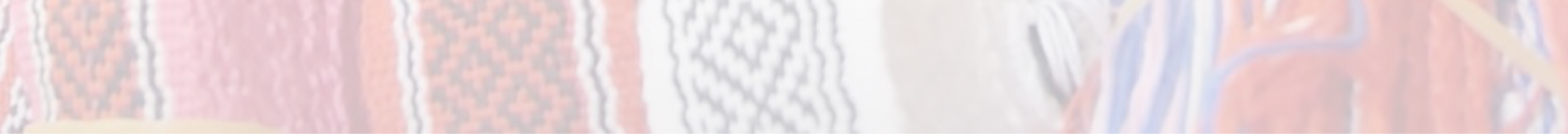


# Prosodic correspondence in Tgdaya Seediq
## insights from corpus and experimental evidence

Jennifer Kuo, Cornell University

Selected Slides

# Research overview

- **Phonological learning.** How do people learn and represent sound patterns?
- **Structure of paradigms.** How do related words influence each other, and how do people encode the relationship between forms of a paradigm?

Corpora

Grabowski & Kuo (2023), Kuo (2023b)

Experimental evidence

Kuo (2023a)

Fieldwork (Seediq, Mam)

Grabowski & Kuo (2023), Elkins & Kuo (2022)

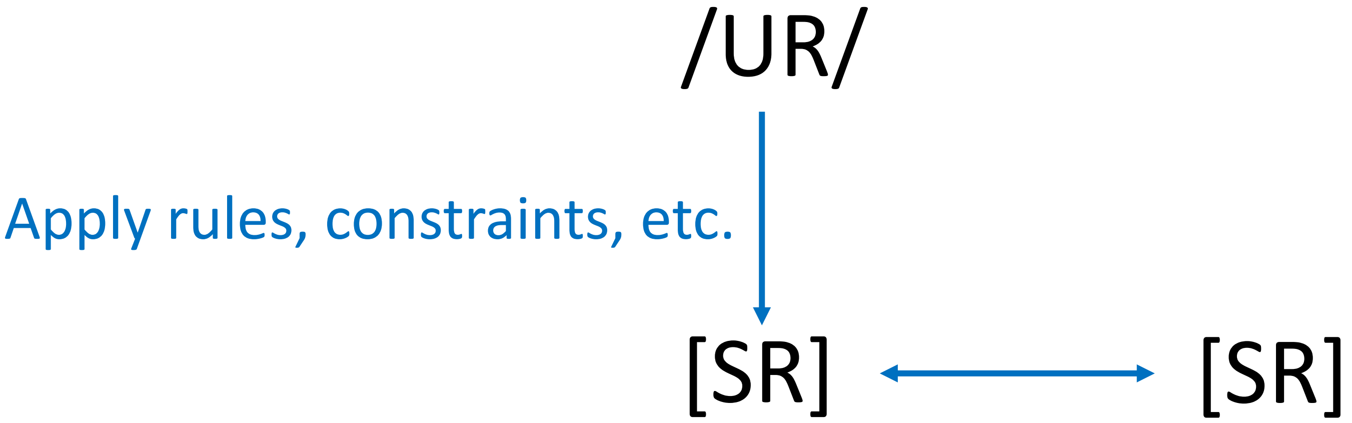
Modeling

Kuo (2020; 2023b)

**Today**: insights about paradigm structure from Tgdaya Seediq

# UR-SR relations

- Typically, models of phonology derive surface representations (SR) from underlying representations (UR)



/UR/

Apply rules, constraints, etc.

[SR]

[SR]

- There is evidence that related surface forms within a paradigm can influence each other, challenging this view.

# Similarity across a paradigm

- Surface forms in a paradigm (i.e. across grammatical contexts) tend to be similar.
  - Example: English past tense

| | | | | | |
|---|---|---|---|---|---|
| **want** | **want**-ed | | sp**ea**k | sp**o**ke | (n=6) |
| **wait** | **wait**-ed | vs. | str**i**ke | str**u**ck | (n=16) |
| **plan** | **plann**-ed | | g**i**ve | g**a**ve | (n=1) |
| … | | | … | | |
| **N=1146 (93%)** | | | | | |

Generalizations from the CELEX database, taken from Albright and Hayes (2003)

