When learning is not statistical: how learning biases have remolded Malagasy sound patterns

Jennifer Kuo

University of California, Los Angeles

Research overview

- Phonological learning: how do people acquire systems of sound patterns?

Computational modeling
Kuo 2020; 2022; diss. in progress

Experimental evidence
Kuo 2020; Kuo, *to appear*

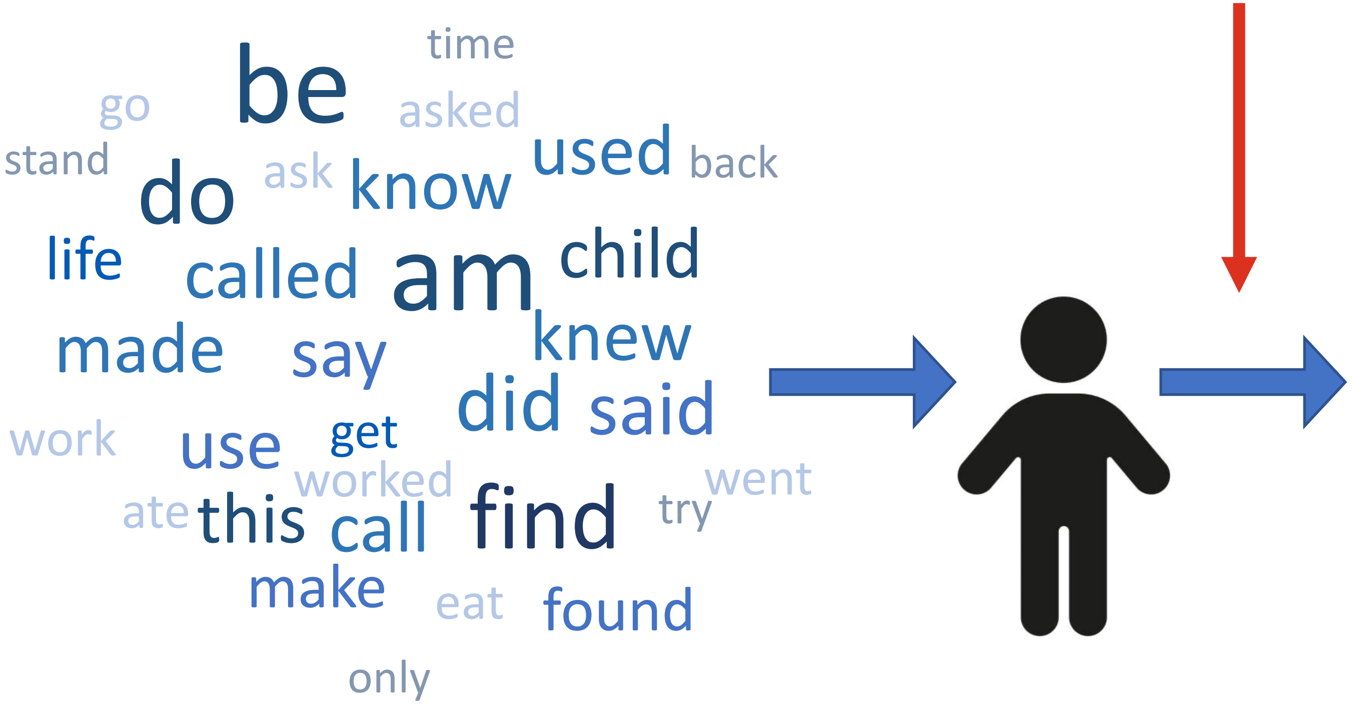
Corpora
Grabowski & Kuo, 2023

Fieldwork (Tgdaya Seediq, Mam)
Kuo 2020; Kuo & Elkins 2022

Insights from understudied languages
Seediq (Taiwan), Malagasy (Madagascar), Samoan (Samoa, American Samoa), Māori (New Zealand), Mam (Guatemala)

Phonological learning

How does this happen?



- /i/ is a sound in English
- "blick" sounds better than "bnick."
- Add "ed" to form past tense.

….and so on!

How does phonological learning happen?

Domain-general or **Language-specific**

Tools:
historical change in Malagasy +
computational modeling

How do learners deal with conflicting patterns?

PRESENT	PAST
laugh	laugh**ed**
dance	danc**ed**
jump	jump**ed**
heed	heed**ed**

Rule: add "ed" to form past tense

….

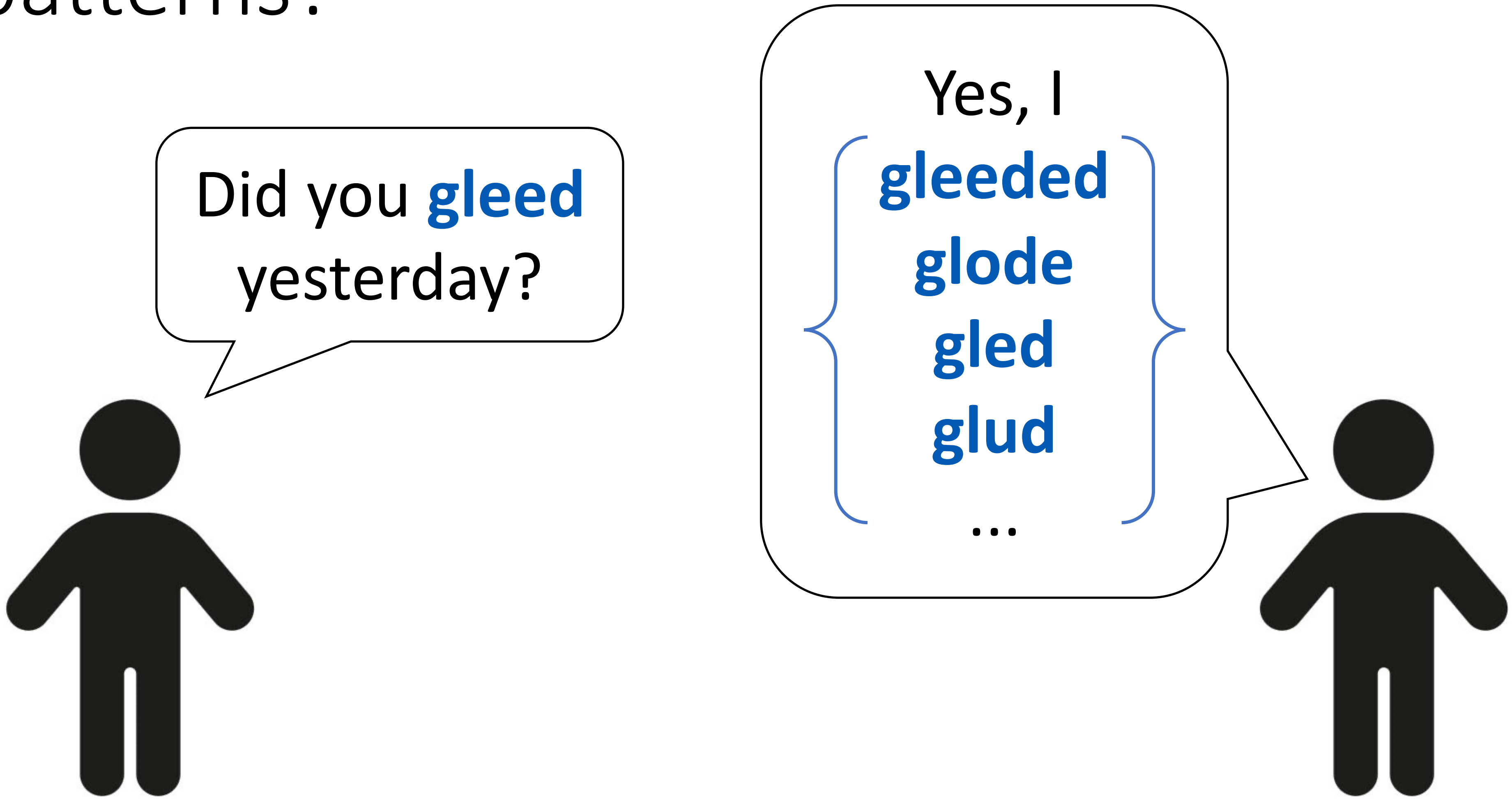
bleed [blid]	bl**e**d [bl**ɛ**d]
read [rid]	r**ea**d [r**ɛ**d]

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**

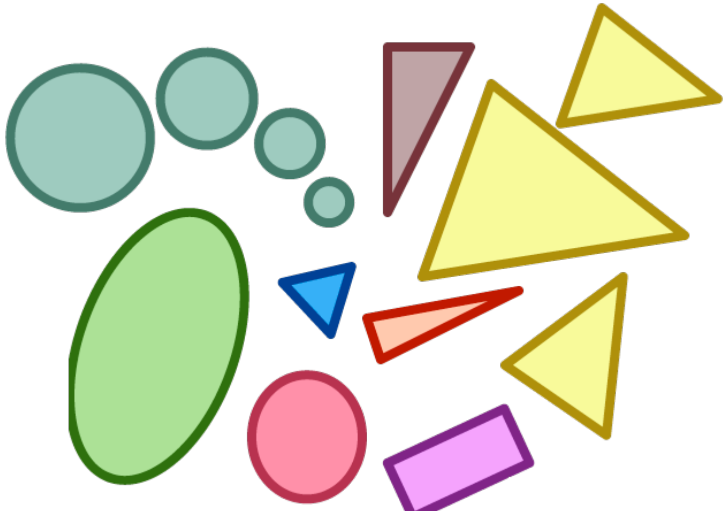
Conflicting data patterns lead to **variance** that is informative

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 say/hear** (markedness bias; Jarosz 2006)

Factors driving reanalysis

- Existing models are frequency-matching

plead → pleaded	**Rule:** add “ed” to form past tense N=1146/1234 (93%)
pled	**Rule:** if a word ends in [id], change [i]→[ɛ] N=6/7 (86%)
plode	**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: In Malagasy, reanalysis is **not** entirely predictable from statistical learning.
- My proposal: reanalysis is sensitive to a **markedness bias**
  - “Marked”= cross-linguistically dispreferred, because of being harder to say or hear.

How does phonological learning work?

Frequency-matching

+

Linguistically-motivated biases?

Tools: historical change (reanalysis) + computational modeling

Goals of the talk

1. Show that reanalysis in Malagasy can be explained as **statistical learning** + **markedness bias**
2. Demonstrate how computational models can be used to test theories about language learning.
3. Outline a model for incorporating markedness effects into reanalysis.

Outline

1

Intro

Facts of Malagasy

2

Malagasy reanalysis

Evidence **against** statistical learning.

3

Model

The cumulative effect of **statistical learning** + **learning biases** over generations of speakers.

4

Results

Demonstrate that Malagasy reanalysis is sensitive to a **markedness bias.**

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.

Malagasy phonology

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

Weak stems (Albro 2005; Keenan and Polinsky 2017)

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

type	**consonant**	**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 has been simplified for ease of presentation, and do not accurately reflect all historical changes

Old pattern → Weak stems

UNSUFFIXED	SUFFIXED	PROCESS
avut	avut-an	
avu**tʂ**	avut-an	t,r→tʂ at the end of words
avutʂ**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

vuhit-ana?

vuhir-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

How did weak stems happen?

Generalizations taken from Dahl (1951, 1988), Mahdi (1988), Adelaar (2013)

Note: Data has been simplified for ease of presentation, and do not accurately reflect all historical changes

Old pattern

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

~600AD—now: Lots of reanalyses

Malagasy today

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.

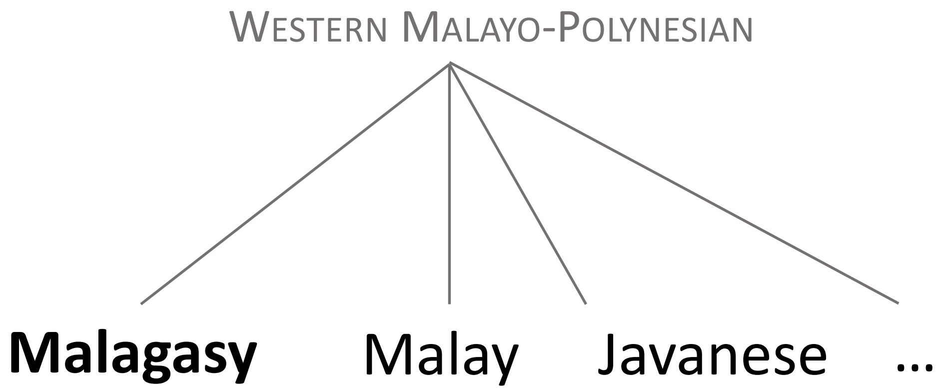
Methodology and data

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

Methodology and data

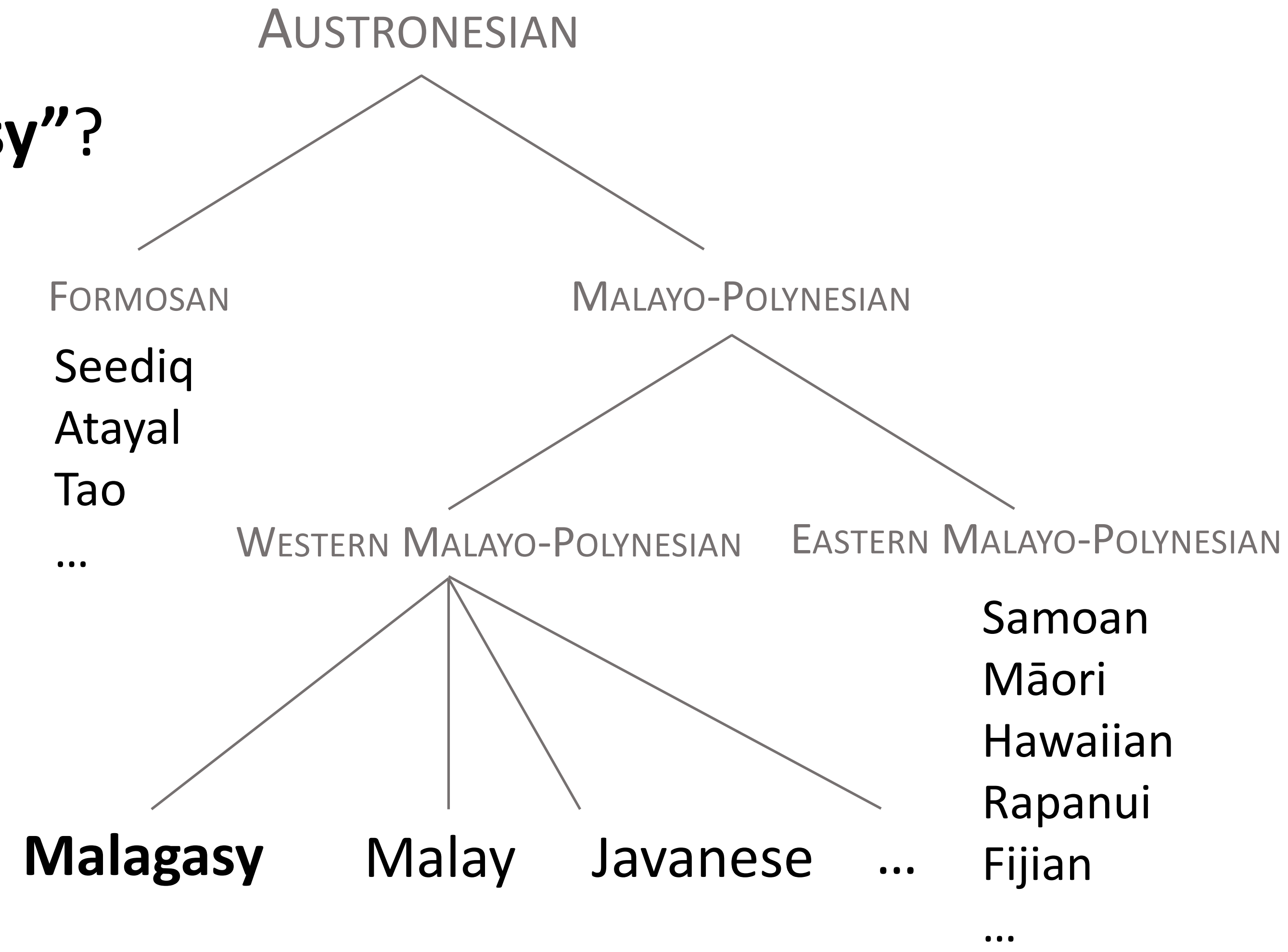
- What is **“old Malagasy”**?
- Insights from historical linguistics

