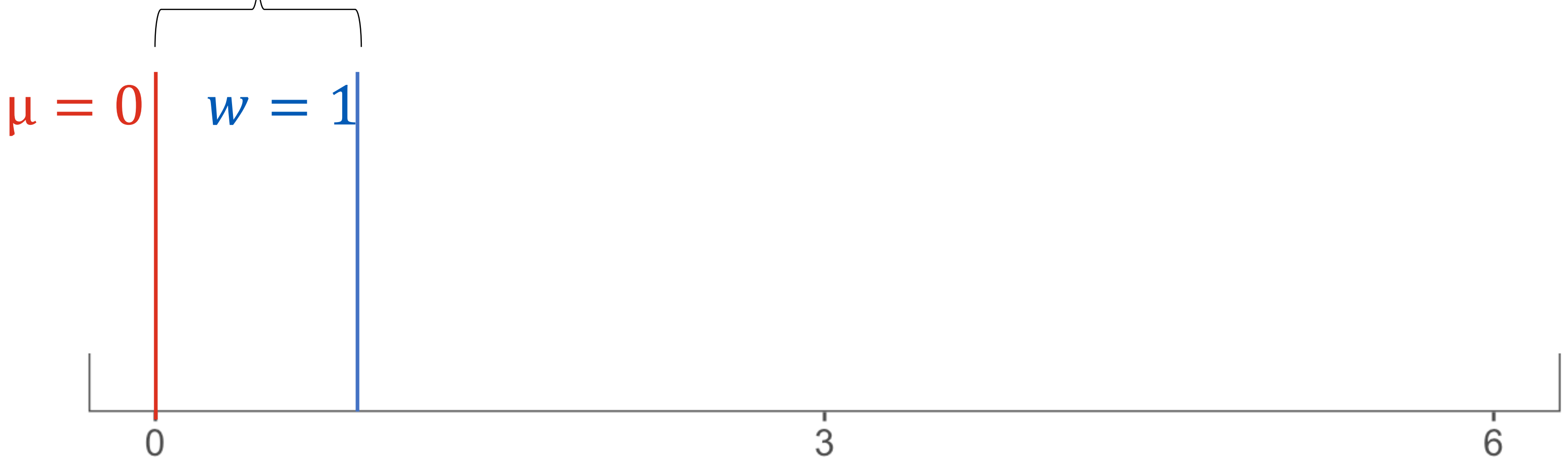
Low μ = low preferred weight

$$\text{Prior} = \frac{(w_m - \mu_m)^2}{2\sigma^2}$$

w = constraint weight

μ = “preferred” weight

$(1 - 0)^2 = 1$

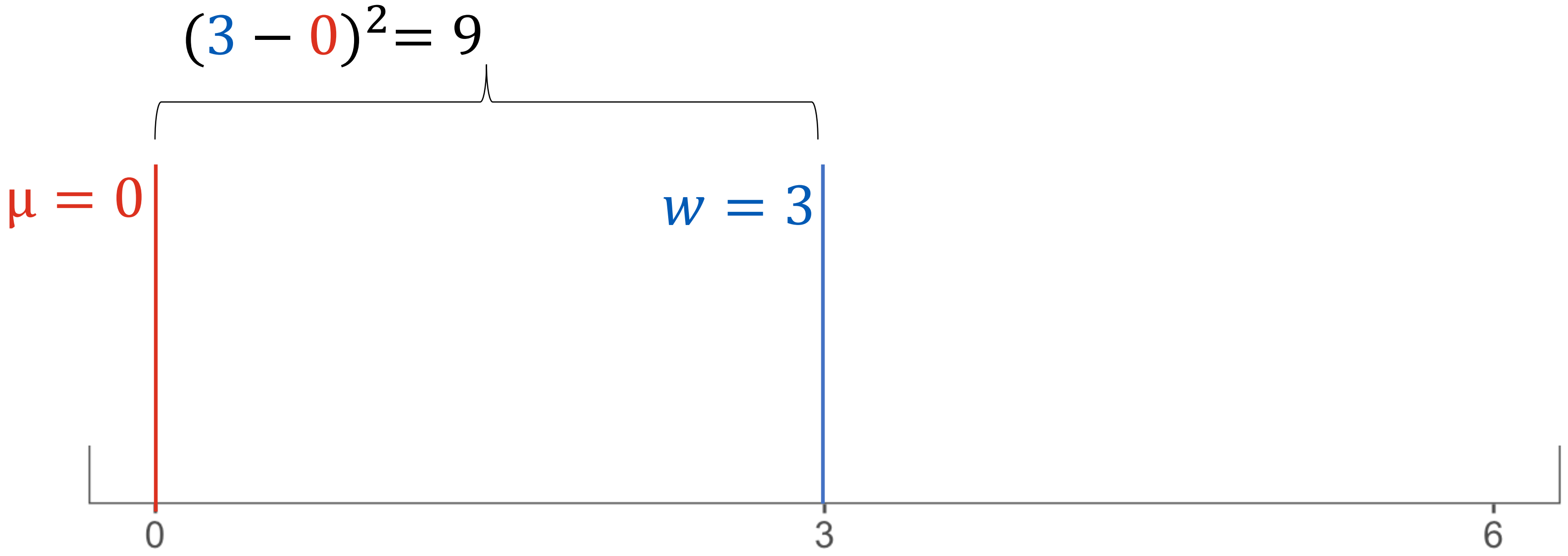


Low μ = low preferred weight

$$\text{Prior} = \frac{(w_m - \mu_m)^2}{2\sigma^2}$$

w = constraint weight

μ = “preferred” weight



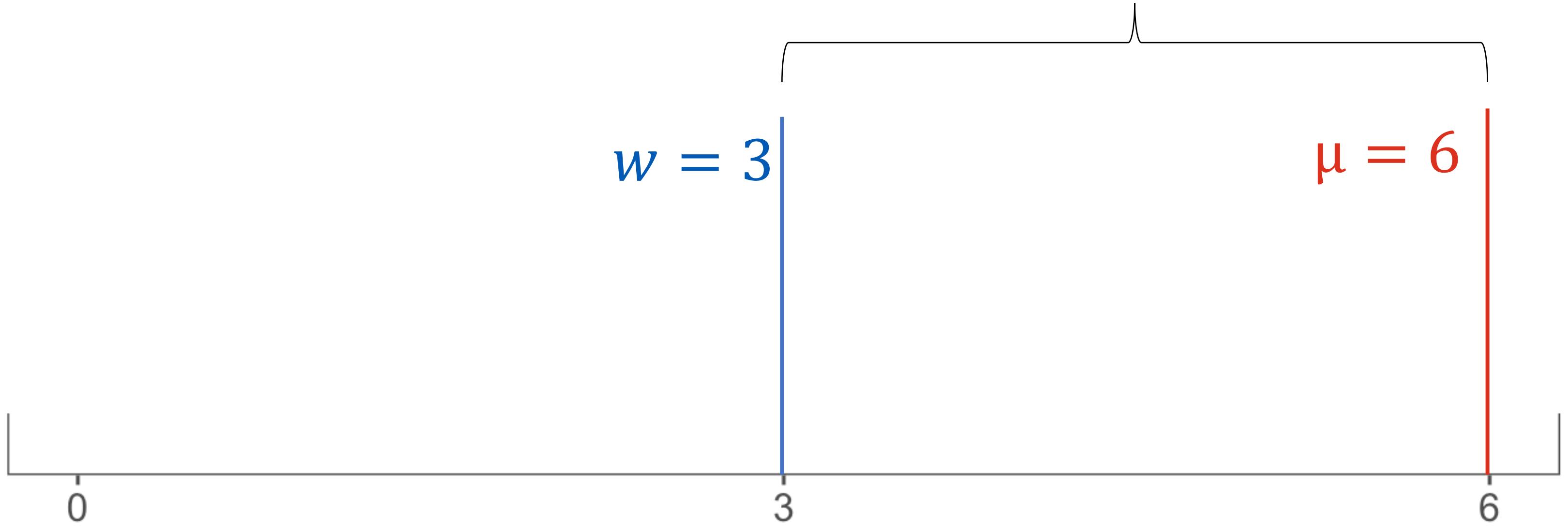
high μ = high preferred weight

$$\text{Prior} = \frac{(w_m - \mu_m)^2}{2\sigma^2}$$

w = constraint weight

μ = “preferred” weight

$$(3-6)^2 = 9$$



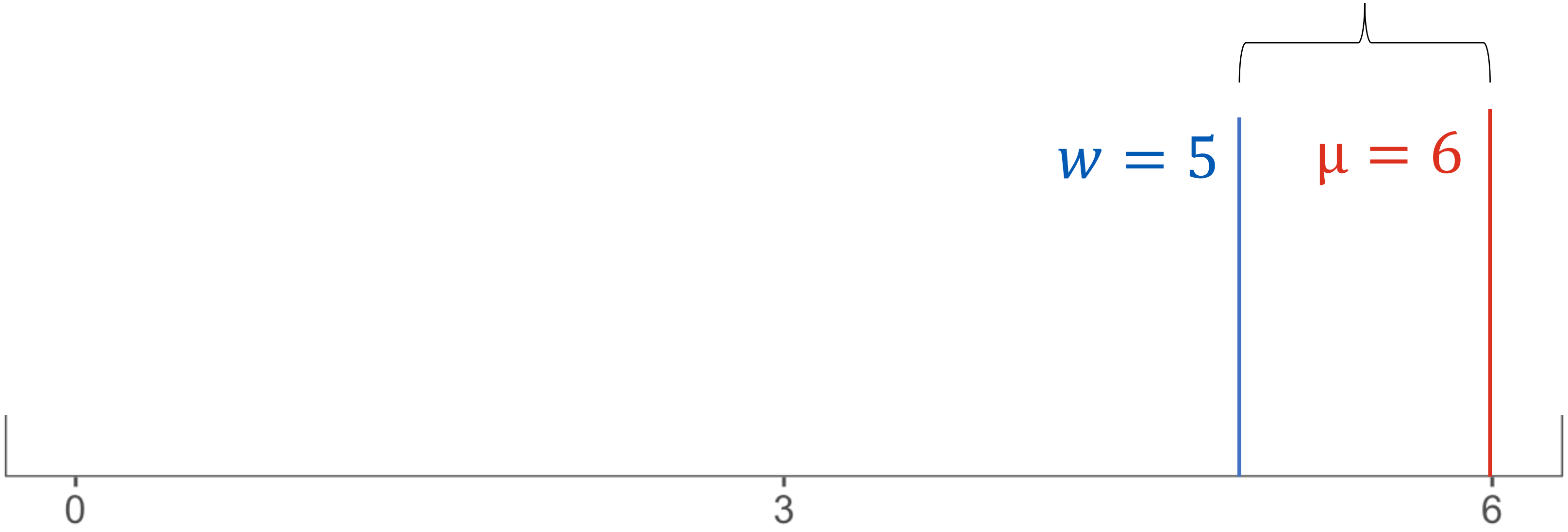
high μ = high preferred weight