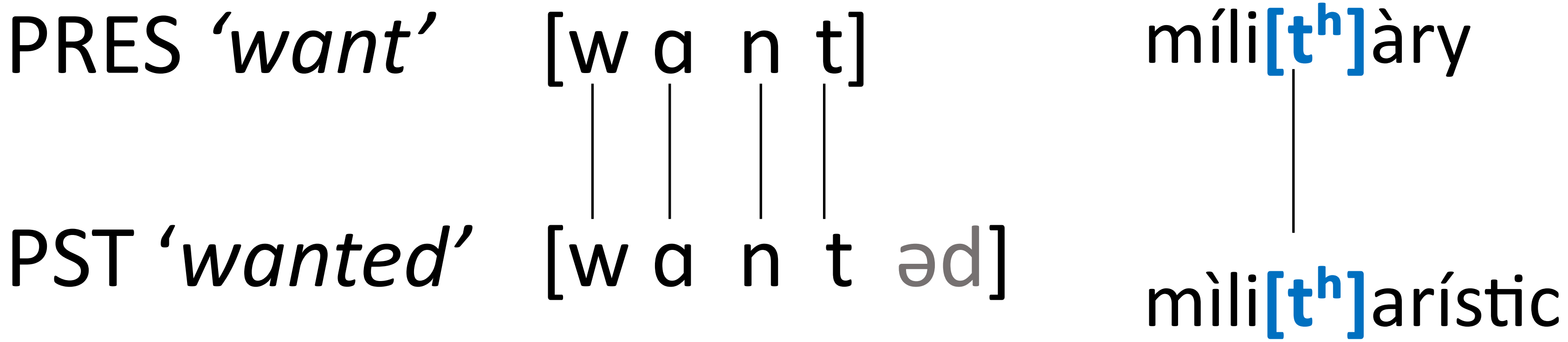
# Similarity across a paradigm

- In fact, surface forms in a paradigm (i.e. across grammatical contexts) can influence each other.
- Example: English aspiration vs. flapping (Withgott 1983)

| | | |
|---|---|---|
| <militaristic> | mìli[$t^h$]arístic | (cf. míli[$t^h$]àry) |
| <capitalistic> | càpi[ɾ]alístic | (cf. cápi[ɾ]al) |

# How do we formalize this generalization?

- **Correspondences** between related forms (Benua 1995; McCarthy & Prince 1995)
  - typically assume a **linear**, 1:1 relationship between segments.



PRES *'want'* [w ɑ n t]

PST *'wanted'* [w ɑ n t əd]

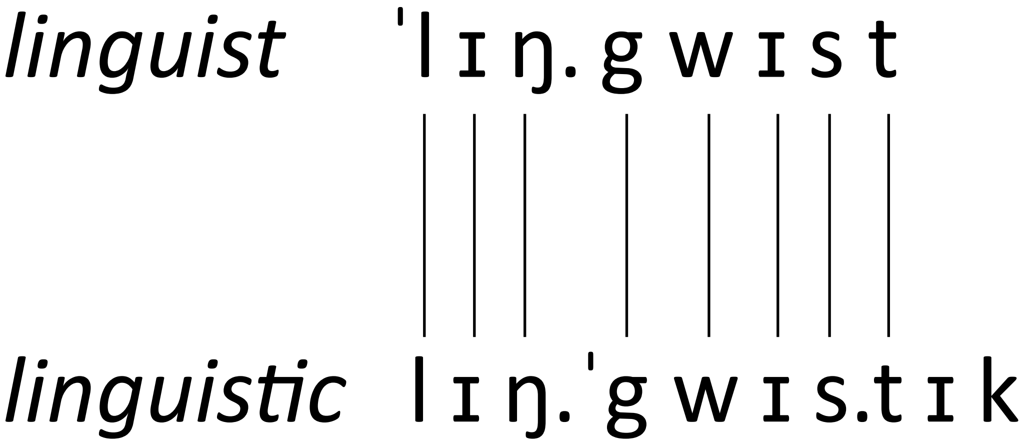
míli[t$^h$]àry

mìli[t$^h$]arístic

e.g. [w]$_{PRES}$ corresponds to [w]$_{PST}$

# Can non-linear correspondences exist?

- Crosswhite (1995) proposes **prosodic correspondence**, where **stressed syllables** of related words correspond to each other