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



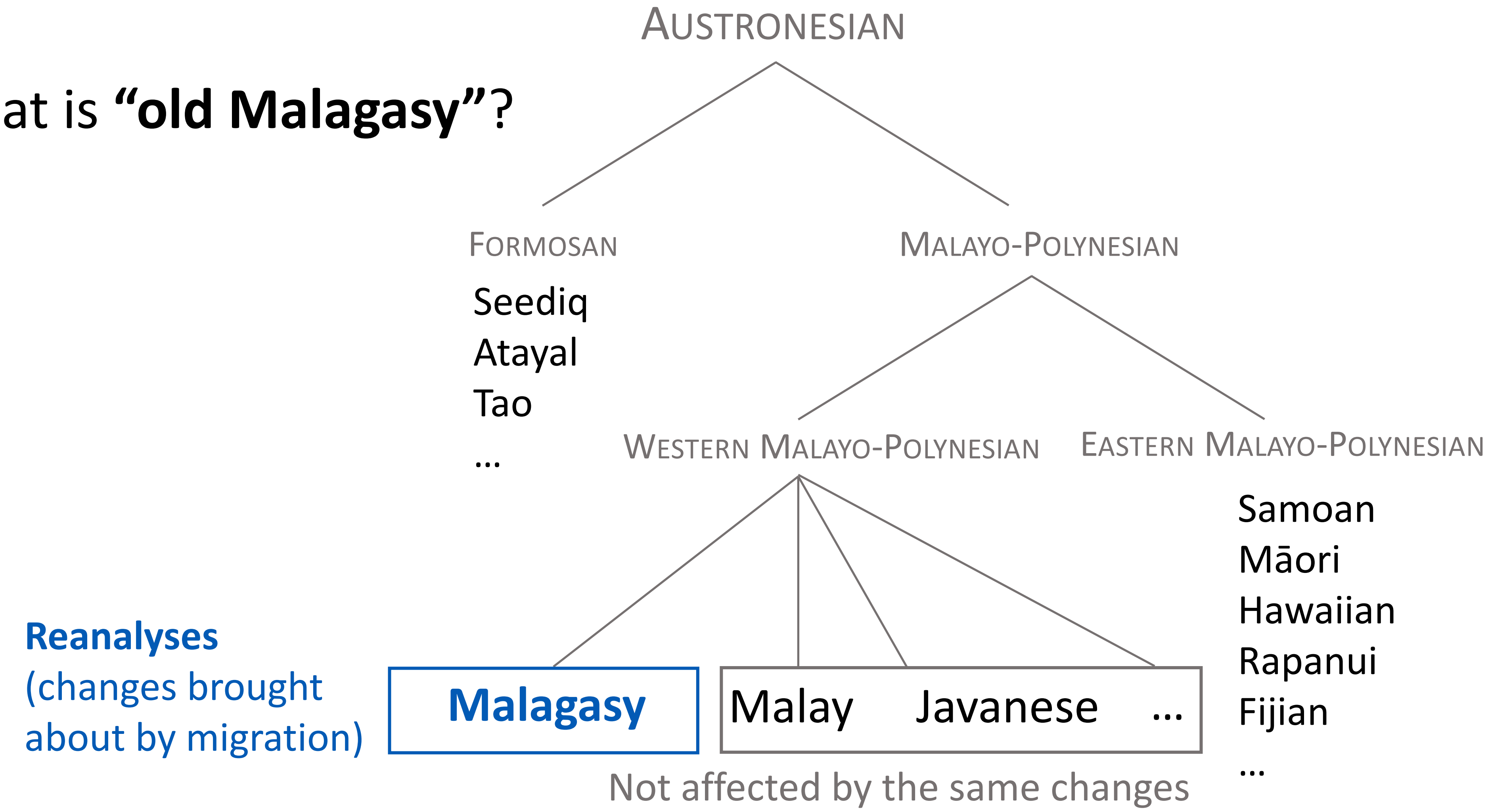
Methodology and data

- What is **“old Malagasy”**?



Methodology and data

- What is **“old Malagasy”**?



Methodology and data

“old” Malagasy	modern Malagasy
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 suffixed consonants in **old** Malagasy

(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 suffixed consonants in **old** Malagasy

(a) all words

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

(b) words with a preceding [r]

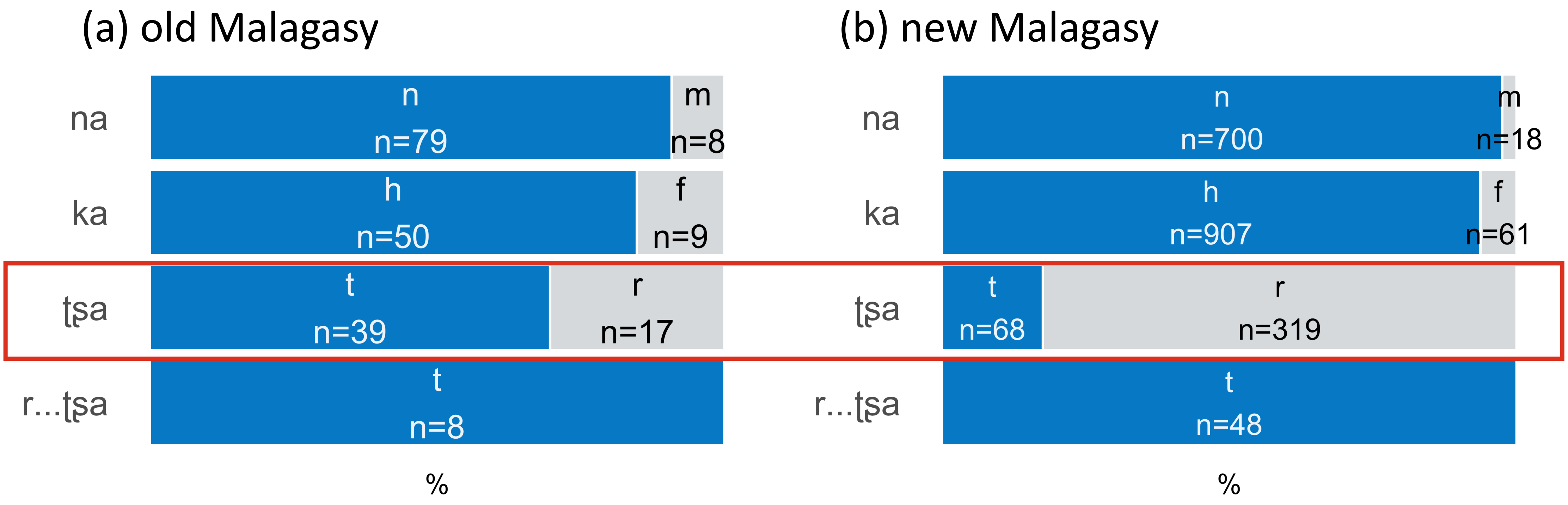
Consonant	n	%	
(ʈʂ~) r	0	0	puriʈʂa~purir-ana
(ʈʂ~) **t**	**8**	**100%**	puriʈʂa~purit-ana

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

Observed directions of reanalysis

- Indirect evidence of reanalysis: comparing old vs. new Malagasy

Distribution of alternants:



Observed directions of reanalysis

- Direct evidence: words that have undergone reanalysis

old→new	count	Example
r → r	18	velaʈʂa ~ velar-ana →velar-ana `to spread out’
r → t	1	saraʈʂ ~ sarar-ana →sarat-ana `to rise up’
t → t	23 (43%)	oroʈʂa ~ orot-ana→ orot-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

old→new	count	Example
r → r	18	velaʈʂa ~ velar-ana →velar-ana `to spread out'
r → t	1	saraʈʂ ~ sarar-ana →sarat-ana `to rise up'
t → t	23 (43%)	oroʈʂa ~ orot-ana→ orot-ana `to massage'
t → r	30 (57%)	akaʈʂa ~ akat-ana → akar-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]

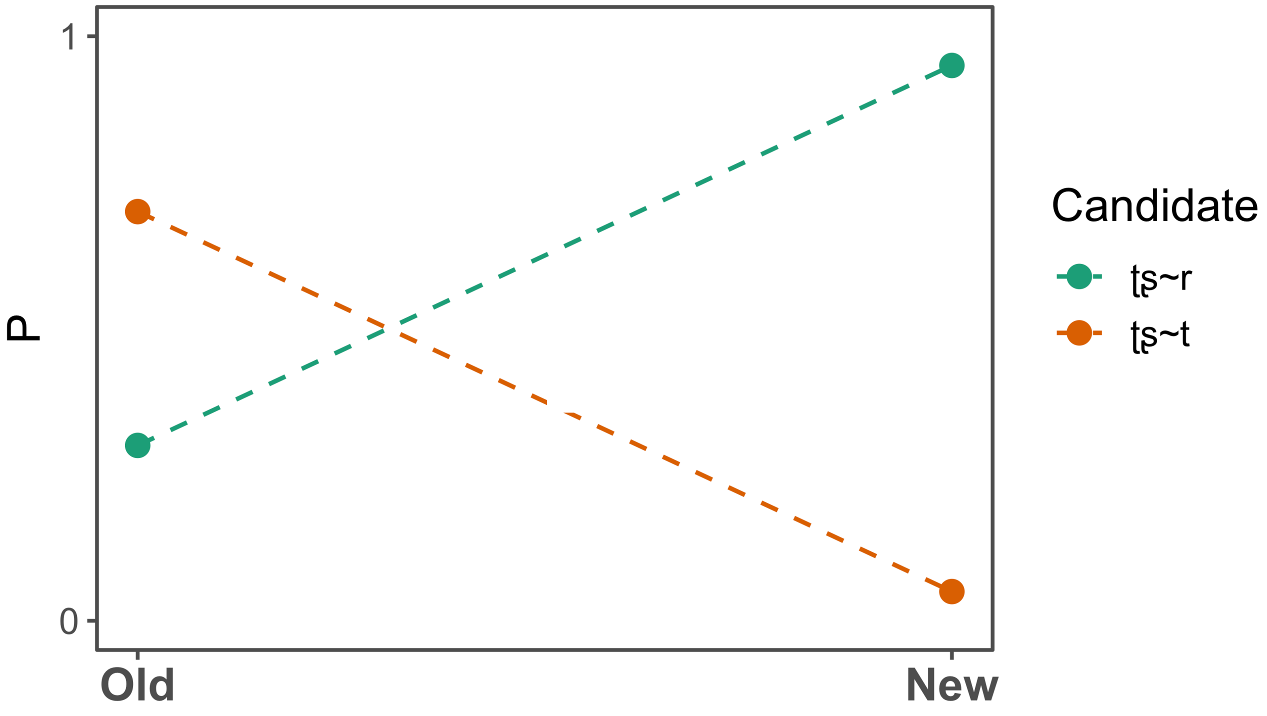
Summary of pattern

	Old	**New**
ka words	prefer [h]	prefer [h]
na words	prefer [n]	prefer [n]
ʈʂa words	prefer [t] avoid r…r	**prefer [r]** avoid r…r

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

input	output	Old	New
vukiʈʂa	vukiɾana	0.3	0.95
	vukiʈana	0.7	0.05

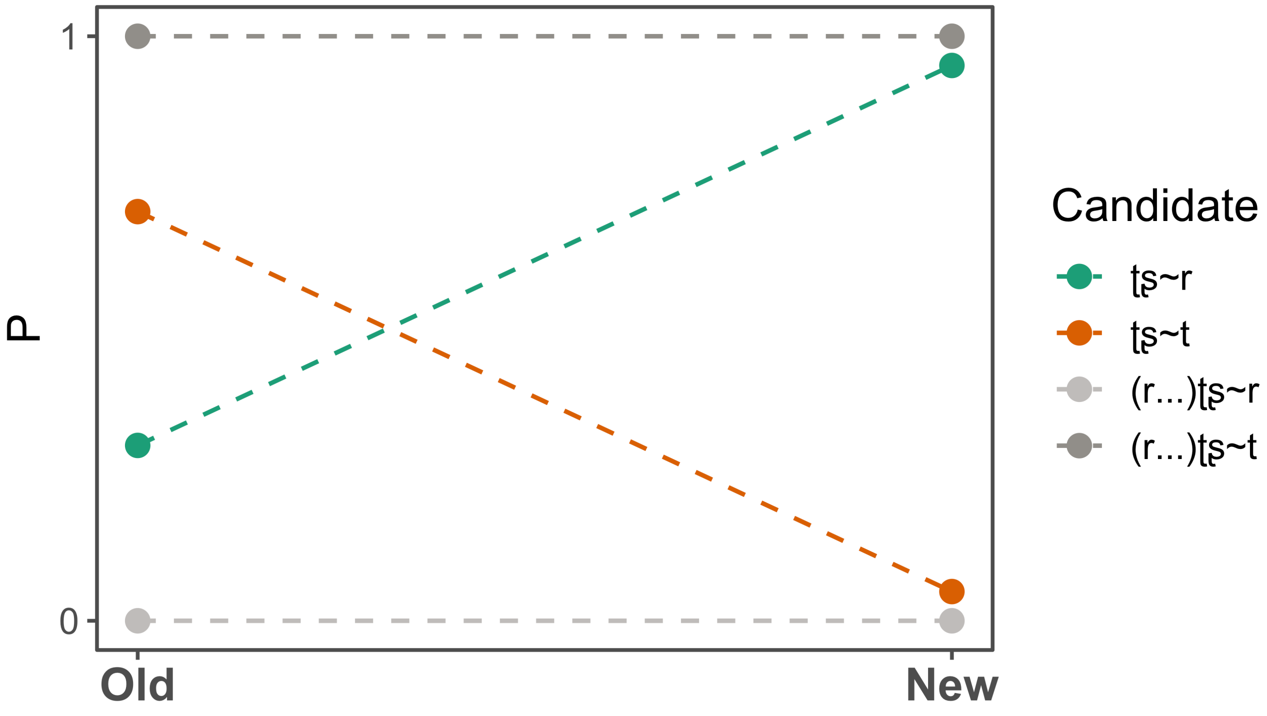
Figure: Proportion of consonant (t vs. r) that surfaces under suffixation:



Graphing the pattern: ʈʂa words (with preceding [r])

input	output	Old	New
vukiʈʂa	vukir̠ana	0.3	0.95
	vukit̠ana	0.7	0.05
vur̠iʈʂa	vur̠ir̠ana	0	0
	vur̠it̠ana	1	1

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



Statistical learning vs. markedness bias

- Reanalysis is not predictable from statistical distributions
- Proposal: Reanalysis is sensitive to a **markedness bias**

Stops between vowels are marked

- Languages disprefer (voiceless) **stops** between vowels /p t k ʈʂ/
  - Bad: atu, faike, papi, beʈʂuka…vulit-ana
  - Good: aro, azi, lumu, tafi, etc… vulir-ana
- harder to say/hear (Kirchner, 1998; Kaplan, 2010; Katz, 2016)
- cross-linguistically dispreferred
  - English ex: tapping! vo[t]e→ vo[ɾ]ing “vote/voting” (Hayes 2011, 143-144)

A model of 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

1 Phonological grammar

- Basic idea: the grammar has…
  - A mechanism for generating candidate outputs given an input
  - A series of constraints on the output