$$\text{Prior} = \frac{(w_m - \mu_m)^2}{2\sigma^2}$$

w = constraint weight

μ = "preferred" weight

$$(5-6)^2 = 1$$



Setting μ for our models

Frequency-matching

Generalization: no markedness bias

Model: μ=0 for all constraints (uniform prior)

Markedness bias

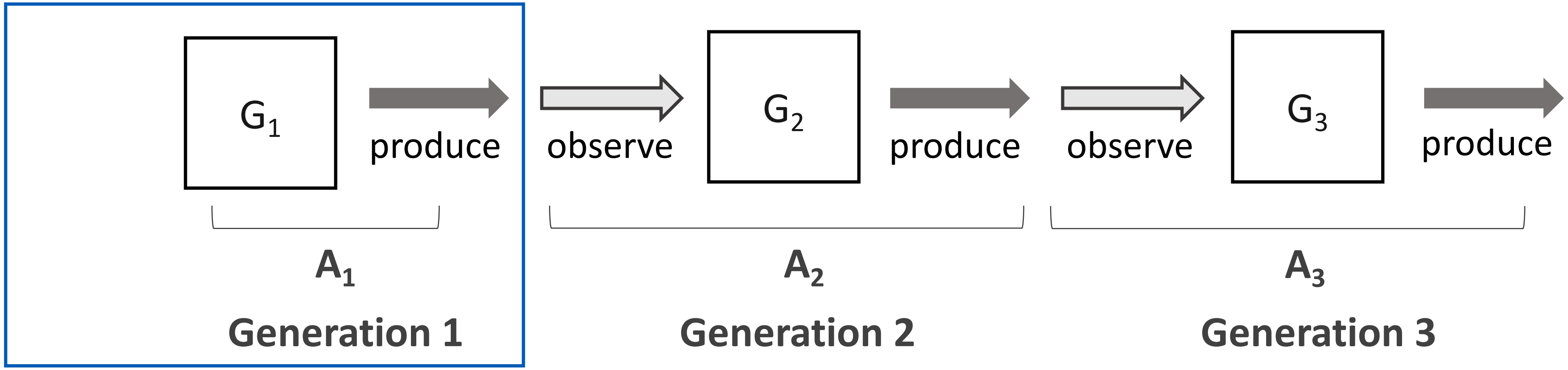
Generalization: dispreference for /p, t, ʈʂ, k/ between vowels

Model: μ(*VTV)=3, otherwise μ=0

Elements in a model of reanalysis

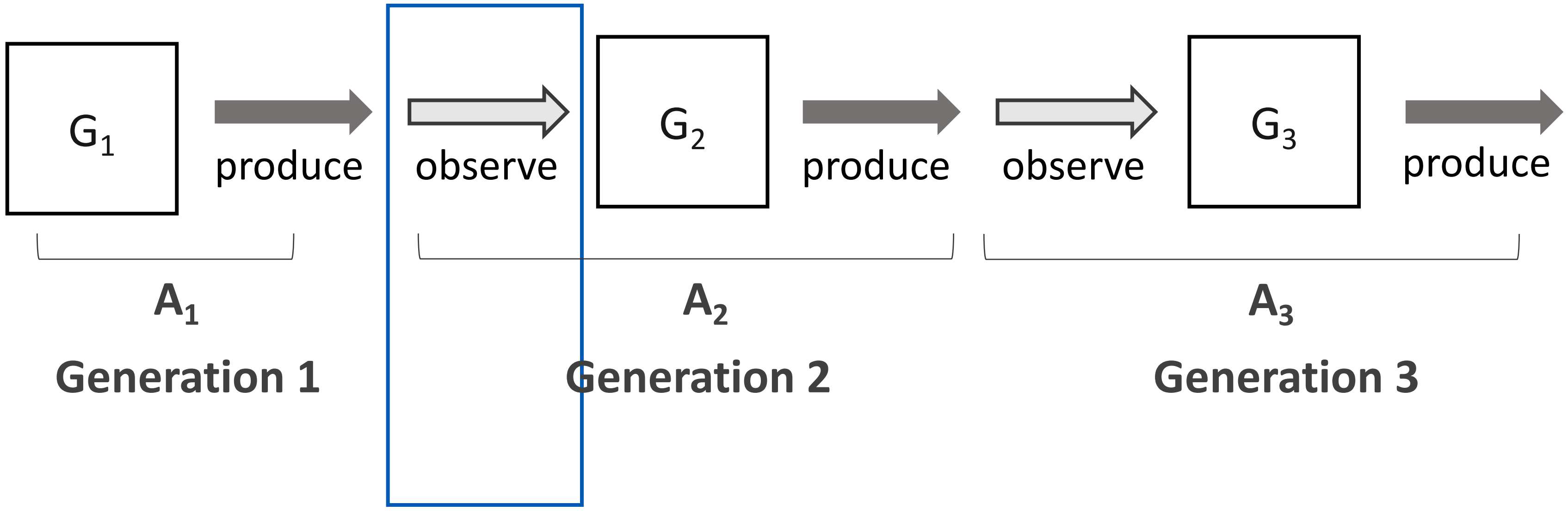
1. A probabilistic phonological grammar
2. Ability to incorporate learning biases
3. **Simulate generations of change**