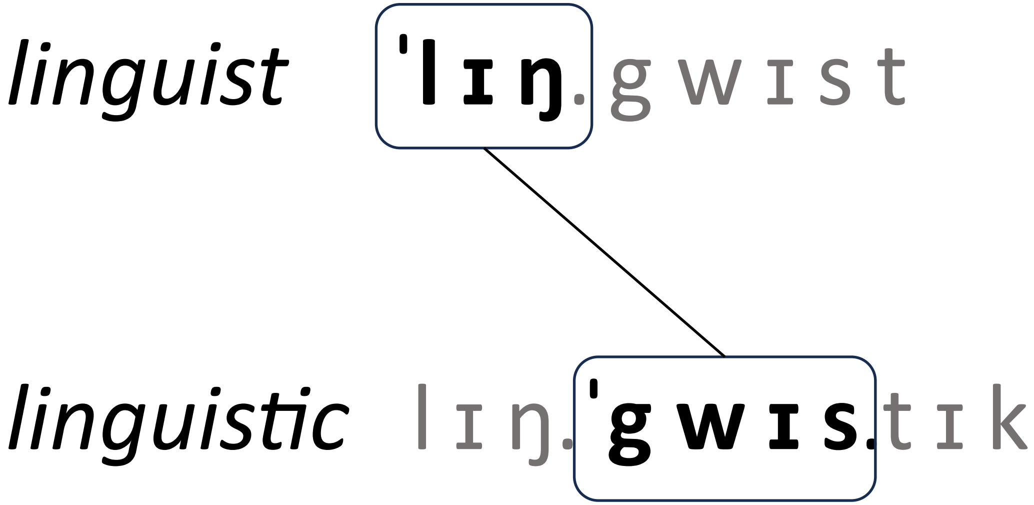
*linguist* ˈl ɪ ŋ. g w ɪ s t

*linguistic* l ɪ ŋ.ˈg w ɪ s.t ɪ k

Typical linear segmental correspondence

# Can non-linear correspondences exist?

- Crosswhite (1995) proposes **prosodic correspondence**, where **stressed syllables** of related words correspond to each other



*linguist* ˈlɪŋ.gwɪst

*linguistic* lɪŋ.ˈgwɪs.tɪk

Prosodic correspondence

# Can non-linear correspondences exist?

- Crosswhite (1995) proposes **prosodic correspondence**, where **stressed syllables** of related words correspond to each other
- Very little empirical evidence to date
    - one case from Chamorro.

# Goals of the talk

1. Present evidence for prosodic correspondence from Tgdaya Seediq.
2. Demonstrate the usefulness of looking at
   - probabilistic patterns
   - experimental evidence

   …. when asking questions about phonological representation.

3. Present a preliminary model of how Seediq speakers learn prosodic correspondence

# Outline of talk

**Intro**

Descriptive facts of Seediq

**Corpus**

Evidence for a gradient prosodic corr. effect

**Experiment**

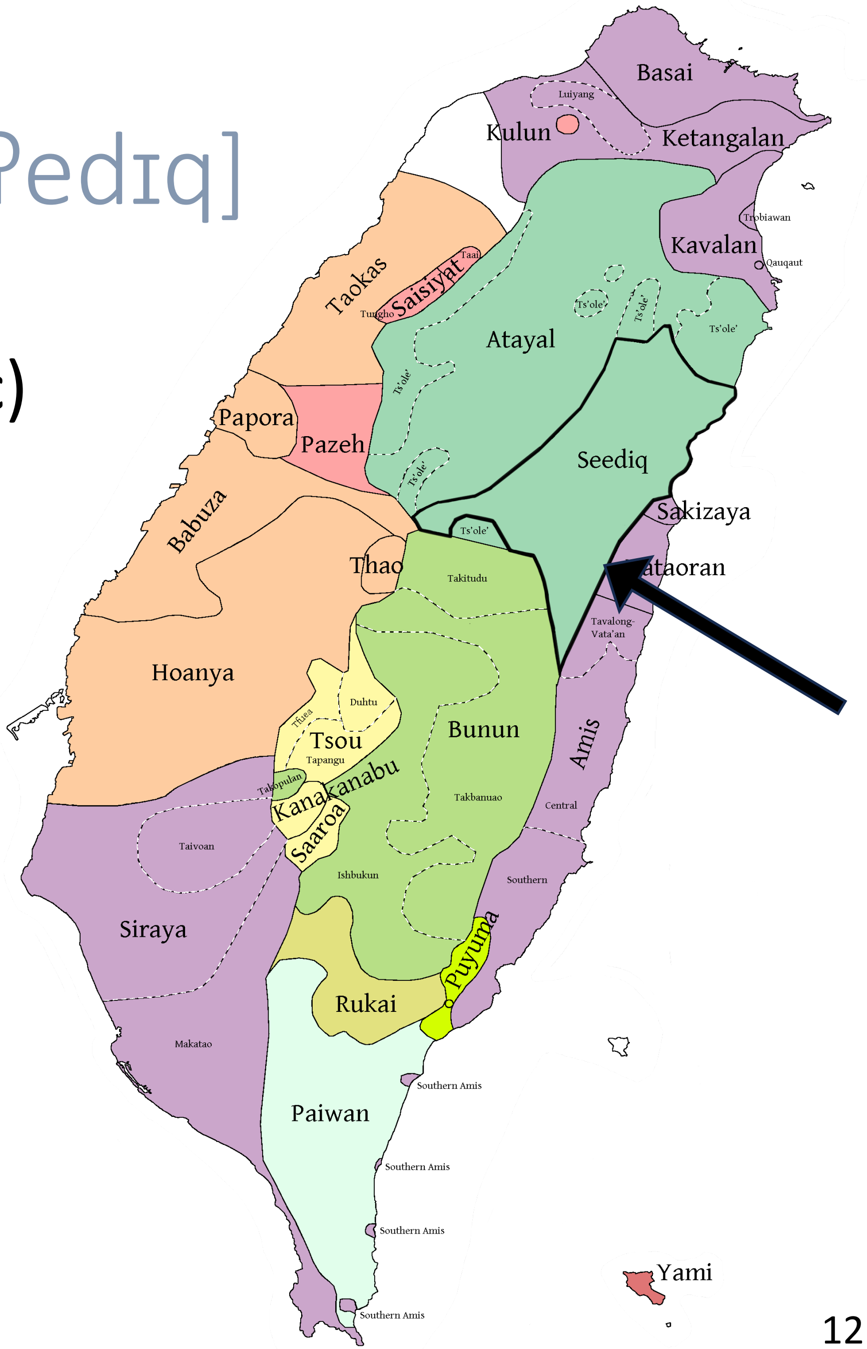
Results of a wug test supporting these findings