  (Optimality Theory; Prince & Smolensky 1993/2004)
- Ex: In English, a "sh" followed by "s" is not allowed (*sh-s)

Example outputs

Input

/dish/ + /s/ 'PLURAL'

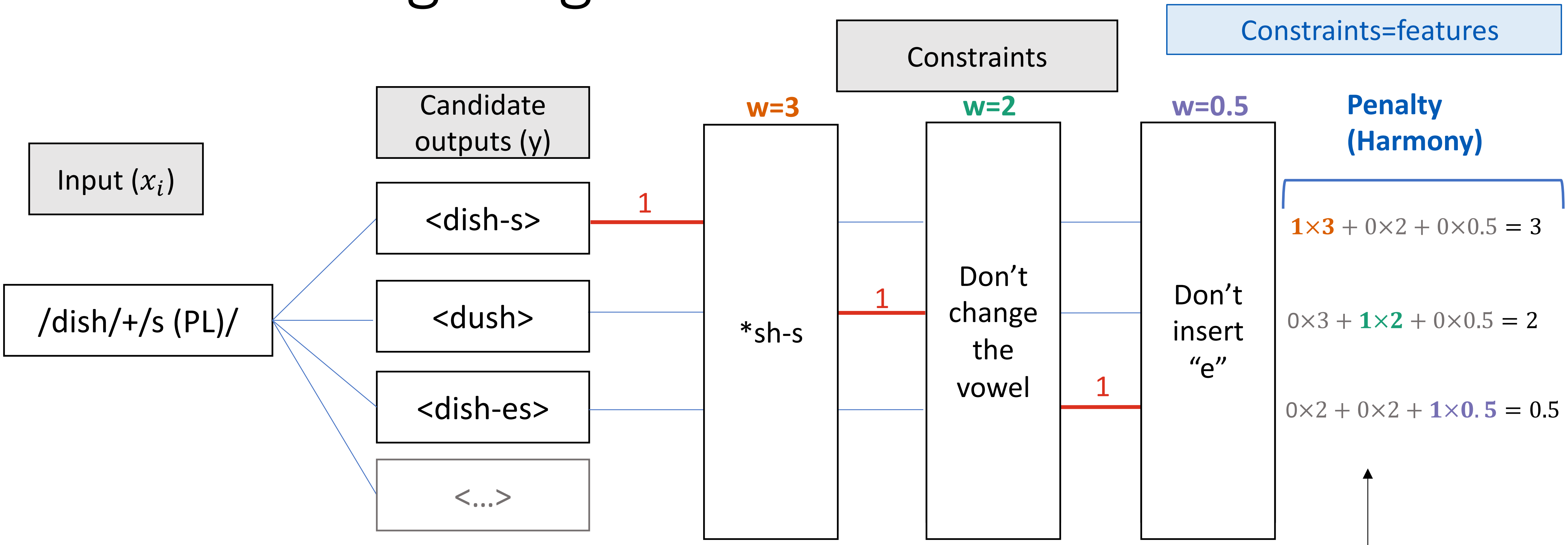
*sh-s X → "dish-s"

→ "dish-es"

1 Phonological grammar

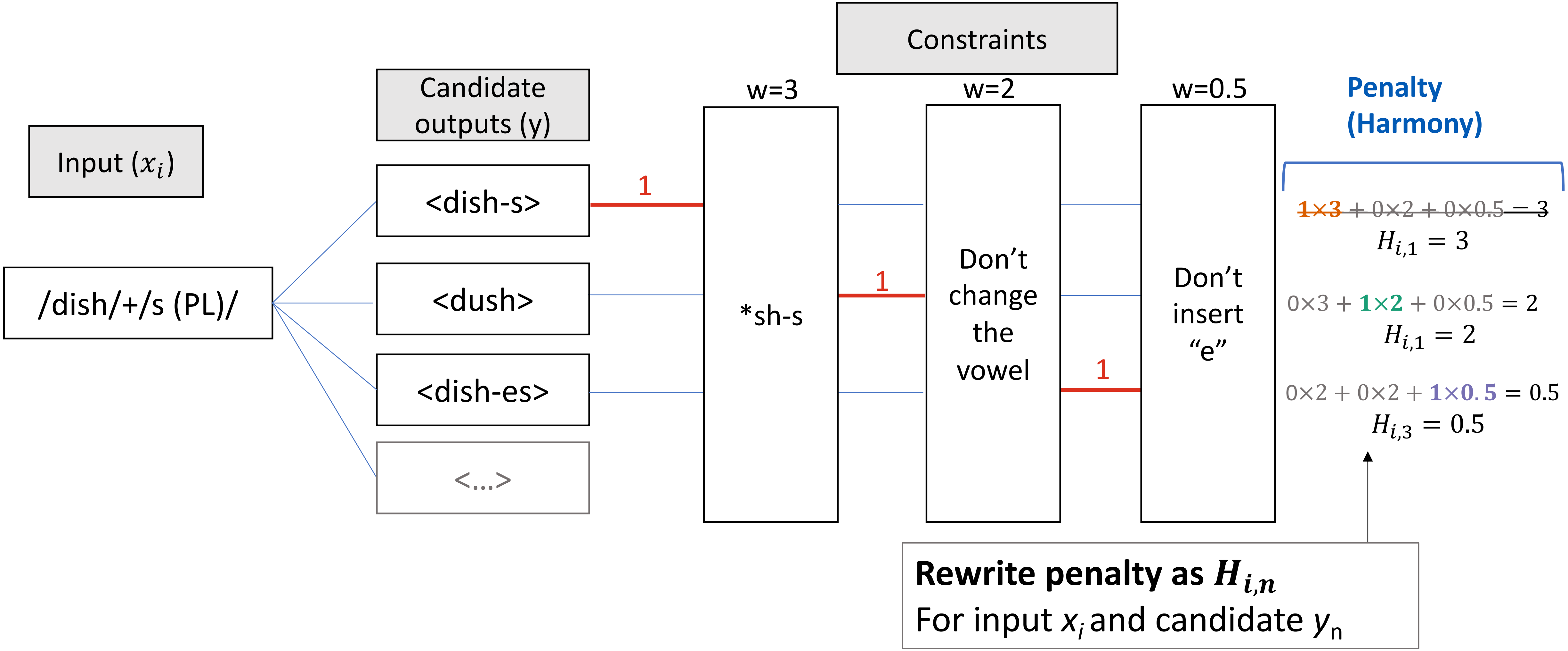
- The grammar also generates probabilities for outputs.
  - Maximum Entropy Harmonic Grammar (e.g., Smolensky 1986; Goldwater & Johnson, 2003)
  - = multinomial logistic regression
- Why Maximum Entropy?
  - convex parameter space, converges to an optimal solution (Della Pietra et al. 1997)
  - Being considered as a model of acquisition

1 Phonological grammar

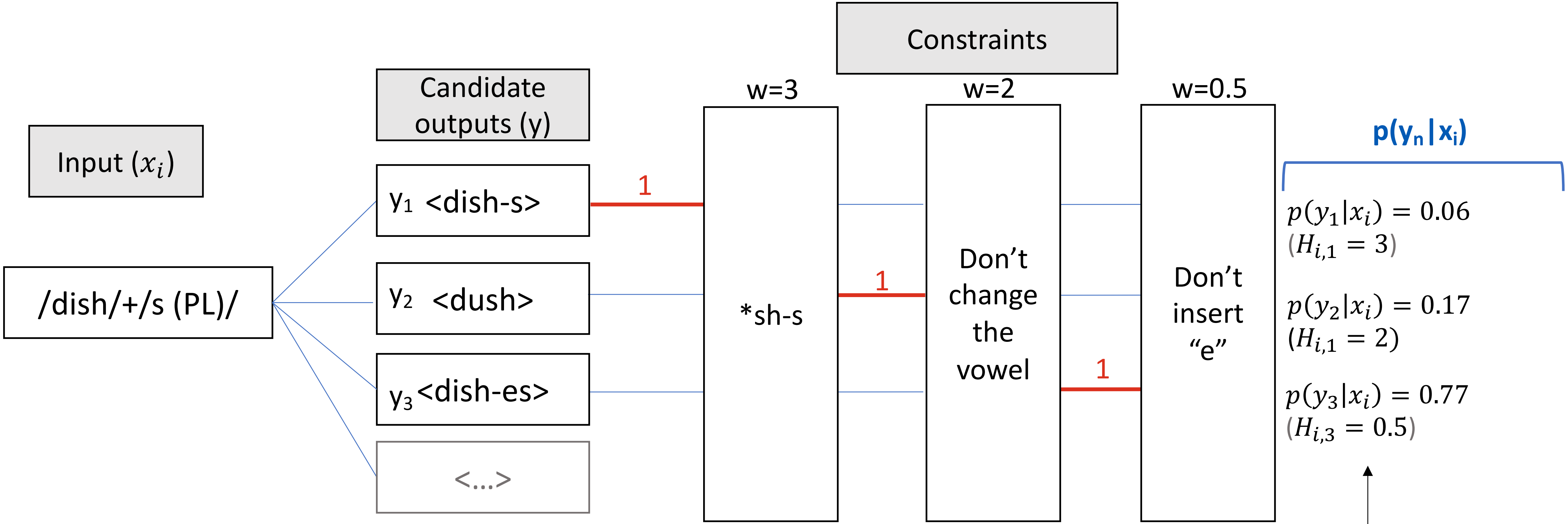


Constraints are weighted, and each candidate receives a penalty score that is the **weighted sum** of all its constraint violations.

1 Phonological grammar



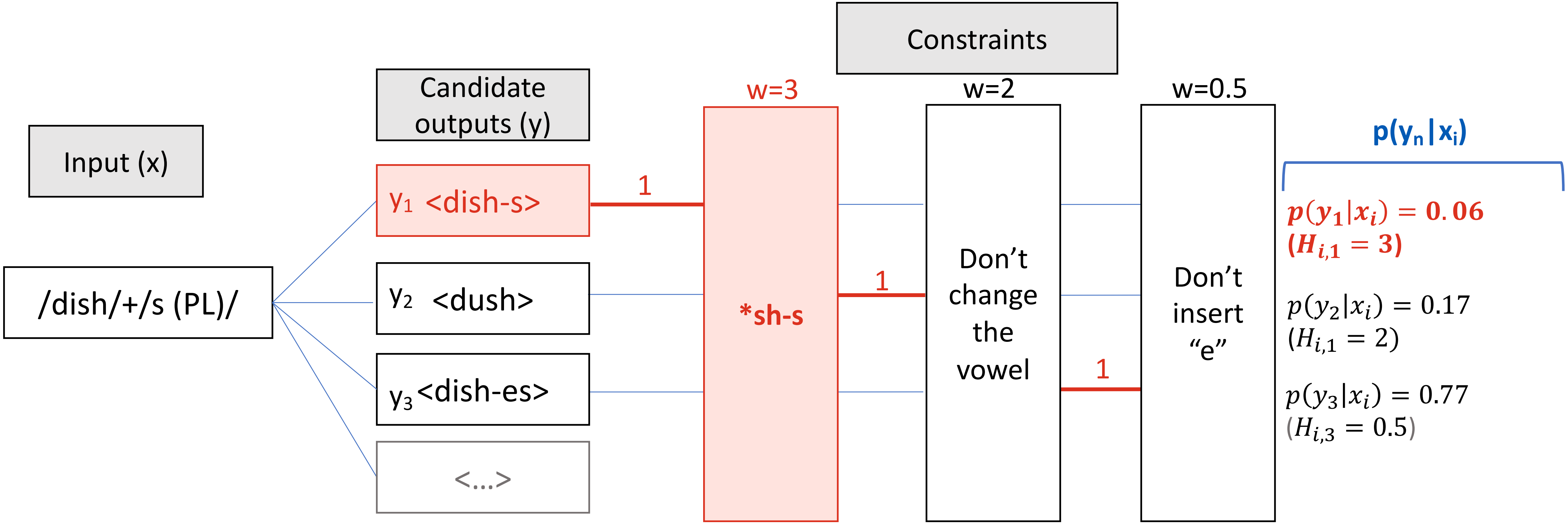
1 Phonological grammar



Convert penalties to probabilities

$$p(y_n|x_i) = \frac{e^{-H_{i,n}}}{Z} \qquad Z = \sum_n \boldsymbol{e^{-H_{i,n}}} = e^{-3} + e^{-2} + e^{-0.5}$$

1 Phonological grammar

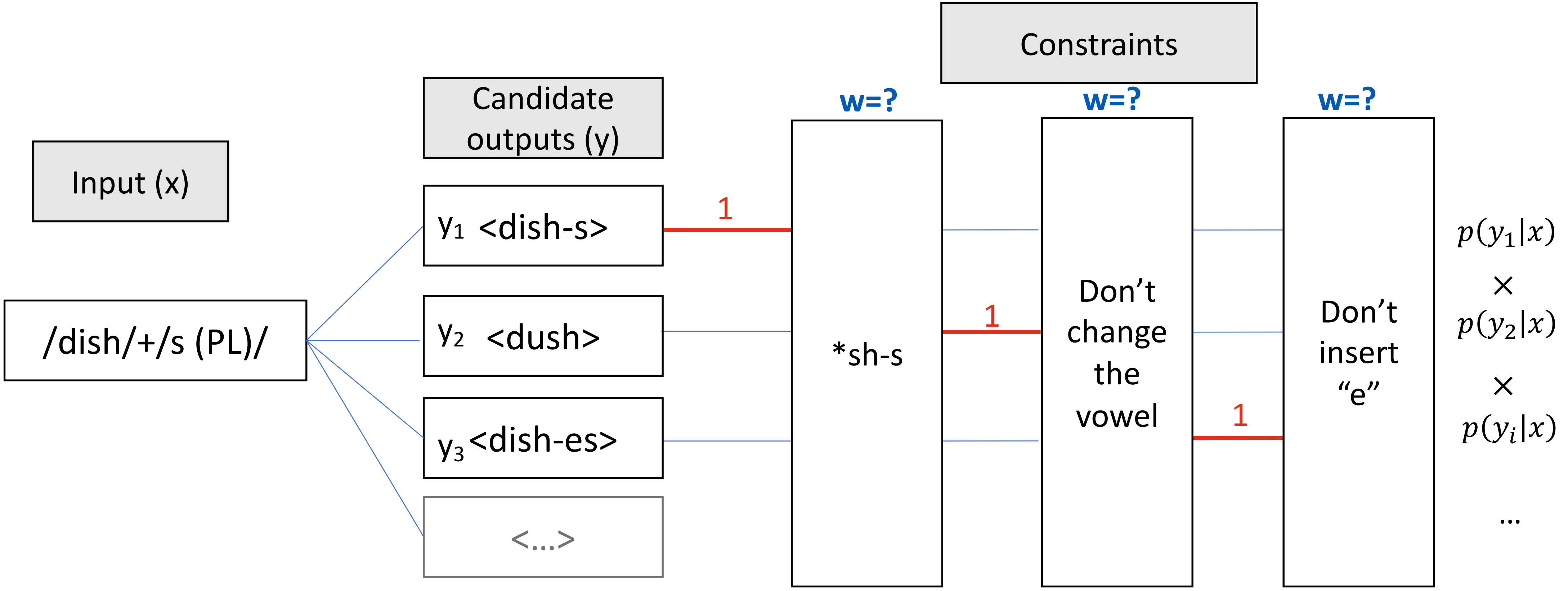


Takeaway: If a candidate output violates a highly weighted constraint, it will receive low probability.