3 Iterated learning



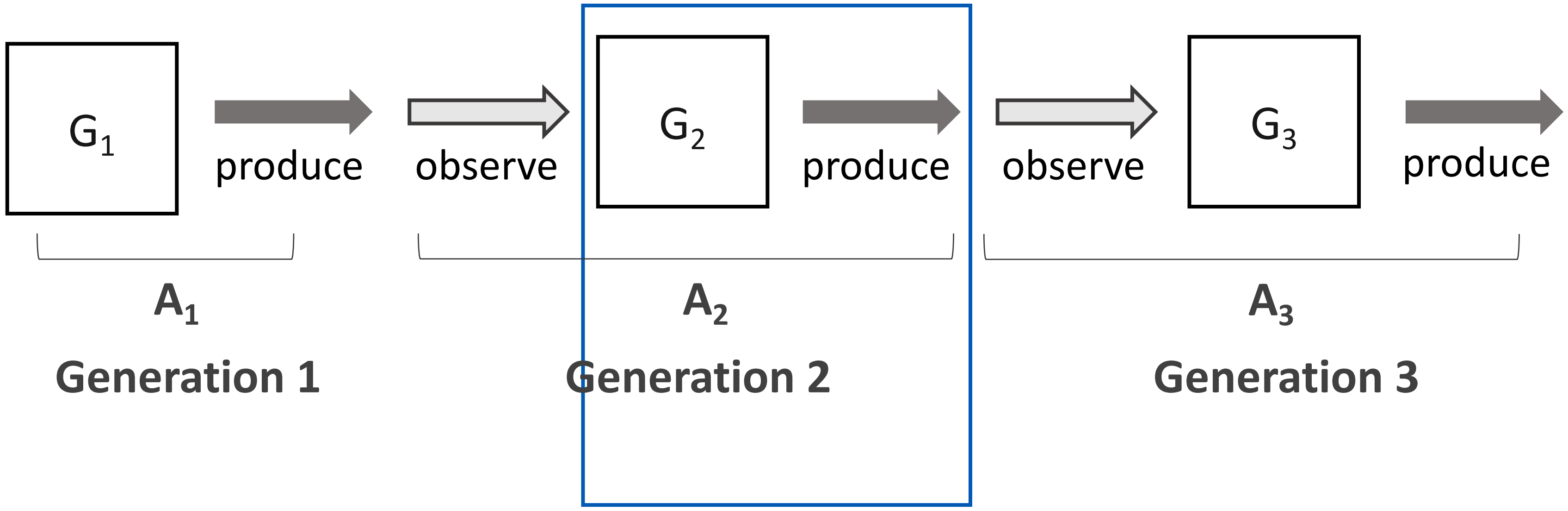
Agent A_1 produces words based on their knowledge of the grammar (G_1), which A_2 observes

3 Iterated learning



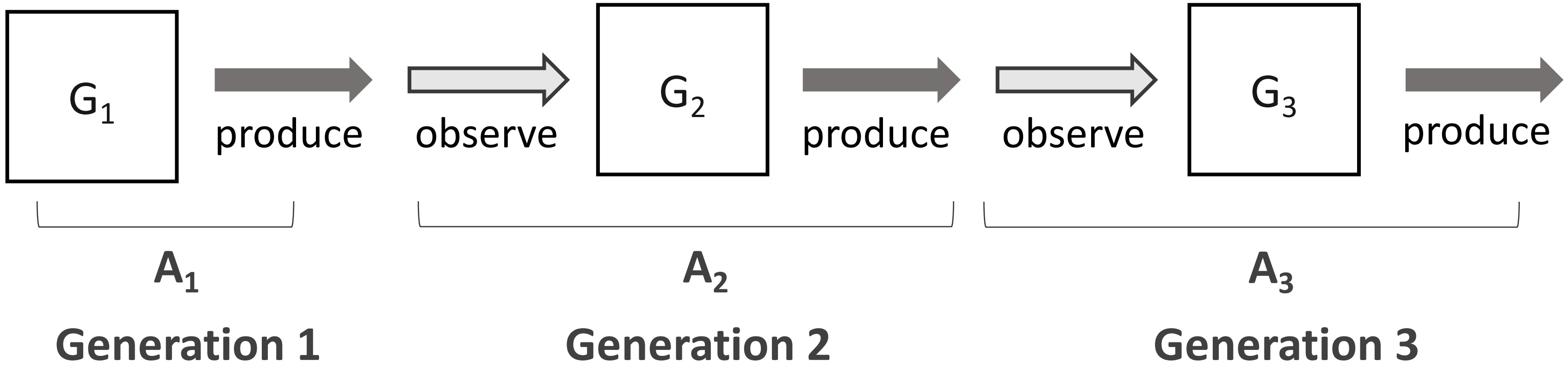
“Bottleneck”: A_2 **forgets** some proportion of words.

3 Iterated learning



Agent A_2 learns a grammar (G_2) based on the remembered words and uses it to generate the forgotten words.

3 Iterated learning



- Used to simulate change/evolution (for a review: Kirby, Griffiths, & Smith 2014)
- Several parameters I won't go into detail on.

Elements in a model of reanalysis

1. A probabilistic phonological grammar ✓
2. Ability to incorporate learning biases ✓
3. Simulate generations of change ✓