Convert penalties to probabilities

$$p(y_n|x_i) = \frac{e^{-H_{i,n}}}{Z} \qquad Z = \sum_n \boldsymbol{e^{-H_{i,n}}} = e^{-3} + e^{-2} + e^{-0.5}$$

1 Phonological grammar

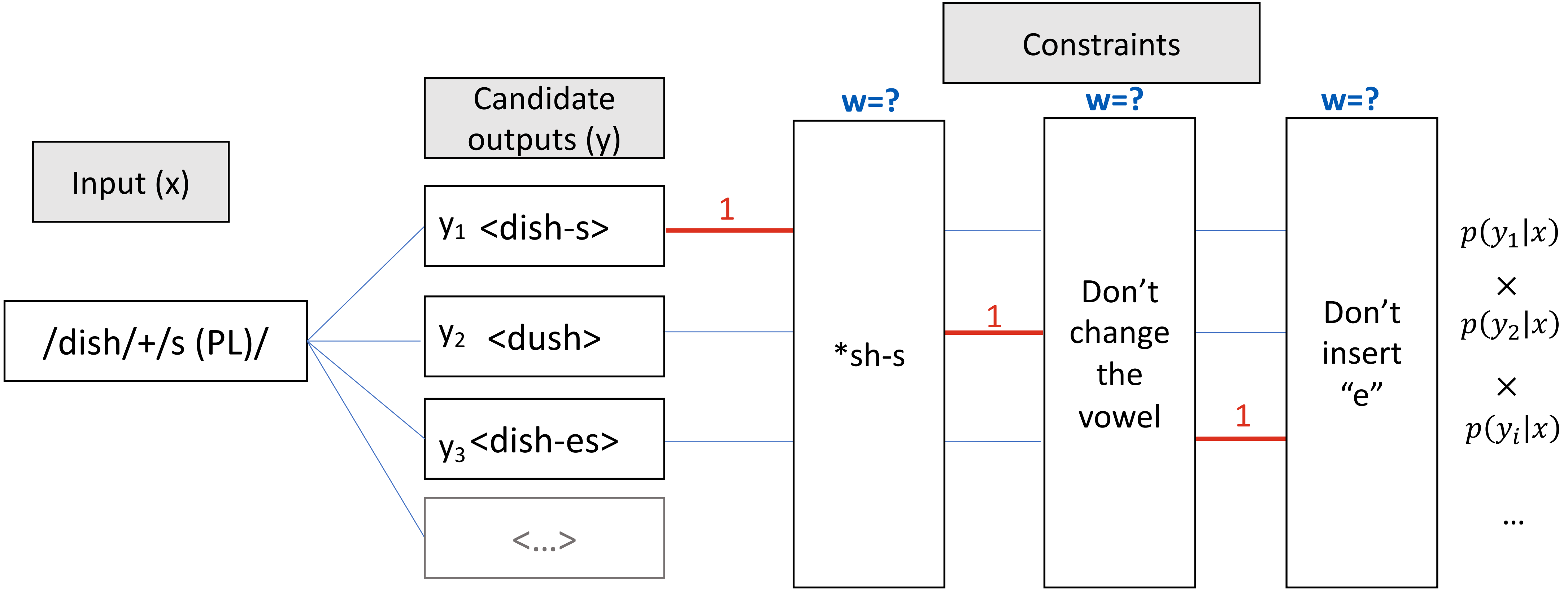


How are weights learned? by maximizing an **objective function** with gradient-based optimization

$$p(y_1|x)p(y_2|x)\dots p(y_n|x)$$

(Goldwater & Johnson, 2003; Lafferty et al., 2001; McCallum, 2003)

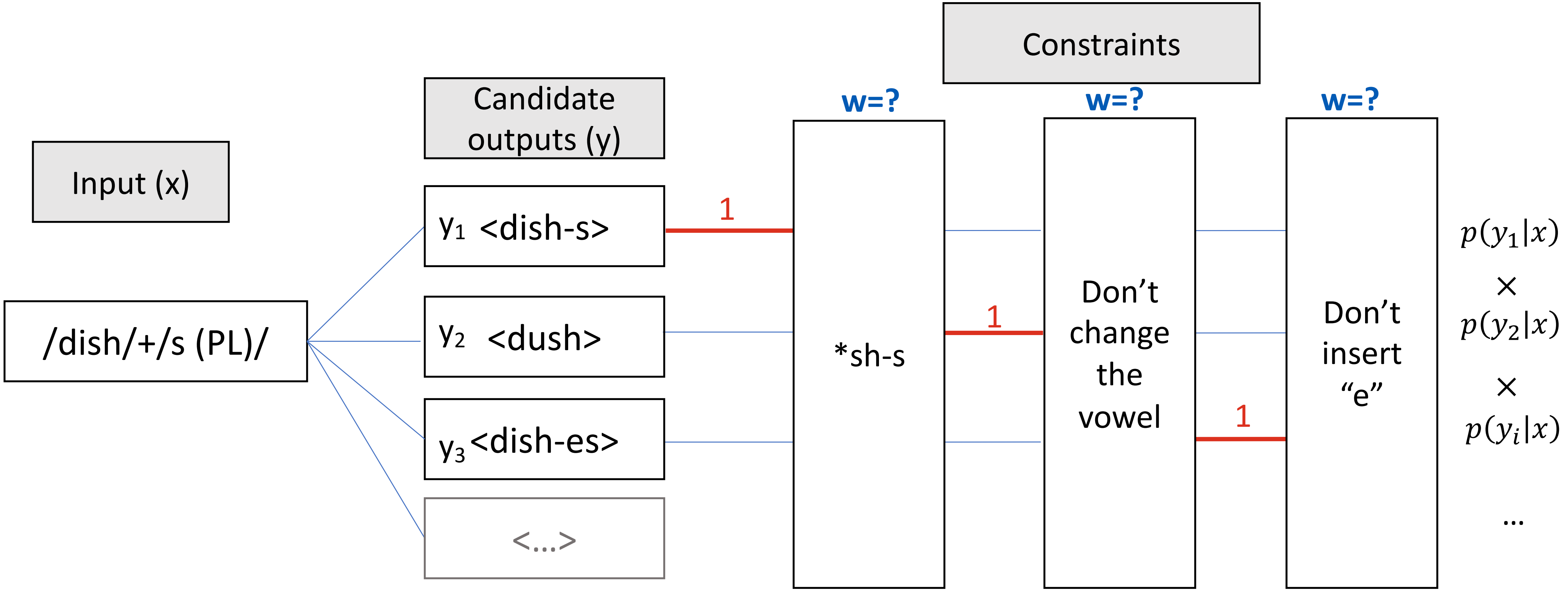
1 Phonological grammar



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

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

1 Phonological grammar



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

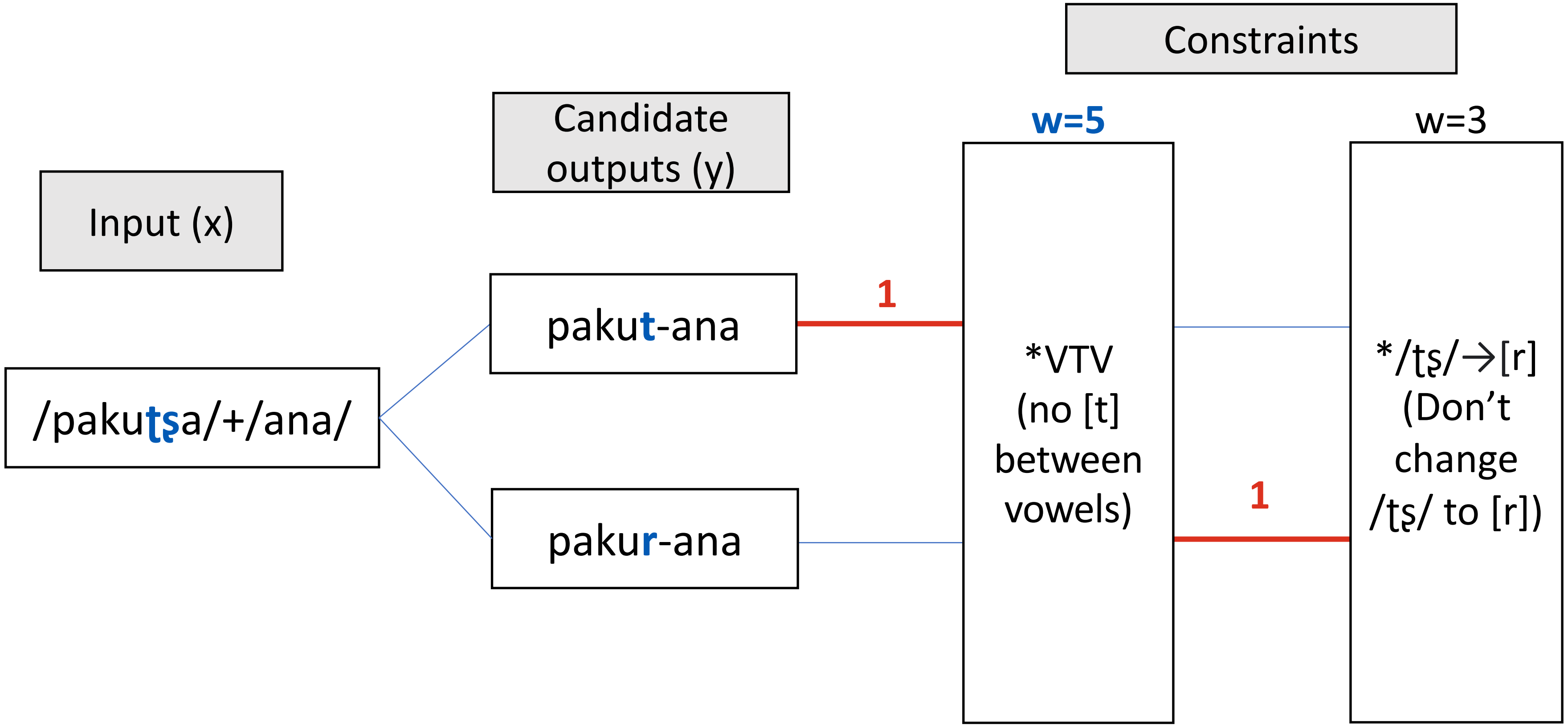
1 Phonological grammar

Now let’s apply this to Malagasy!

1 Phonological grammar: constraints

- ***VTV: no voiceless stops between vowels (p,t,k,ʈʂ)**
  - *[+syllabic][-continuant,-voice][+syllabic]
- Other
  - */a/→[b]: an input /a/ should **not** become [b] in the output (*MAP; Zuraw 2010; 2013)
    - e.g. vukiʈʂa→ vukit-ana violates */ʈʂ/→[t]
  - *ʈʂ]V, *k]V, *n]V (Pater 2007; Chong 2020).
  - *r...r: penalizes sequences of r…r

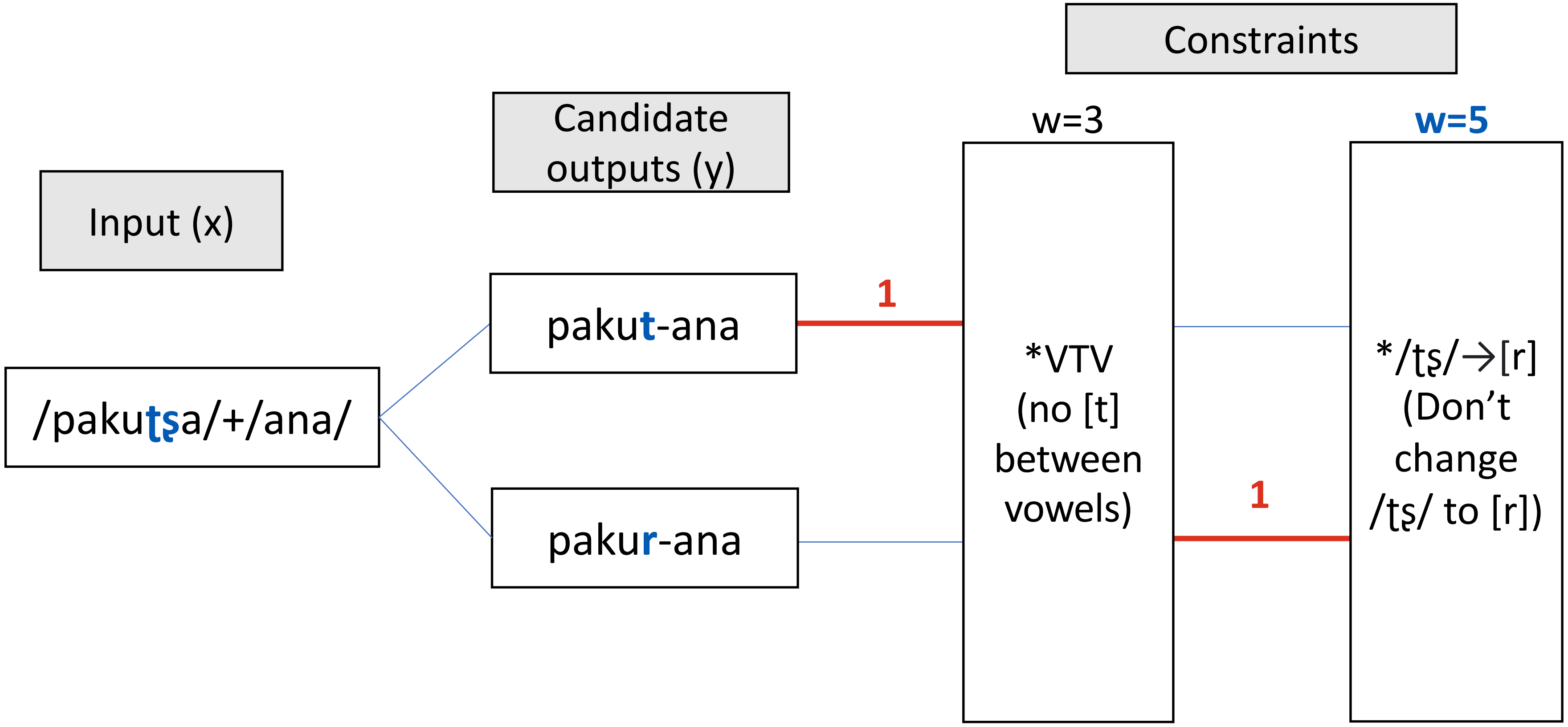
A Malagasy example (simplified)



If w(*VTV) > w(*/ʈʂ/⟶[r]), the grammar will prefer [pakur-ana]

If w(*/ʈʂ/⟶[r]) > w(*VTV), the grammar will prefer [paku**t**-ana]

A Malagasy example (simplified)



If w(*VTV) > w(*/ʈʂ/→[r]), the grammar will prefer [pakur-ana]

If w(*/ʈʂ/→[r]) > 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)