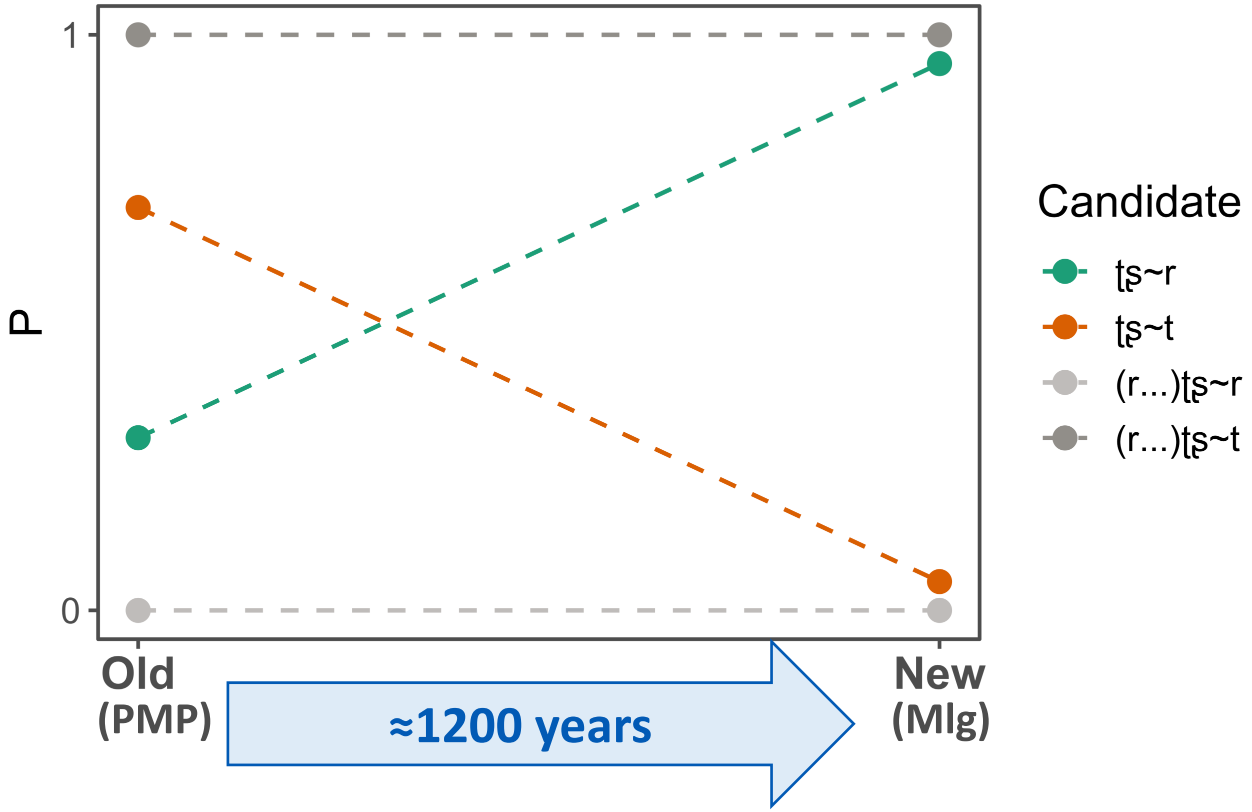
Results

1	2	3	4
Intro	Malagasy reanalysis	Model	**Results**

Reviewing the Malagasy data: all stem types

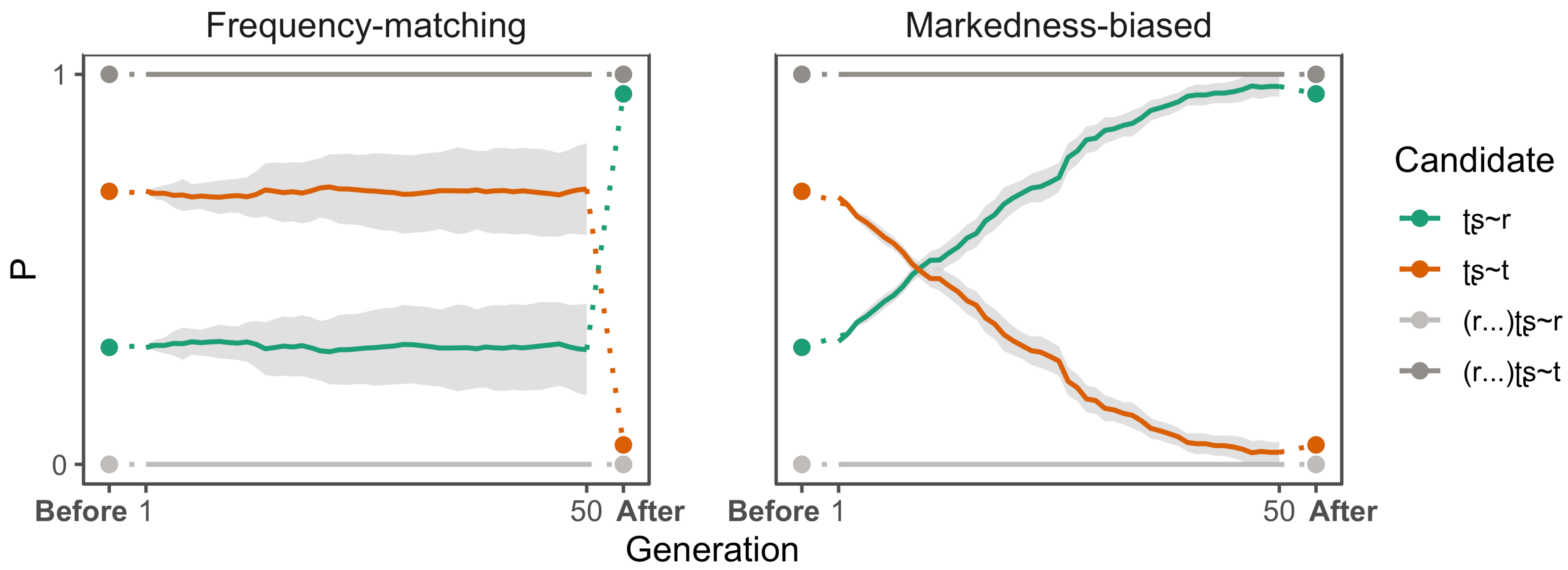
	PMP	**Malagasy**
ka words	prefer [h]	prefer [h]
na words	prefer [n]	prefer [n]
ţşa words	prefer [t] avoid r…r	**prefer [r]** avoid r…r

Reviewing the Malagasy data: ʈʂa stems



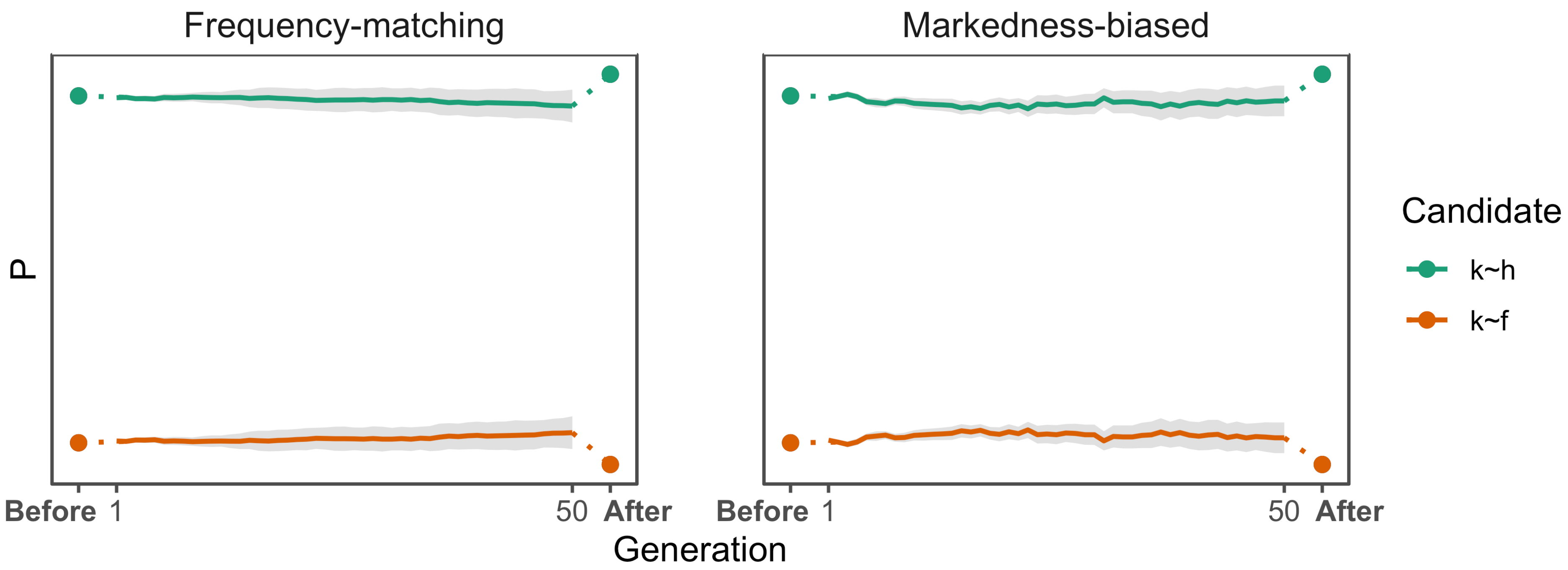
Markedness bias improves model predictions

Figure: Predicted proportion of suffixed form outputs for **ʈʂa** weak stems
(forget rate = 0.2)



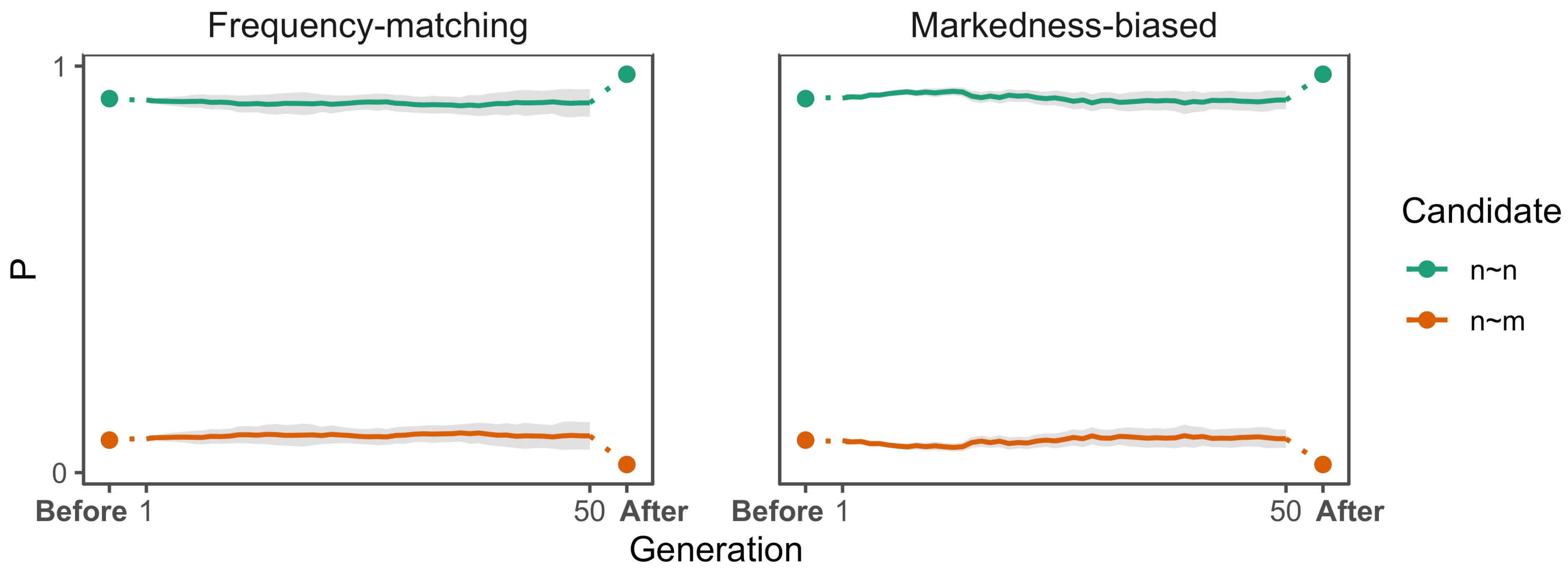
Markedness model performs well on **all** weak stems

Figure: Predicted proportion of suffixed outputs for **ka** weak stems (forget rate = 0.2)



Markedness model performs well on all weak stems

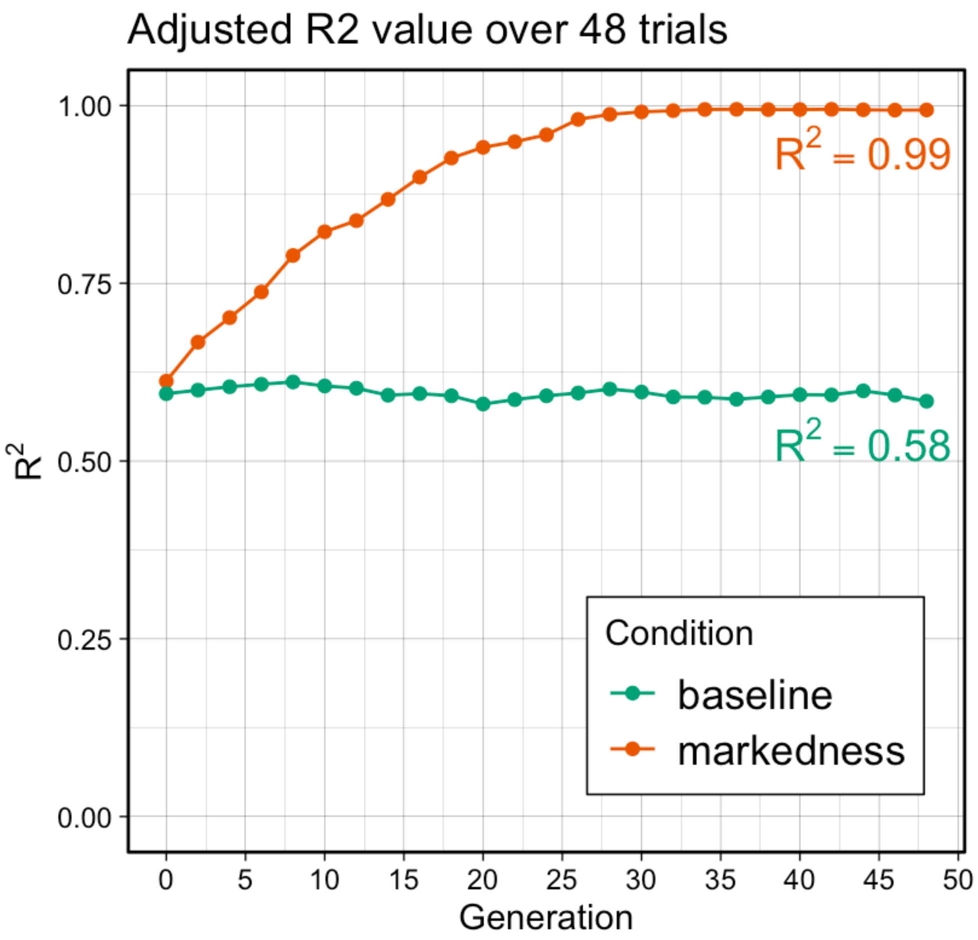
Figure: Predicted proportion of suffixed form outputs for **na** weak stems



Markedness model performs better overall

Figure: proportion variance accounted for (R^2), fit to **modern Malagasy** data

	Log likelihood
baseline	-9273
markedness	-6033



Summing up

1. Show that reanalysis in Malagasy can be explained as statistical learning + (active) markedness bias
    - ➢In ʈʂa words, t→r is motivated by giving *VTV a bias towards higher weight
    - ➢Markedness effect is active in the root phonotactics.

Summing up

1. Show that reanalysis in Malagasy can be explained as statistical learning + (active) markedness bias
   - ➢In ʈʂa words, t→r is motivated by giving *VTV (UsePhonotactics) a bias towards higher weight
   - ➢Markedness effect is active in the root phonotactics
2. Outline a model for incorporating markedness effects into reanalysis.
   - ➢MaxEnt HG with Gaussian prior + iterated learning.

Summing up

1. Show that reanalysis in Malagasy can be explained as statistical learning + (active) markedness bias
   - ➢In ʈʂa words, t→r is motivated by giving *VTV (UsePhonotactics) a bias towards higher weight
   - ➢Markedness effect is active in the root phonotactics
2. Outline a model for incorporating markedness effects into reanalysis.
   - ➢Maximum Entropy HG with Gaussian prior + iterated learning.
3. Demonstrate how quantitative models can be used to test theories about language learning in the absence of direct evidence.

Summing up

- Theories of reanalysis should be supplemented by markedness bias.
- Language change can be a **"natural laboratory"** for how humans learn (Kiparsky 1965; 1968; 1978, et seq)
- Where quantitative techniques are particularly helpful!

Thank you!

Thank you to…

My consultant Vololona Rasolofoson for her time and contribution.

Bruce Hayes, Kie Zuraw, Claire Moore-Cantwell, David Goldstein, members of the UCLA Phonology seminar, and other colleagues and friends for their time and feedback.

References I

- Adelaar, A. (2012). {Malagasy} phonological history and {Bantu} influence. *Oceanic Linguistics*, *51*(1), 123–159.
- Adelaar, A. (2009). Malagasy. In M. Haspelmath & U. Tadmor (Eds.), *World loanword database ({WOLD})*. Max Planck digital library.
- Albright, A., & Hayes, B. (2002). Modeling English past tense intuitions with minimal generalization. In *Proceedings of the {ACL-02} workshop on Morphological and phonological learning* (pp. 58–69).
- Albro, D. M. (2005). *Studies in computational {Optimality Theory}, with special reference to the phonological system of {Malagasy}*. University of California, Los Angeles.
- Blust, R. (1984). The Austronesian homeland: a linguistic perspective. *Asian Perspectives*, *26*(1), 45–67.
- Blust, R., & Trussel, S. (2010). Austronesian comparative dictionary, web edition.
- Boersma, P. (1998). *Functional phonology*. Netherlands Graduate School of Linguistics.
- Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*(1), 25–54.
- Chen, S. F., & Rosenfeld, R. (1999). *A Gaussian prior for smoothing maximum entropy models*.
- Chong, A. J. (2020). Exceptionality and derived-environment effects: A comparison of {Korean and Turkish}. *Phonology*.
- Chong, A. J. (2021). The effect of phonotactics on alternation learning. *Language*, 97(2), 213-244.
- Dahl, O. C. (1988). Bantu Substratum in {Malagasy}. *Études Océan Indien*, *9*, 91–132.
- Dahl, O. C. (1951). {Malgache} et {Maanjan}: une comparaison linguistique. *Avhandlinger Utgitt Av Egede Instituttet*.
- Dahl, O. C. (1988). Bantu Substratum in Malagasy. *Études Océan Indien*, *9*, 91–132.
- De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*(4), 441–465.
- de La Beaujardière, J.-M. (2004). Malagasy dictionary and encyclopedia of {Madagascar}.