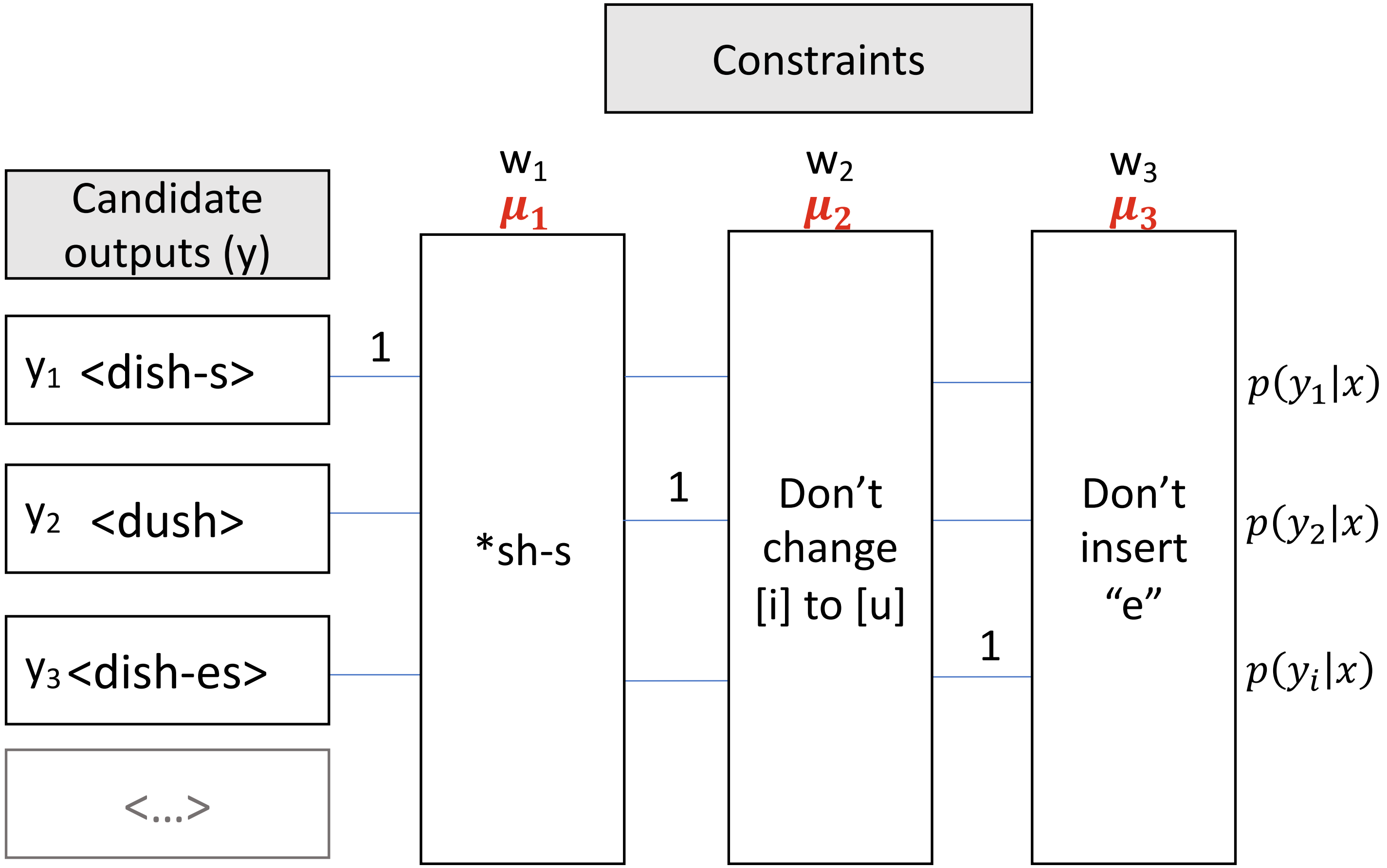
- Functionally equivalent to L2 regularization

The intuition: 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

Implementing a **Gaussian prior**

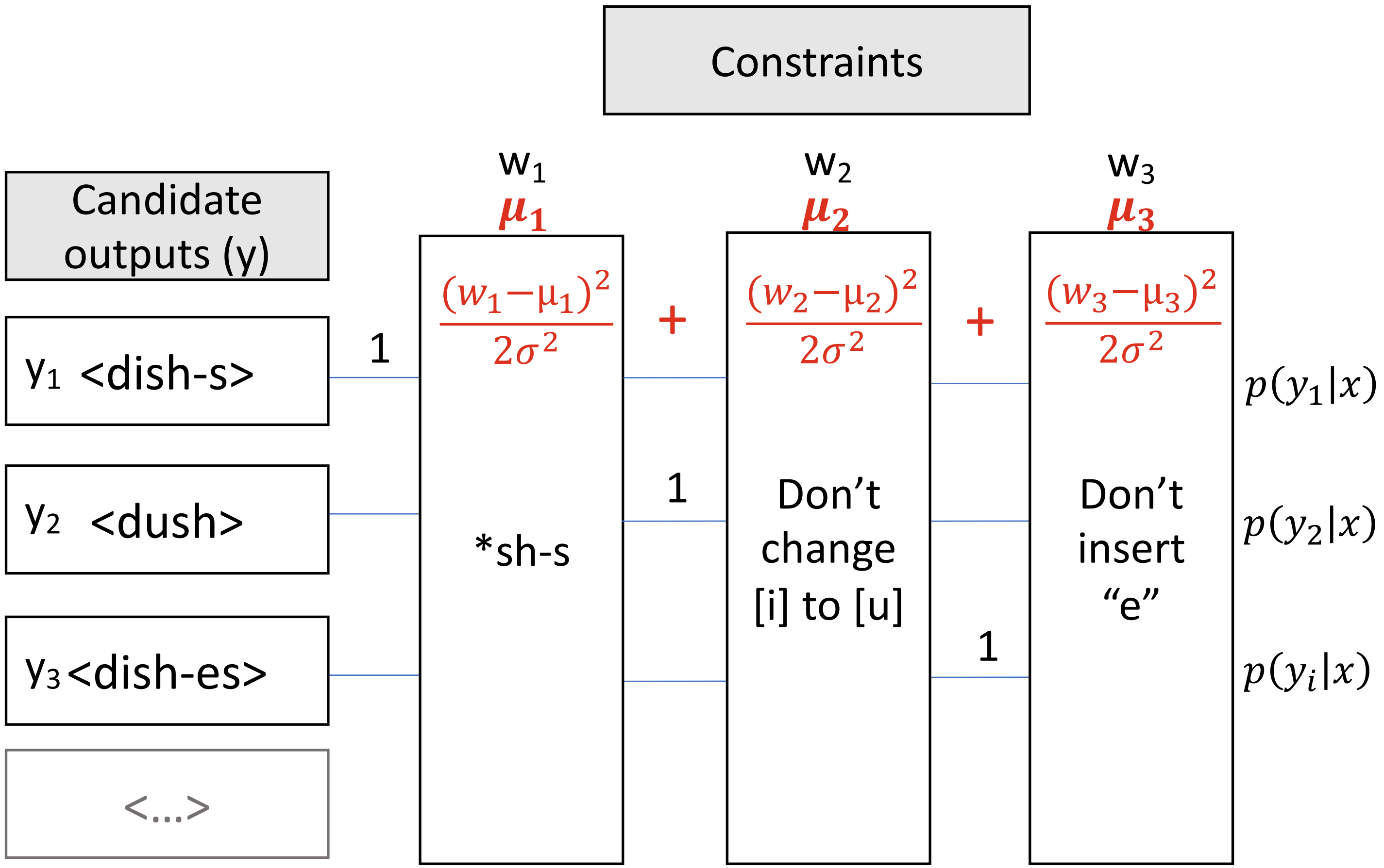


Old objective function

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

2 Learning biases

Implementing a **Gaussian prior**



New objective function

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

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

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

- σ set to 1.0 for all constraints

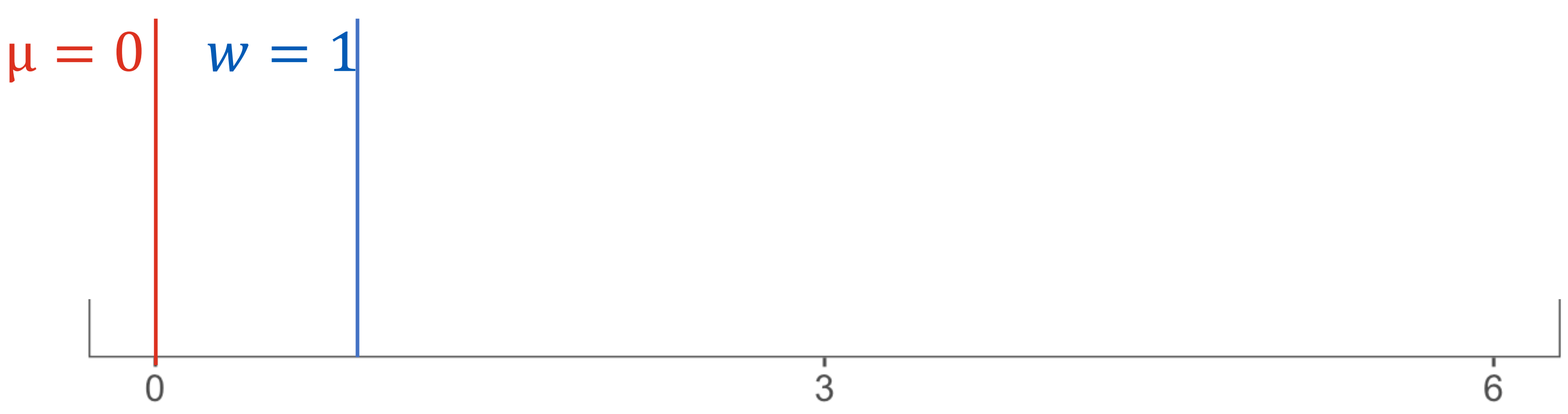
Low μ = low preferred weight

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

w = constraint weight

μ = “preferred” weight

$$(1 - 0)^2 = 1$$

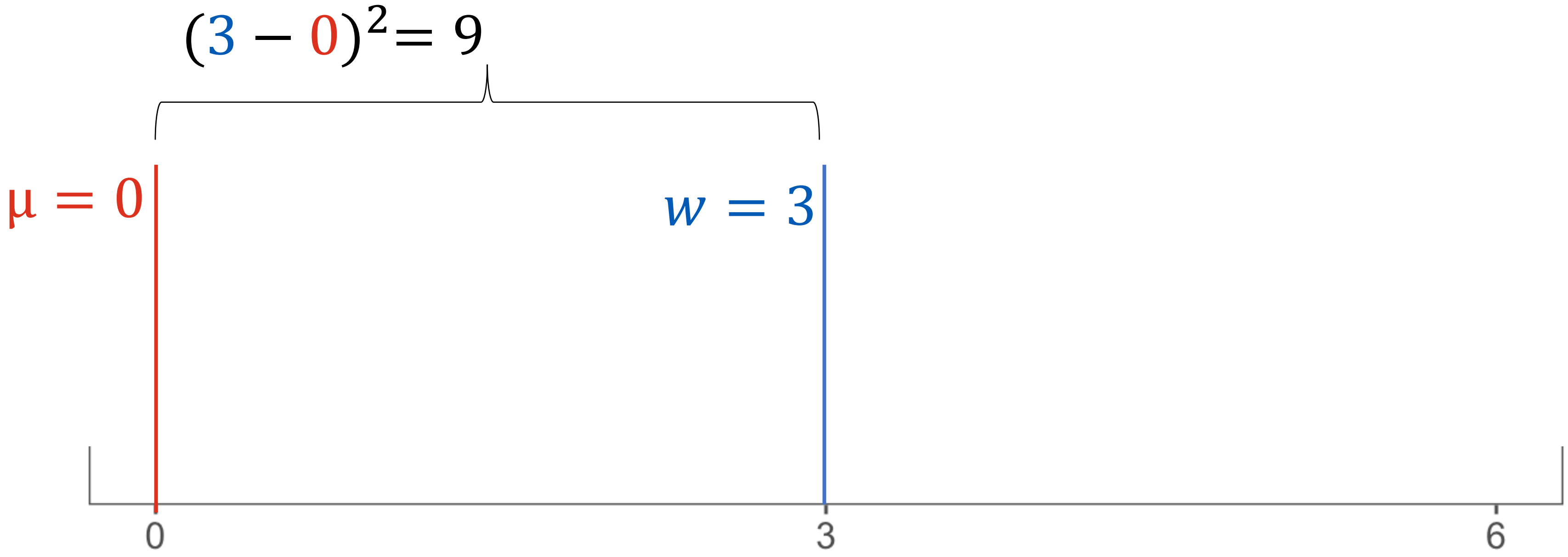


Low μ = low preferred weight

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

w = constraint weight

μ = “preferred” weight



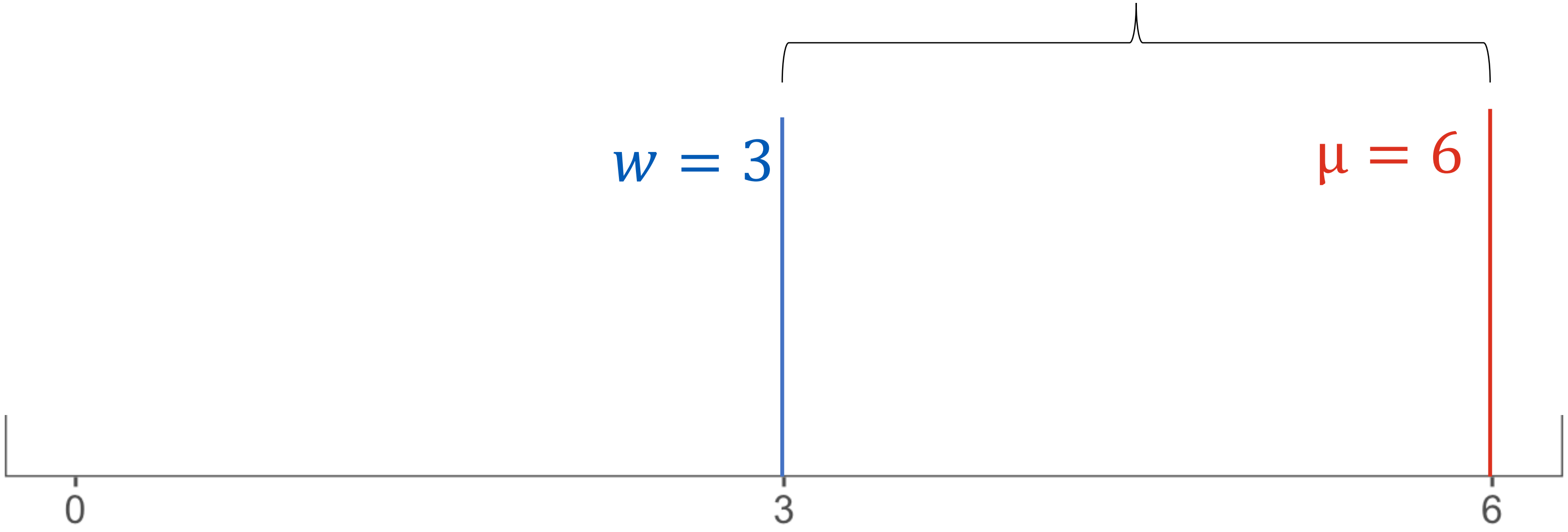
high μ = high preferred weight

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

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

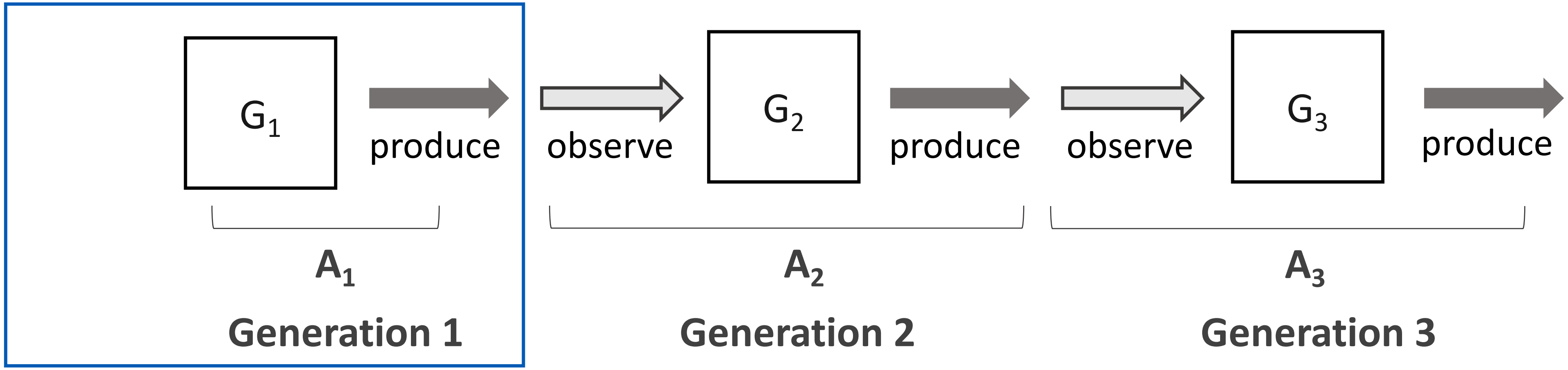
Generalization: dispreference for /p, t, ʈʂ, k/ between vowels