References II

- Della Pietra, S., Della Pietra, V., & Lafferty, J. (1997). Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(4), 380–393.
- Elkins, N., & Kuo, J. (2023). A prominence account of the Northern Mam weight hierarchy. In *Supplemental Proceedings of AMP 2022*.
- Ernestus, M. T. C., & Baayen, R. H. (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language*, *79*(1), 5–38.
- Gallistel, C. R. (1990). The organization of behavior. *Wiley, New York Gillner S, Mallot HA (1998) Navigation and Acquisition of Spatial Knowledge in a Virtual Maze. J Cogn Neurosci*, *10*, 445–463.
- Goldwater, S., & Johnson, M. (2003). Learning {OT} constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within {Optimality Theory}* (Vol. 111120).
- Grabowski, E., & Kuo, J. (2023). *Comparing {K-means and OPTICS} clustering algorithms for identifying vowel categories*.
- Jarosz, G. (2006). *Rich lexicons and restrictive grammars: maximum likelihood learning in Optimality Theory*.
- Keenan, E. L., & Polinsky, M. (2017). Malagasy (Austronesian). *The Handbook of Morphology*, 563–623.
- Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation*, *5*(2), 102–110.
- Kuo, J. (2020). *Evidence for base-driven alternation in Tgdaya Seediq*. UCLA.
- Kuo, J. (to appear). Evidence for prosodic correspondence in the vowel alternations of Tgdaya Seediq. *Phonological Data and Analysis*.
- Mahdi, W. (1988). *Morphophonologische Besonderheiten und historische phonologie des Malagasy* (Vol. 20). D. Reimer.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101--B111.
- McCarthy, J., & Prince, A. (1995). Faithfulness and reduplicative identity. In J. Beckman, L. W. Dickey, & S. Urbanczyk (Eds.), *Papers in Optimality Theory* (pp. 249–384).
- Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part {I}: Structure. *Language and Linguistics Compass*, *6*(11), 686–701.

References III

- Pater, J., & Tessier, A. M. (2005). Phonotactics and alternations: Testing the connection with artificial language learning. *University of Massachusetts Occasional Papers in Linguistics [UMOP]*, 31, 1-16.
- Moreton, E., & Pater, J. (2012). Structure and substance in artificial-phonology learning, part {II}: Substance. *Language and Linguistics Compass*, *6*(11), 702–718.
- Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, *49*(2), 85–117. https://doi.org/https://doi.org/10.1016/j.cogpsych.2003.12.002
- Prince, A., & Smolensky, P. (1993/2004). Optimality Theory: Constraint interaction in generative grammar. *Optimality Theory in Phonology*, 3.
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914.
- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52. https://doi.org/https://doi.org/10.1016/S0010-0277(98)00075-4
- Smolensky, P. (1986). *Information processing in dynamical systems: Foundations of harmony theory*.
- Steriade, D. (2001). The phonology of perceptibility effects: the {P-map} and its consequences for constraint organization. In K. Hanson & S. Inkelas (Eds.), *The nature of the word: Studies in honor of Paul Kiparsky* (pp. 151–180). Cambridge, MA: MIT Press.
- Stites, J., Demuth, K., & Kirk, C. (2004). Markedness vs. frequency effects in coda acquisition. In *Proceedings of the 28th annual Boston University conference on language development* (Vol. 2, pp. 565–576).
- Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552.
- White, J. (2017). Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a P-map bias. *Lg*, *93*(1), 1–36.
- Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982.
- Zuraw, K. (2010). A model of lexical variation and the grammar with application to Tagalog nasal substitution. *NLLT*, *28*(2), 417–472.

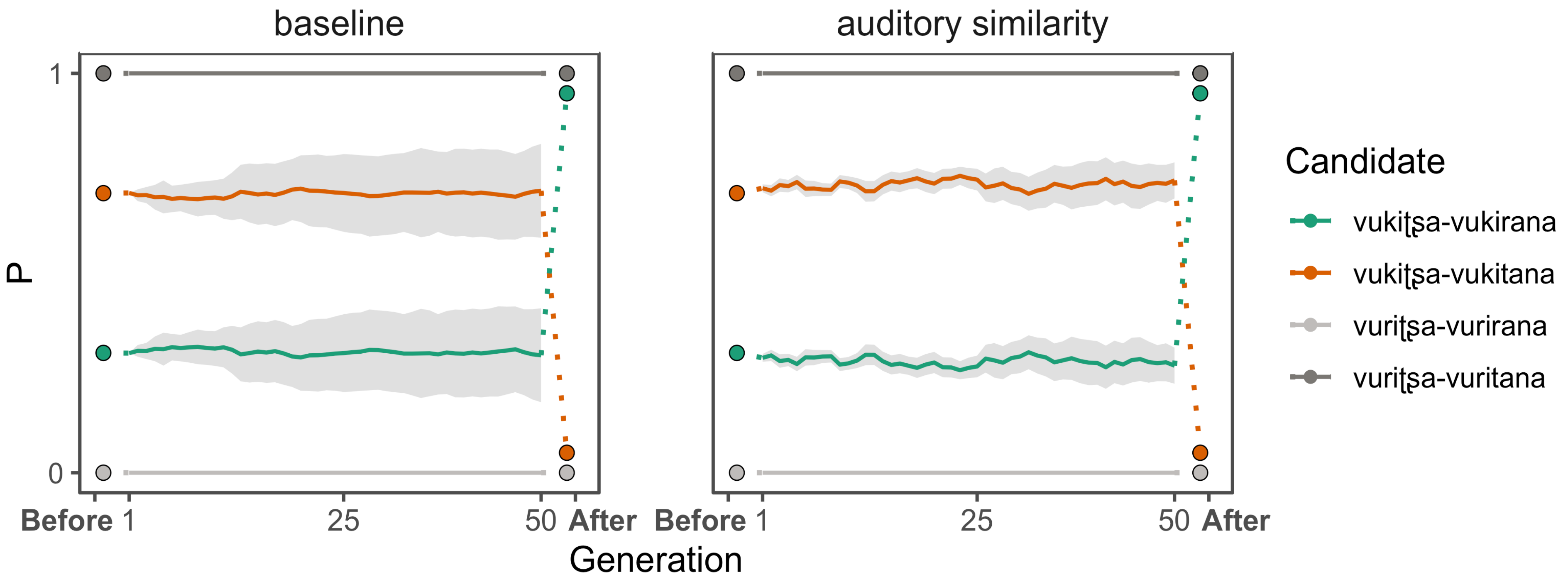
Perceptual similarity bias

- Constraints *map(*a, b)* penalizes changes from input *a* to output *b*
- μ(*/*a*/→[*b*]) > 0, otherwise μ=0
- The more dissimilar two sounds ***a*** and ***b*** are, the higher the μ of the corresponding */*a*/→[*b*]
  - i.e. bigger changes are penalized more

input	**output**	**Similarity**	**Constraint**	μ
vuki**ʈʂ**a+ana	vuki**r**-ana	low	*/ʈʂ/→[r]	4
	vuki**t**-ana	medium	*/ʈʂ/→[t]	1
	vuki**ʈʂ**-ana	high	NA	

Similarity derived from Warner, McQueen & Cutler (2014)