Model: μ(*VTV)=1, otherwise μ=0

Elements in a model of reanalysis

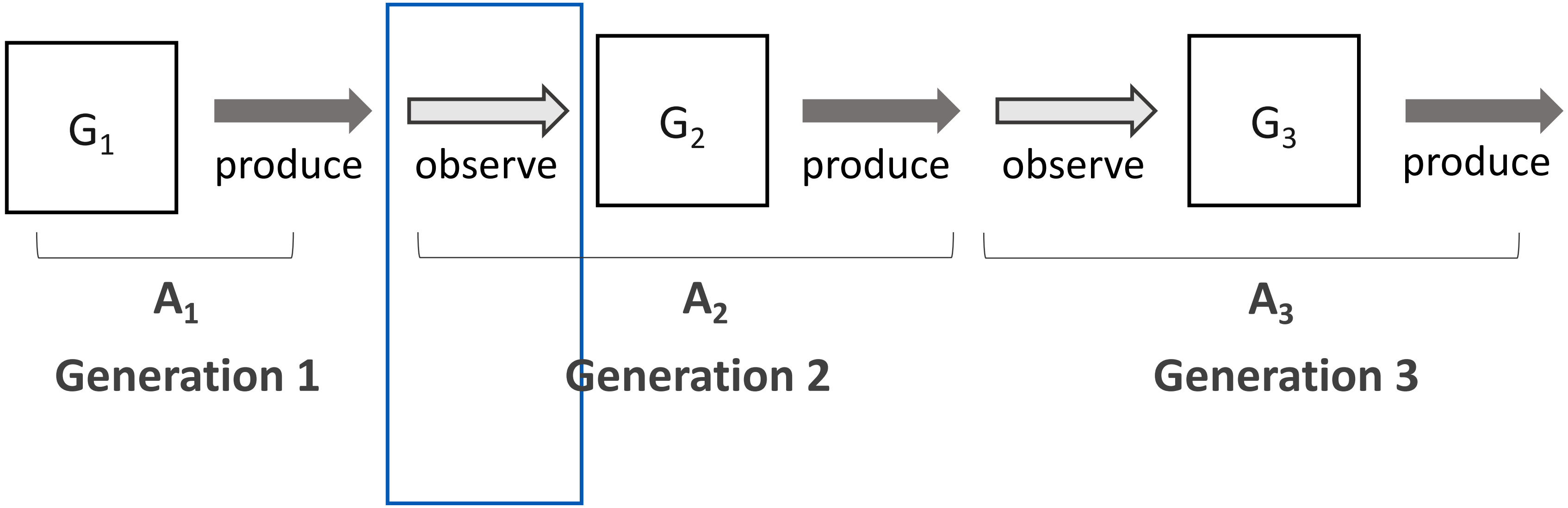
1. A probabilistic phonological grammar
2. Ability to incorporate learning biases
3. **Simulate generations of change**

3 Iterated learning



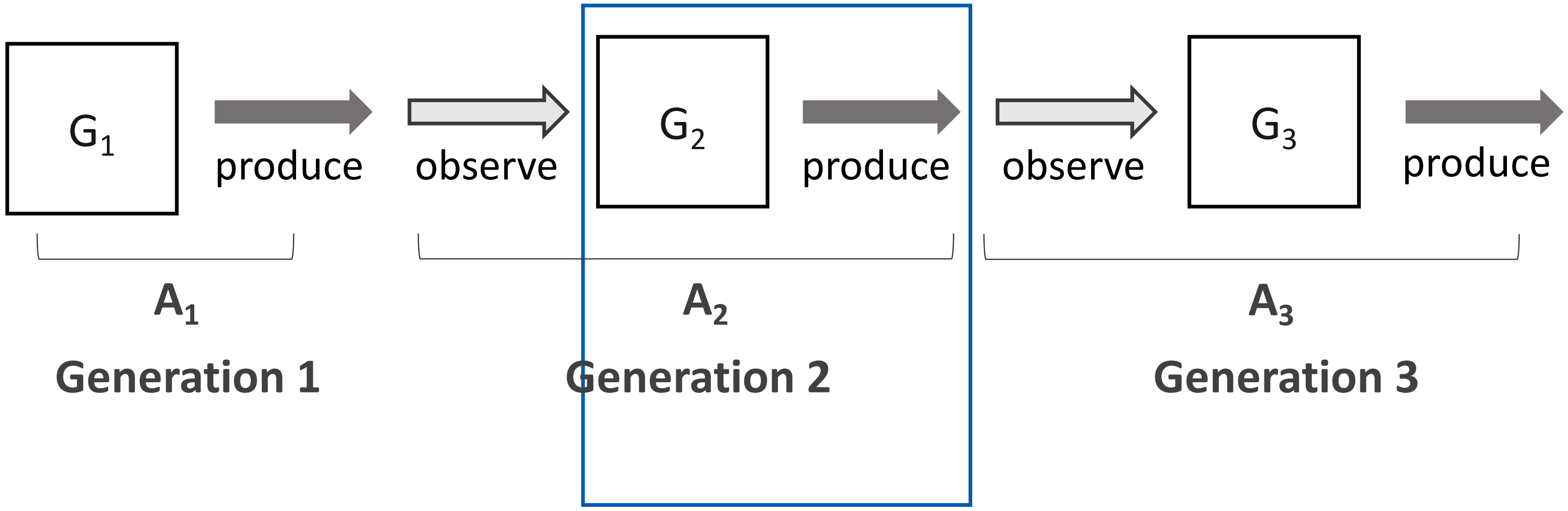
Agent A_1 produces words based on their knowledge of the grammar (G_1), which A_2 observes

3 Iterated learning



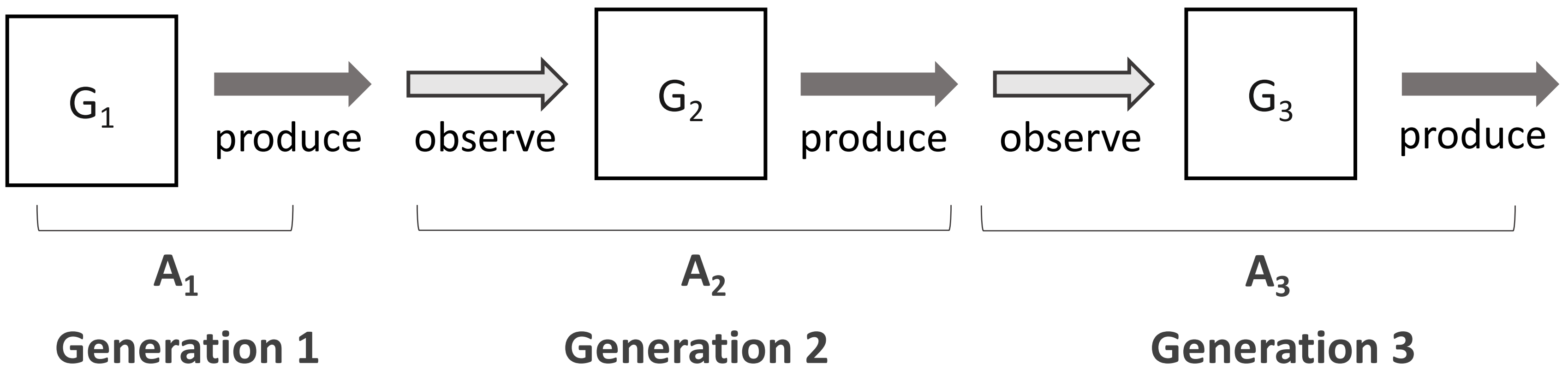
"Bottleneck": A_2 **forgets** some proportion of words.

3 Iterated learning



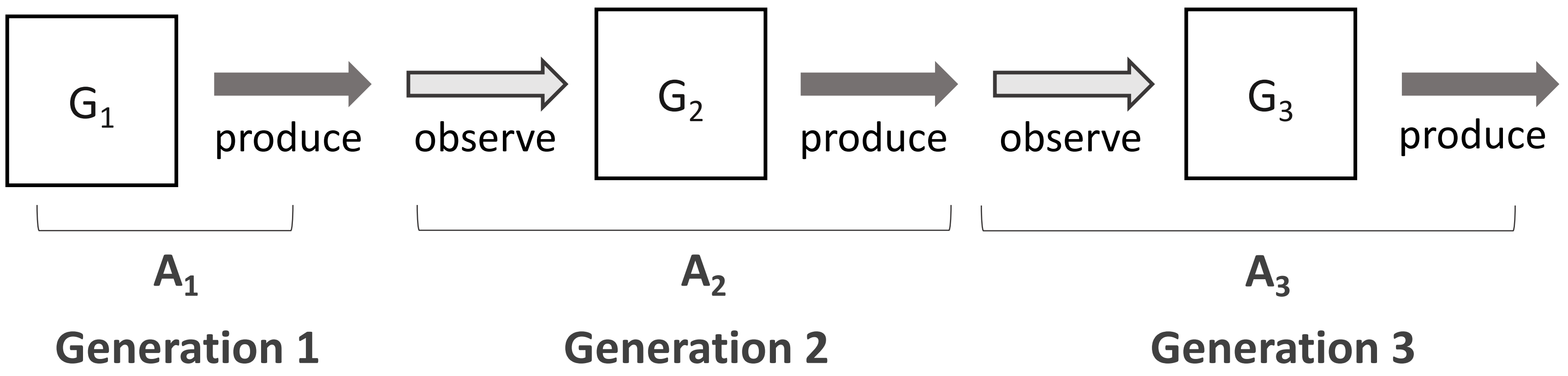
Agent A_2 learns a grammar (G_2) based on the remembered words and produces it.

3 Iterated learning



- Used to simulate change/evolution (for a review: Kirby, Griffiths, & Smith 2014)
- Few applications to detailed patterns of language change.
  - See: Ito & Feldman (2022) on accent change in Sino-Korean.
  - Other work on iterated learning: de Boer 2000; Kirby 2001; Brighton 2002, etc.

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 20 runs.

Elements in a model of reanalysis

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

Results

1

Intro

2

Malagasy reanalysis

3

Model

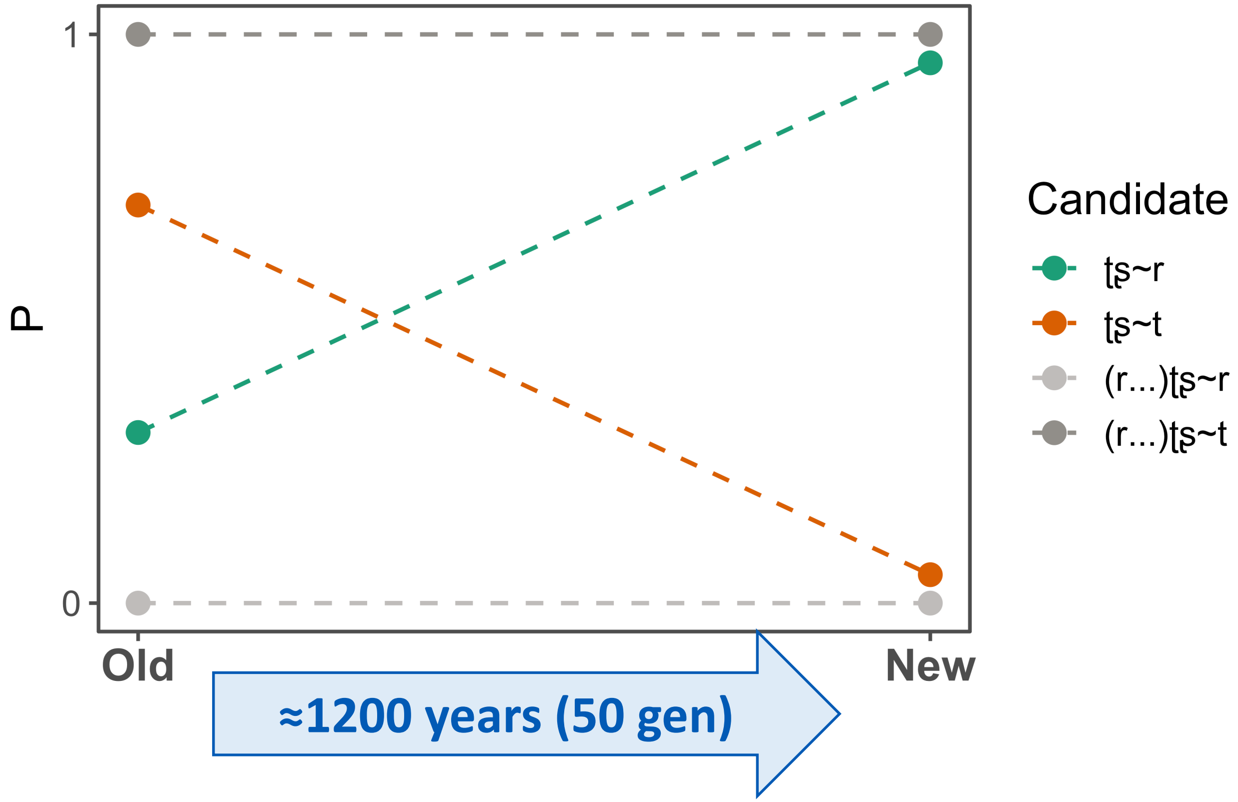
4

Results

Reviewing the Malagasy data: all stem types

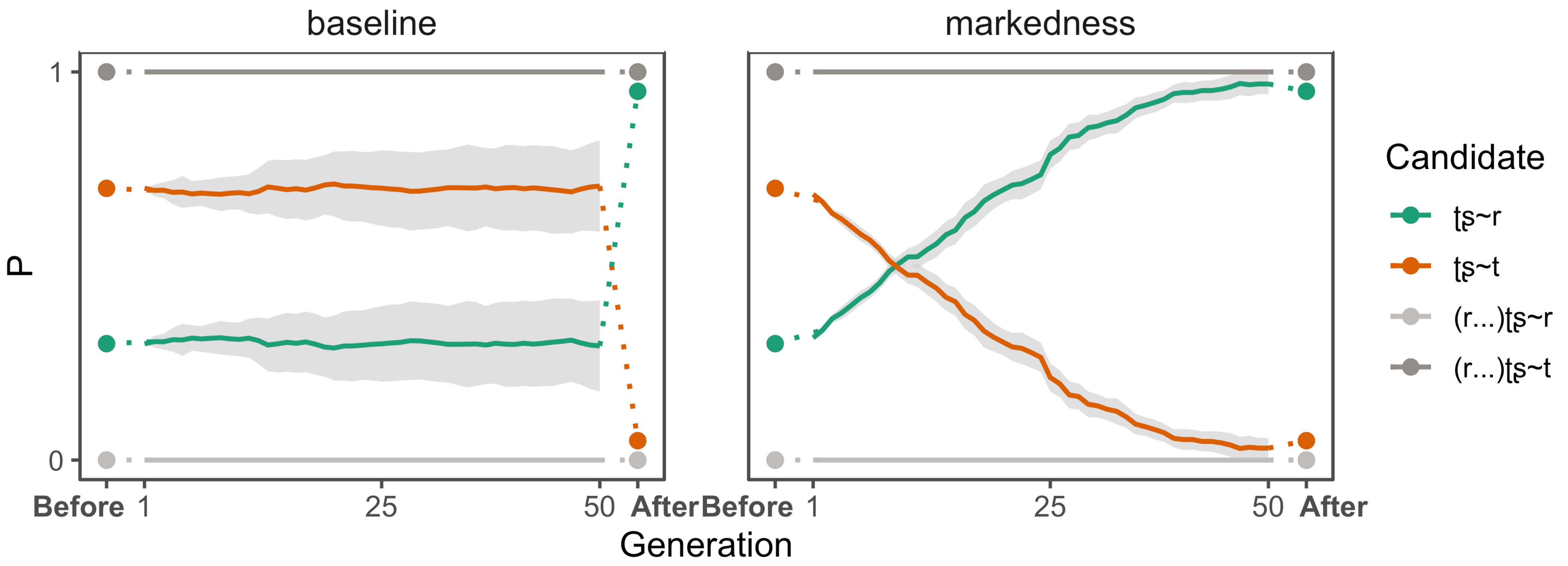
	old Malagasy	new Malagasy
ka-stems	prefer [h]	prefer [h]
na-stems	prefer [n]	prefer [n]
ʈʂa-stems	prefer [t] avoid r…r	**prefer [r]** avoid r…r