Perceptual similarity bias

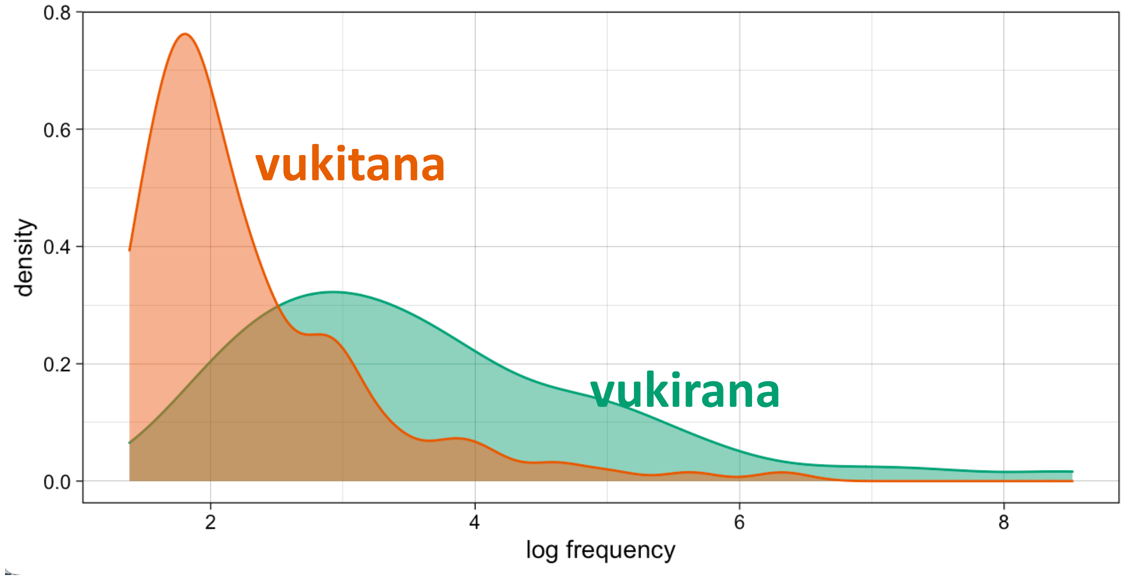


Token frequency

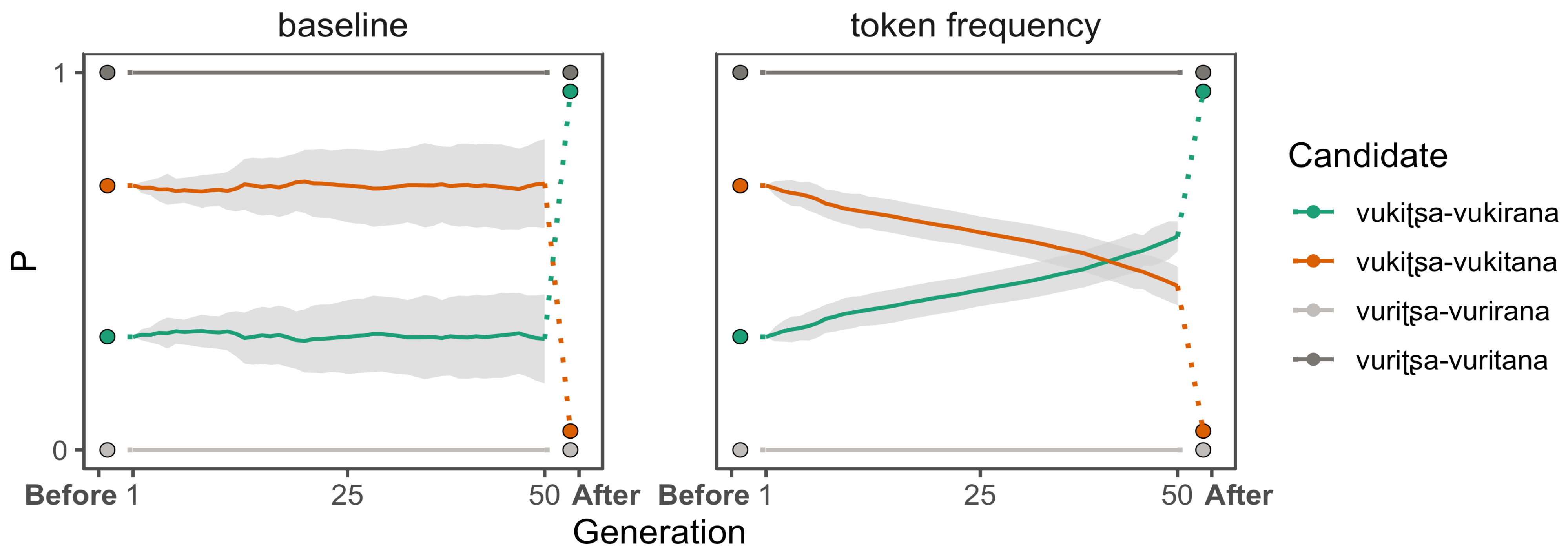
- In phonology, type frequency is a better predictor of phonological patterns (Bybee 1995; Bybee, 2001; Pierrehumbert 2001; Albright & Hayes, 2003)
- However, words with high token frequency:
  - Are more likely to be learned/passed down through generations
  - And may end up influencing a pattern (Albright, 2006).
- If ʈʂa~r forms have higher token frequency than ʈʂa~t forms, reanalysis could be from t→r

Token frequency

- Simulated input lexicon where ʈʂa-r words have high token frequency.
  - Zipfian distribution (Zipf, 1935/2013)
- Scale to log frequency (Marcus et al., 1992; Jackson & Cottrell 1997; Polinsky & Everbroek, 2003)



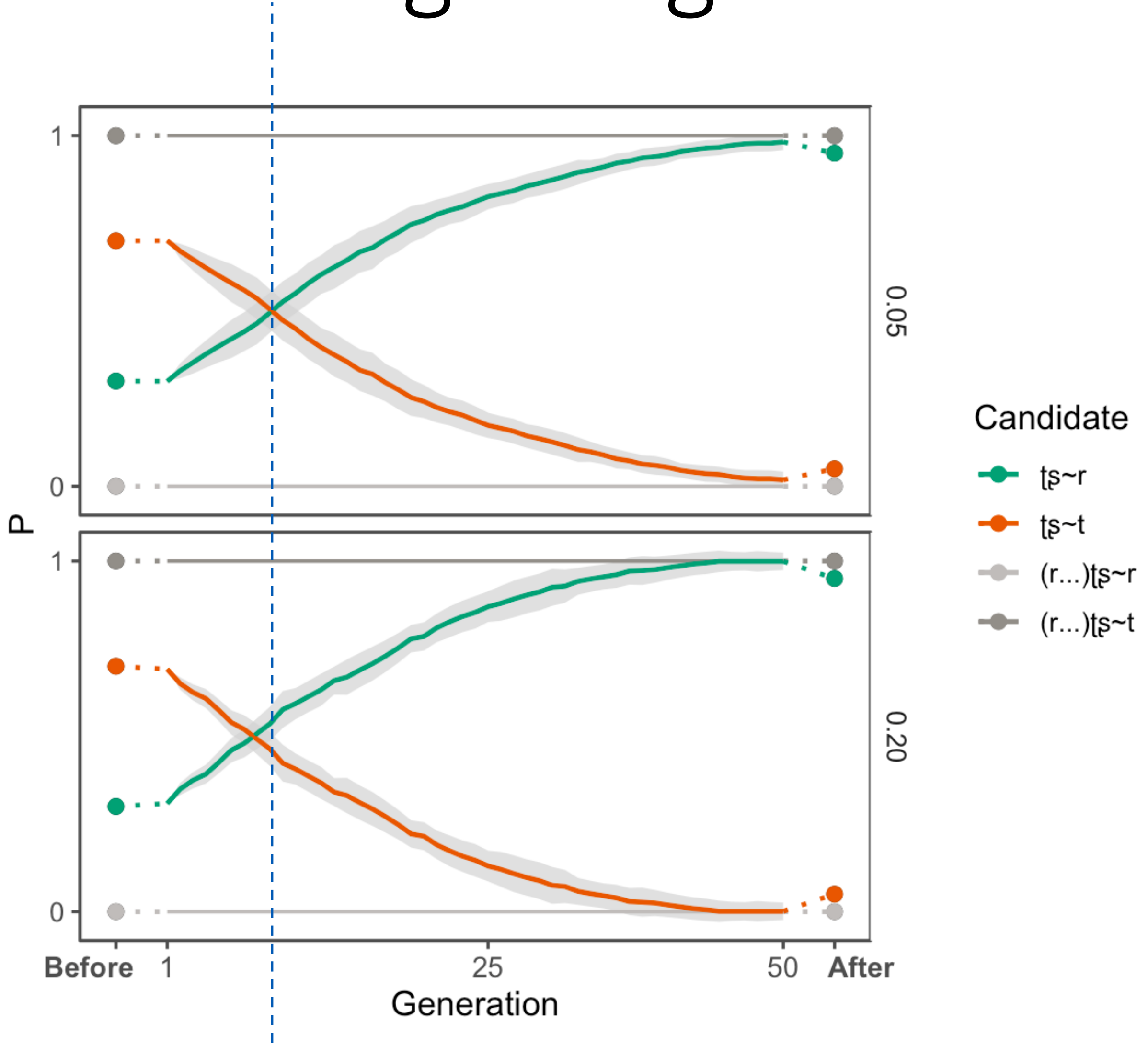
Token frequency



Underspecification analysis

/vuki**c**+ana/~[ˈvukitsa]	*ts]V	*t]V	*r]V	IO-FAITH	OO-FAITH
a. vukiʈʂana	*				
b. vukitana		*			*
c. vukirana			*		*

Effect of forgetting rates



Malagasy phonology

- Vowels: /a e i u (o)/
- Consonants:

	bilabial	labiodental	dental	alveolar	retroflex	velar	glottal
plosives*	p, b		t, d			k, g	
	^{m}p, ^{m}b		^{n}t, ^{n}d			^{ŋ}k, ^{ŋ}g	
affricates*				ts, dz	ʈʂ, ɖʐ		
				^{n}ts, ^{n}dz	^{n}ʈʂ, ^{n}ɖʐ		
nasals	m		n			(ŋ)	
trills/flaps				r∼ɾ			
fricatives		f, v		s z			h
lat. approximants				l			

- (C)V syllables structure (no codas)

2 Learn markedness from stem phonotactics

***VTV**	Assess violation for voiceless stops/affricates between vowels (/p, t, k, ʈʂ/) *[+syllabic][-continuant,-voice][+syllabic]
*ʈʂ]V, *k]V, *n]V	Assess violations for every C]V, where C is at a *morpheme boundary* (Pater 2007; Chong 2020) • within stems, prevocalic ʈʂ, k, and n are allowed • (e.g. beʈʂuka 'to swell up')
*r…r	Assess violation for sequences of r…r
*IDENT-IO[F]	The specification for [F] in an input segmen[t is] preserved in its output correspondent (McCarthy & Prince 1995)

Can we use a constraint directly derived from stem phonotactics?

2 Learn markedness from stem phonotactics

1. Stem phonotactics

[paka]
[manu]
[rara]
[tai]
…

Phonotactic grammar

UCLA Phonotactic Learner (Hayes & Wilson 2008)

Based in MaxEnt, assigns words penalty scores based on phonotactic well-formedness.

2 Learn markedness from stem phonotactics

1. Stem phonotactics

[paka]
[manu]
[rara]
[tai]
…

→ Phonotactic grammar

2. Model of reanalysis

Candidates
[vukirana]
[vukitana]
[vukiʈʂana]
…

→ Phonotactic grammar →

Penalty (H)	
[vukirana]	0.23
[vukitana]	1.13
[vukiʈʂana]	1.27
…	

→

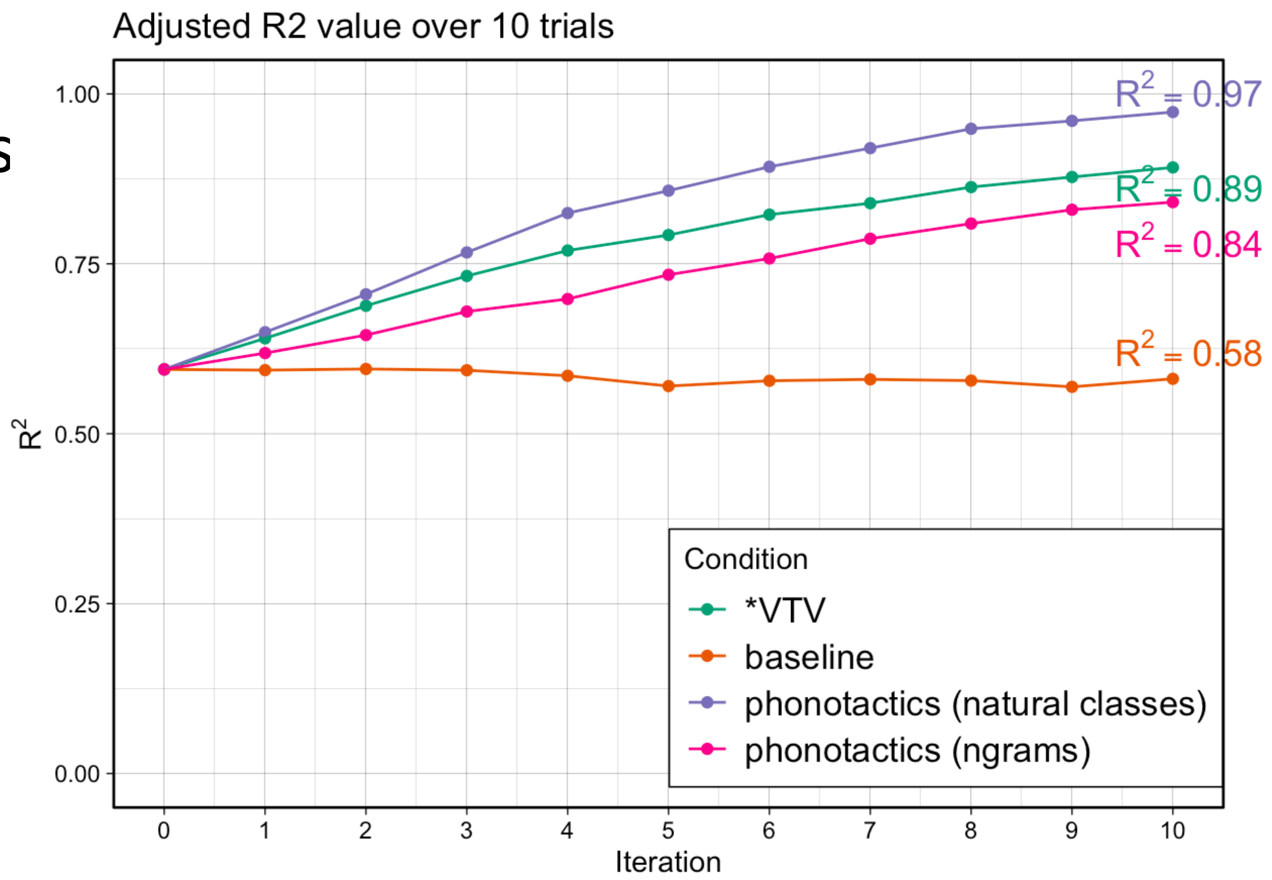
	PHONOTACTICS	**C1**	**C2 …**
[vukirana]	0.23	1	0
[vukitana]	1.13	0	1
[vukiʈʂana]	1.27	0	0
…			

2 Learn markedness from stem phonotactics

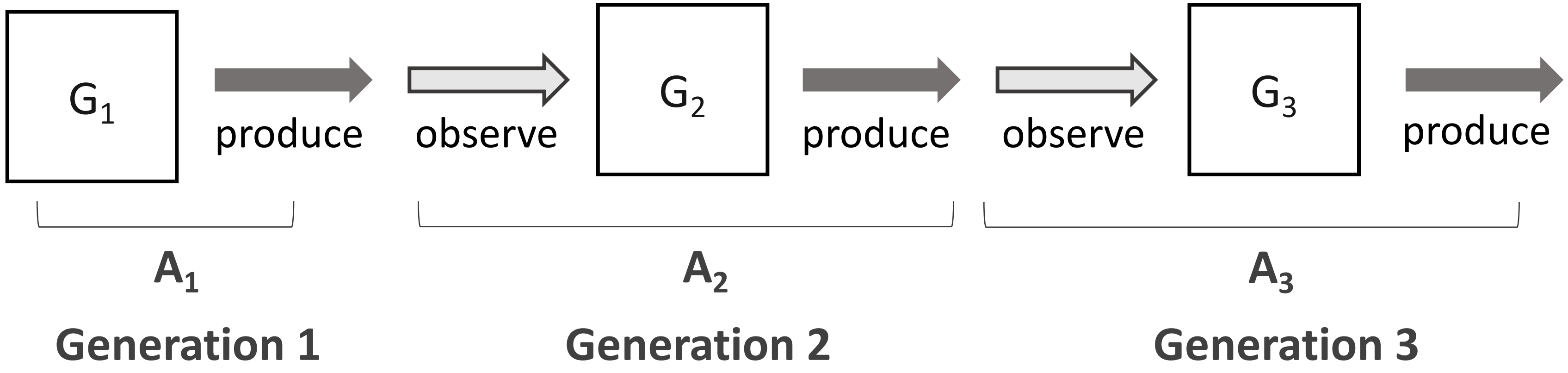
USEPHONOTACTICS	Constraint violations are derived directly from stem phonotactics
*ʈʂ]V, *k]V, *n]V	Assess violations for every C]V, where C is at a *morpheme boundary* (Pater 2007; Chong 2020) • within stems, prevocalic ʈʂ, k, and n are allowed • (e.g. beʈʂuka ‘to swell up’)
*r…r	Assess violation for sequences of r…r
*IDENT-IO[F]	The specification for [F] in an input segment must be preserved in its output correspondent (McCarthy & Prince 1995)

Markedness bias and phonotactics

- The phonotactic model that generalizes to natural classes performs the best
- All three models outperform the baseline



3 Iterated learning



Parameters:

- Forgetting rate [0, 1]
  - values: 0.05, 0.1, 0.15, **0.2**, 0.25
- 50 (25 years/generation, from 600-1800AD)
- Mean of 30 runs.