Reviewing the Malagasy data: ʈʂa stems



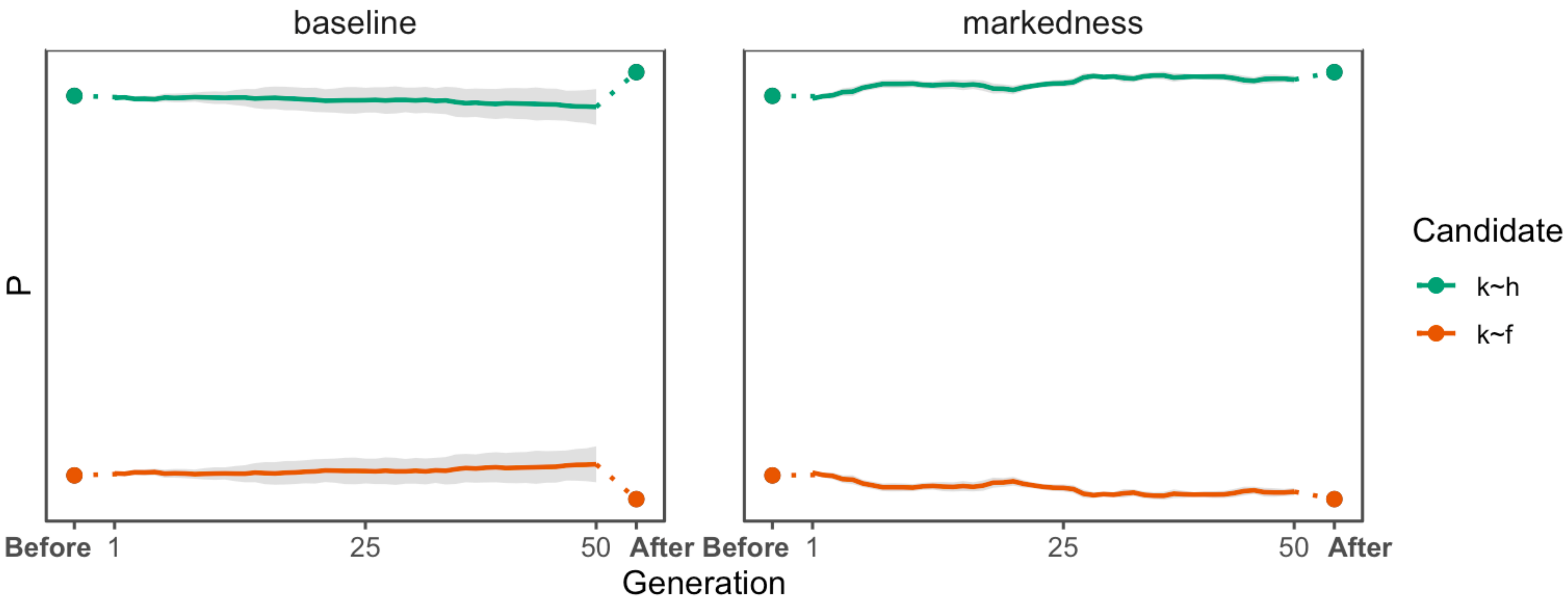
Markedness bias improves model predictions

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



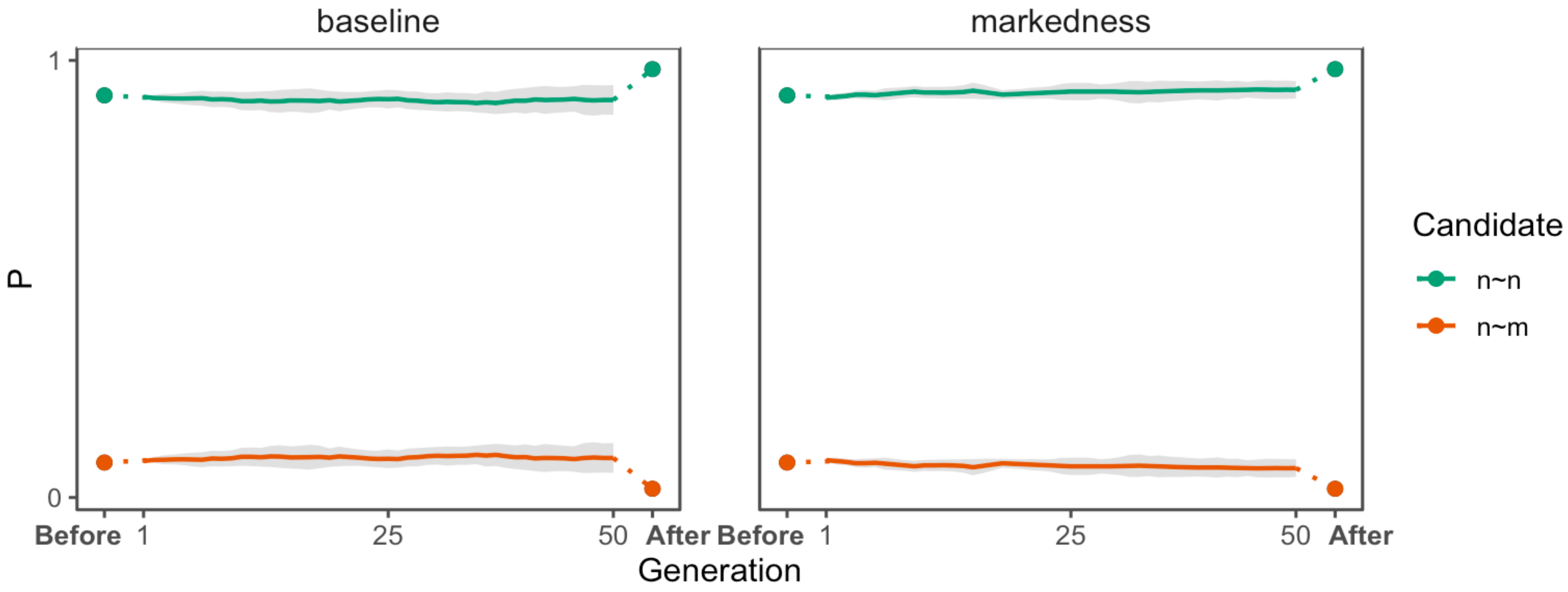
Markedness model performs well on **all** weak stems

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



Markedness model performs well on all weak stems

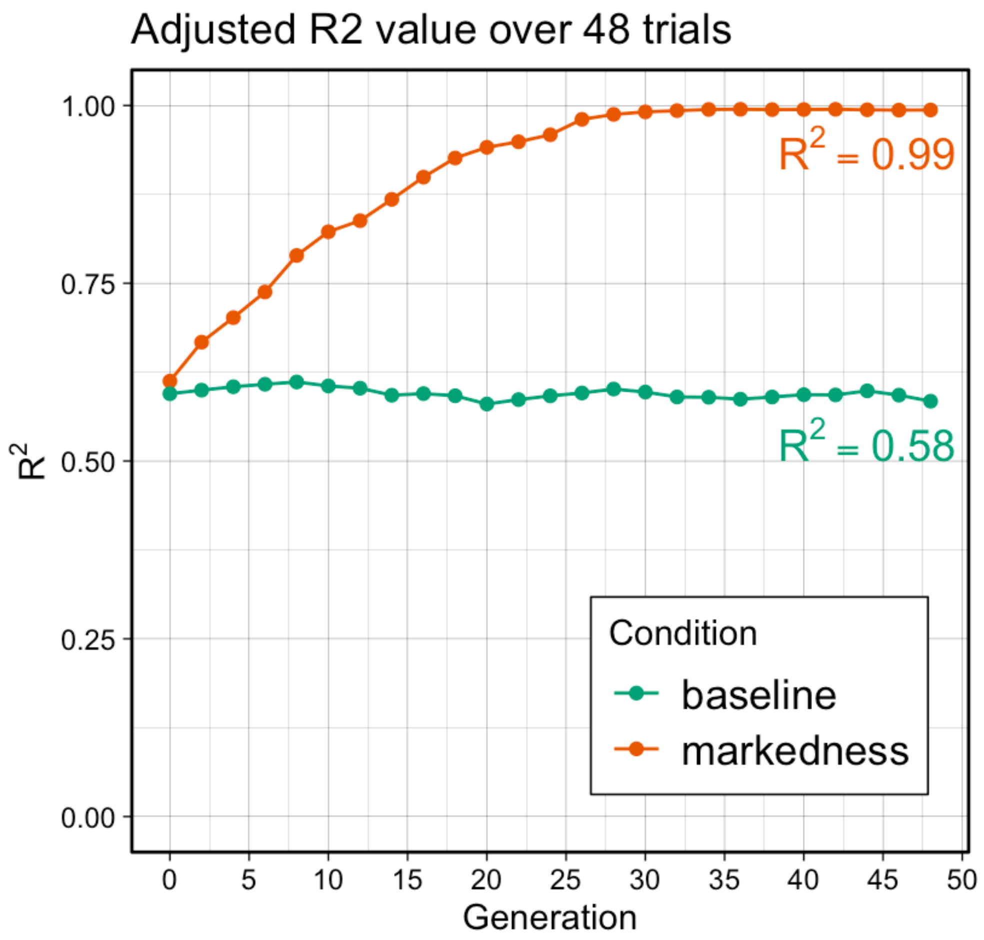
Figure: Predicted proportion of suffixed form outputs for **na** weak stems (forget rate = 0.2)



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 + markedness bias
   - ➢In ʈʂa words, t→r is motivated giving *VTV a bias towards higher weight

Summing up

1. Show that reanalysis in Malagasy can be explained as statistical learning + markedness bias
   - ➢In ʈʂa words, t→r is motivated giving *VTV a bias towards higher weight
2. Outline a model for incorporating markedness effects into reanalysis.
   - ➢Maximum Entropy HG with Gaussian prior + iterated learning.

Summing up

1. Show that reanalysis in Malagasy can be explained by the interaction of **statistical learning** and a **markedness bias**
   - ➢Reanalysis of t→r is motivated giving *VTV a bias towards higher weight
2. Outline a model for incorporating markedness effects into reanalysis.
   - ➢Maximum Entropy HG with Gaussian prior + iterated learning.
3. Demonstrate how computational 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 computational techniques are particularly helpful!

What’s next?

- More evidence from Māori, Samoan, etc…
- What restricts the range of possible markedness effects?
  - Universal markedness
  - Learned from language-specific trends, already present in the lexicon- probably true for Malagasy!
- Expanding the empirical domain:
  - Artificial Grammar Learning studies (in progress)?
  - L2 acquisition, heritage languages

Thank you!

Thank you to…

My consultant Vololona Rasolofoson for her time and contribution.

Bruce Hayes, Kie Zuraw, Claire Moore-Cantwell, Emily Grabowski, members of the UCLA Phonology seminar, other colleagues and friends for their insightful 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*.
- 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}.
- 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.

References II

- 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.
- 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

- 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.

Markedness bias and phonotactics

- Maximum Entropy models, differing in constraint set
  - Assigns **penalty scores** that can then be used as constraint violations in the reanalysis model
  - (UCLA Phonotactic Learner; Hayes & Wilson, 2008)
- "Natural class model": generalizes to 'natural classes' of sounds
  - Ex: *[-syllabic] applies to all non-vowel sounds.
- "N-gram model": segment bigrams and trigrams
  - Ex: *p, *t, *k, *m, *n .....
- Trained on 3573 words
  - No complex words (e.g. compounds, suffixed words, etc.)
  - No reduplicated forms (e.g. [paka-paka])

Markedness bias and phonotactics

- Tested on candidate outputs of the **reanalysis model**
  - Generates penalty values that are then input to the reanalysis model

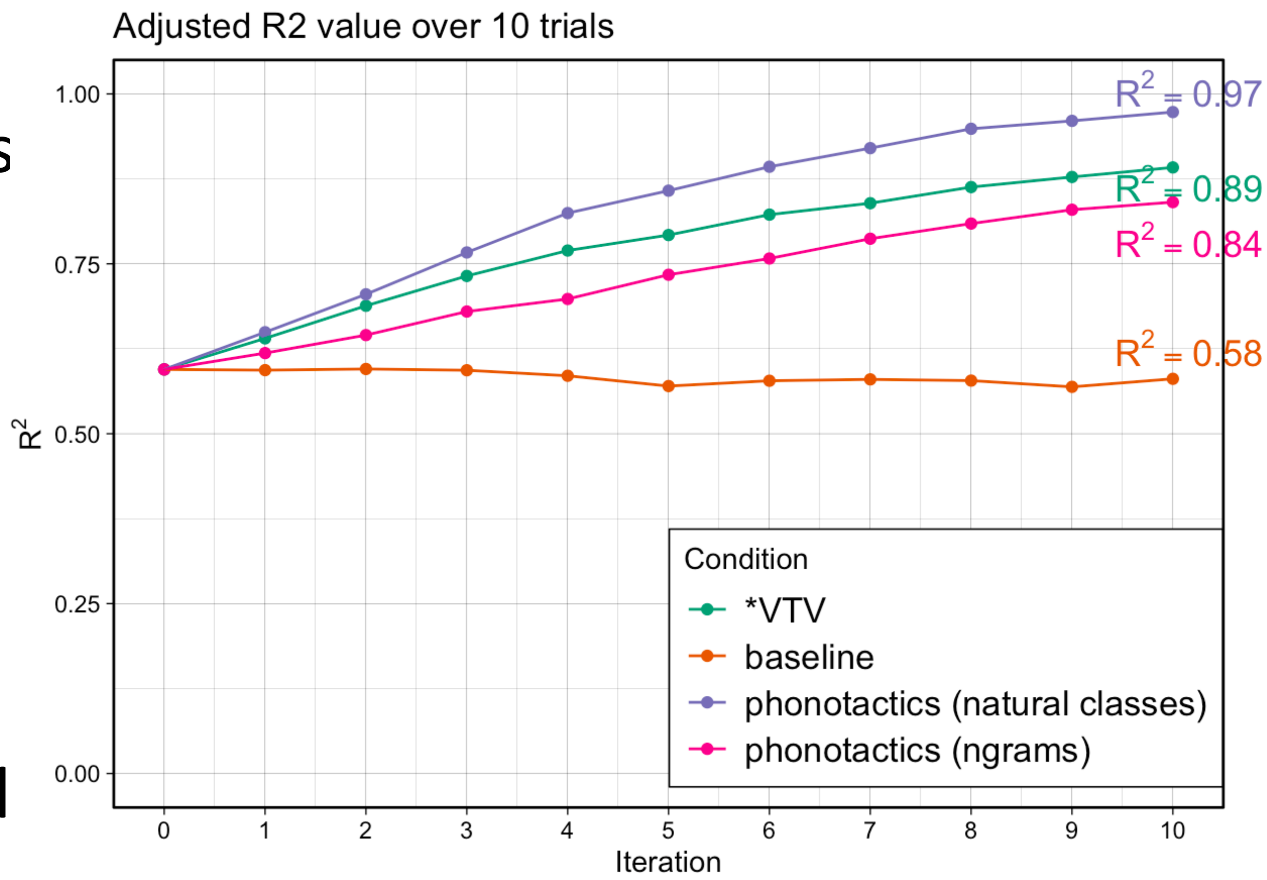
Ex: penalty assigned by natural class model

input	output	H (penalty)
vukiʈʂa+ana	vukir-ana	12.25
	vukit-ana	13.34
	vukiʈʂ-ana	13.76

Input into reanalysis model as constraint violations of a constraint ‘ObeyPhonotactics’

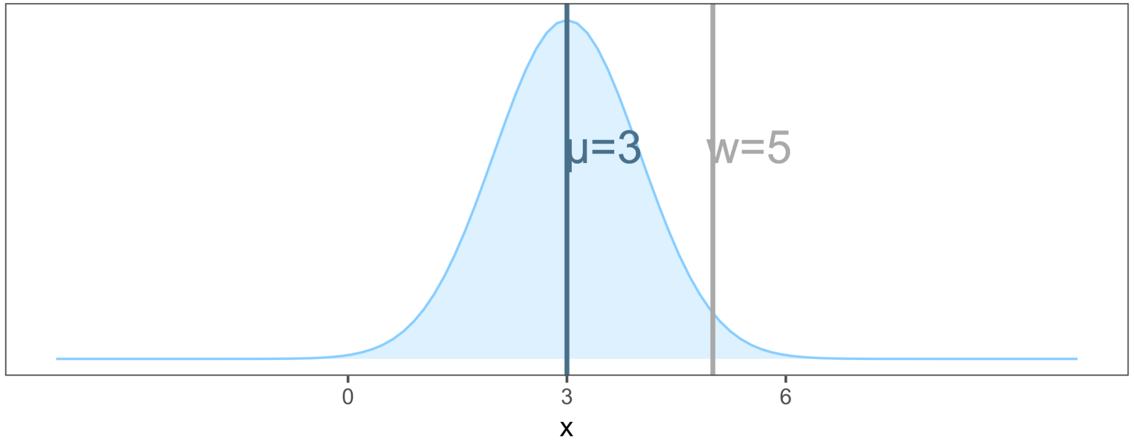
Markedness bias and phonotactics

- The phonotactic model that generalizes to natural classes performs the best
- N-gram model performs the worst
- All three models outperform the baseline
- Subsequent comparisons use results from the "natural class" model

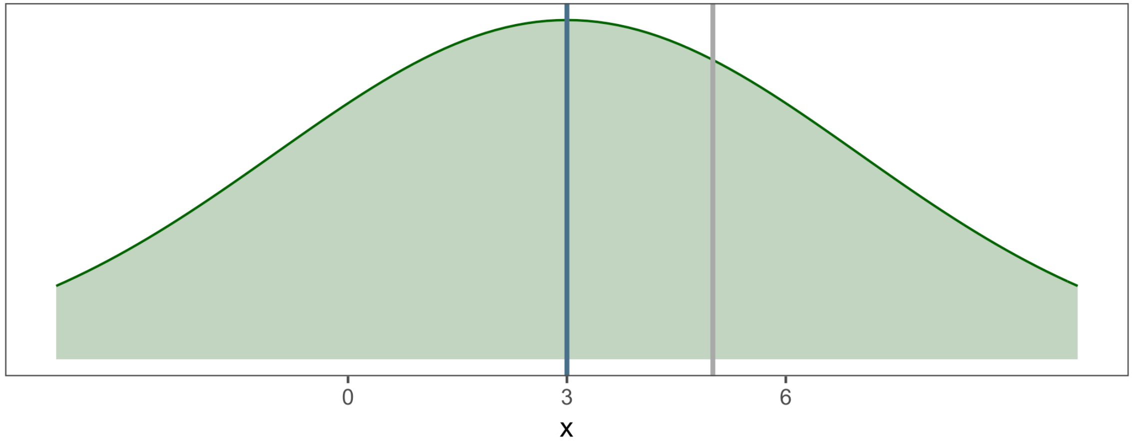


2 Learning biases: sigma

σ=1



σ=4



$$\sum_{m=1}^{M} \frac{(w_m - 3)^2}{2\boldsymbol{\sigma}^2}$$

- σ = standard deviation

Can in principle be varied to implement bias

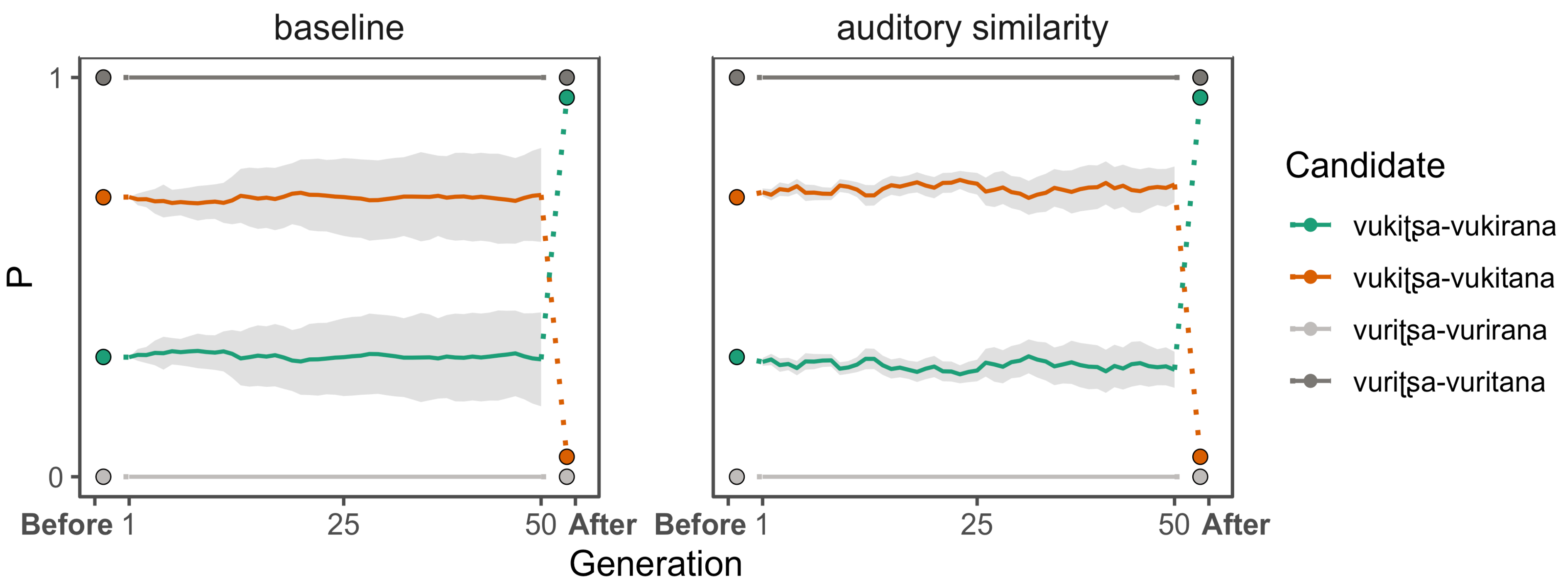
Perceptual similarity bias

- 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

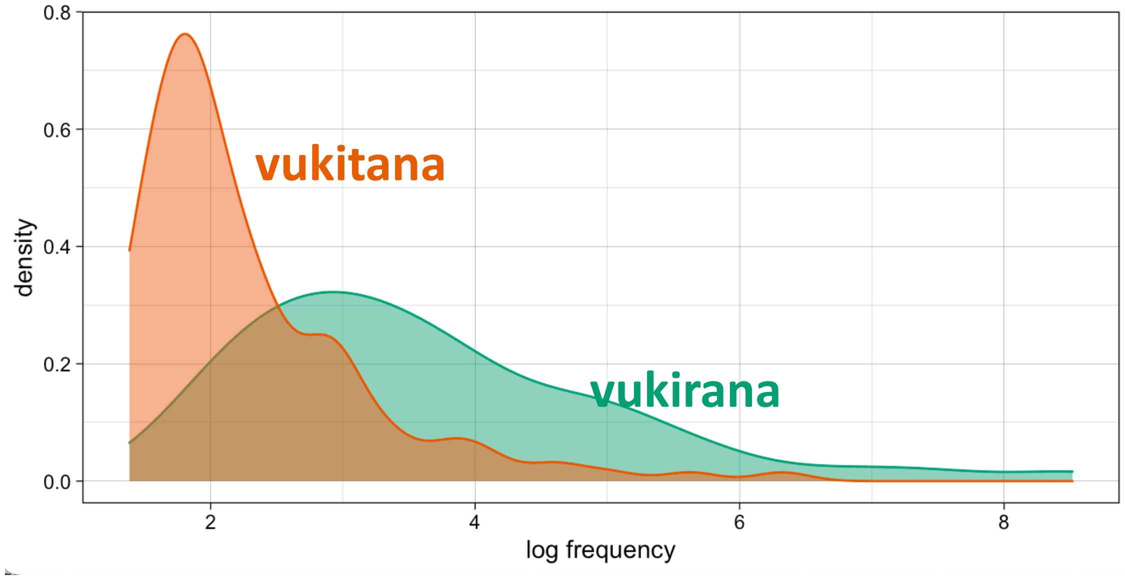


Token frequency

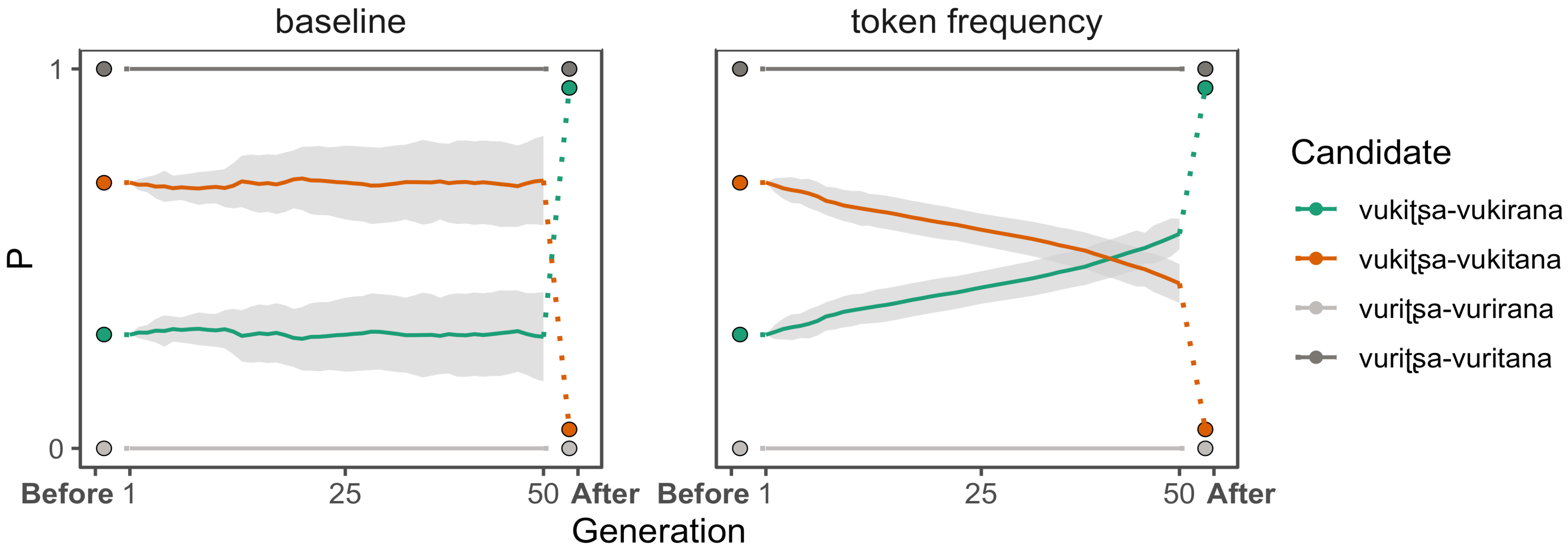
- In phonology, type frequency is a better predictor of phonological patterns (Bybee 1995; Bybee, 2001; Pierrehumbert 2001; Albright & Hayes, 2003)
- However, words with high token frequency:
  - Are more likely to be learned/passed down through generations
  - And may end up influencing a pattern (Albright, 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



UR analysis

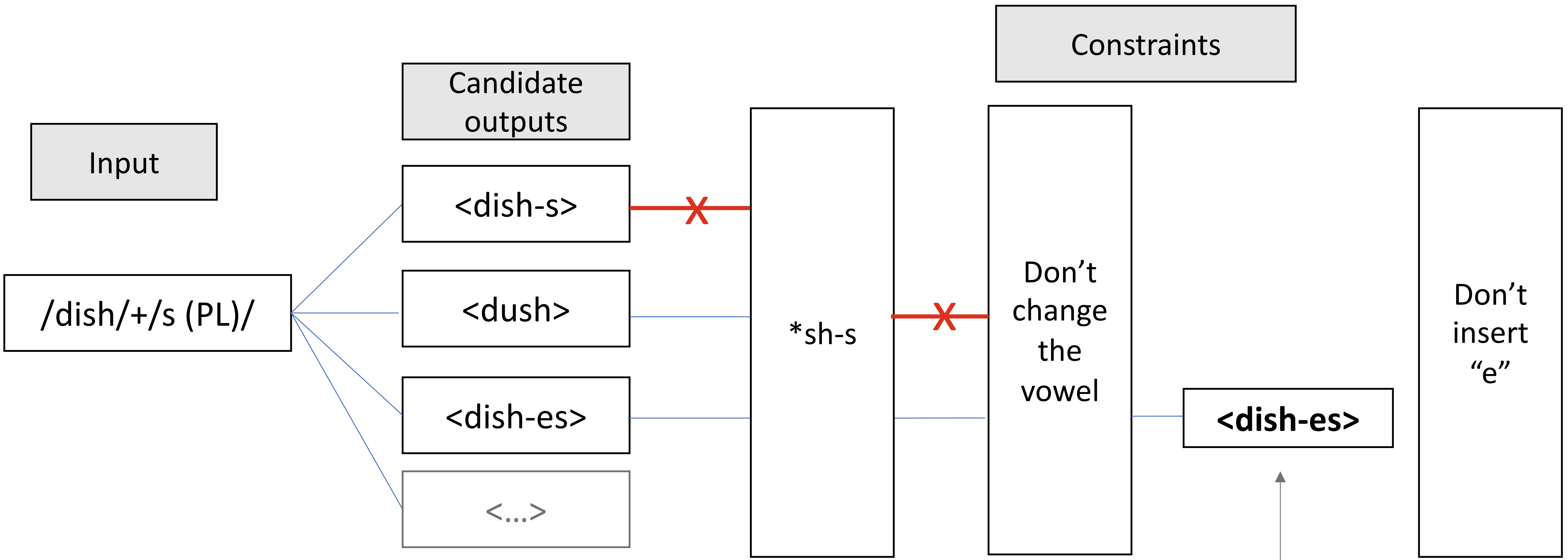
/vuki**c**+ana/∼[ˈvukitsa]	*ts]V	*t]V	*r]V	IO-FAITH	OO-FAITH
a. vukiʈʂana	*				
b. vukitana		*			*
c. vukirana			*		*

Markedness bias

- Makes the correct predictions for **all** weak stems

TYPE	SUFF CONS.	V-STOP-V	EXAMPLE SUFFIXED FORM	TARGET OF REANALYSIS
ʈʂa	**r**		pulirana	✓
	t	*	pulitana	
ka	h		pulihana	✓
	f		pulifana	
na	n	(*)	pulinana	✓
	m	(*)	pulimana	

Categorical phonological grammar (OT)



In a categorical grammar, constraints are ranked, and the winner is selected once competing candidates are ruled out.

Effect of forgetting rates