**Modeling**

A model of how speakers can learn and extend pros. corr

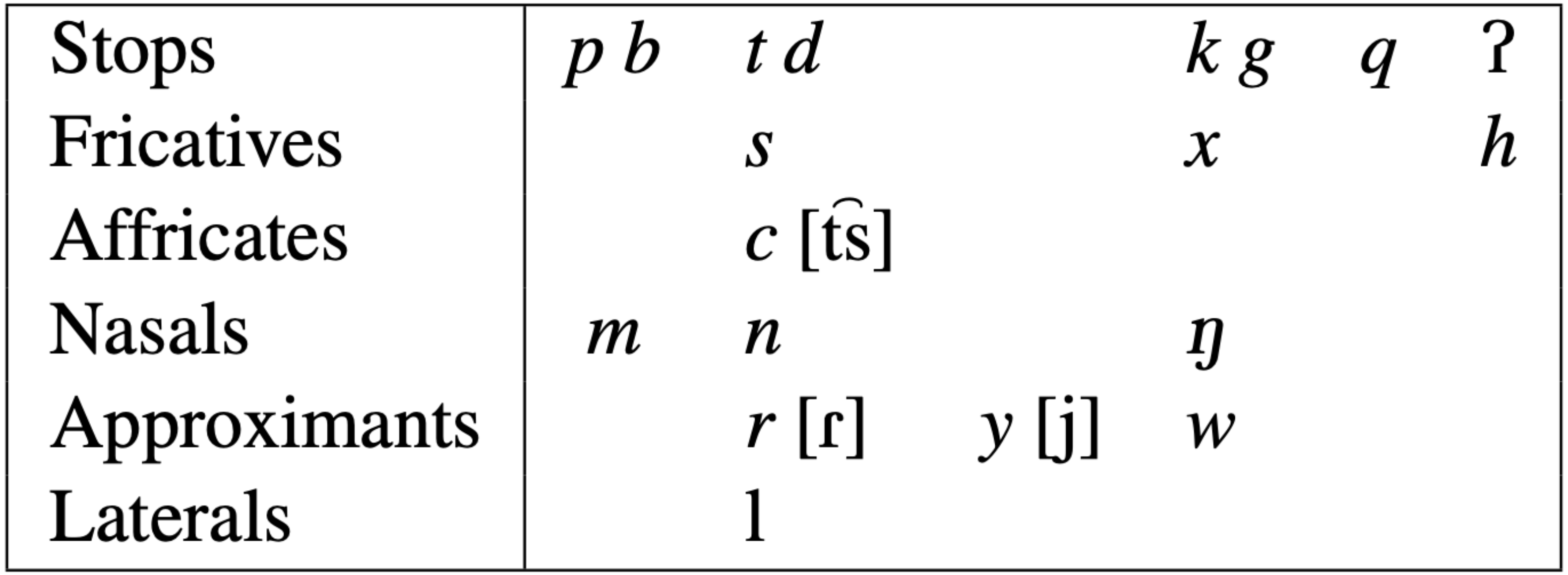
# Tgdaya Seediq [tuguˈdaya seˈʔedɪq]

- a dialect of Seediq (Austronesian, Atayalic)
- Located in central Taiwan
  - ~2,500 members including non-speakers
  - critically endangered



# Phoneme inventory

- 5 vowels /a e i o u/
- Consonants:



| | | | | | | |
|---|---|---|---|---|---|---|
| Stops | *p b* | *t d* | | *k g* | *q* | ʔ |
| Fricatives | | *s* | | *x* | | *h* |
| Affricates | | *c* [t͡s] | | | | |
| Nasals | *m* | *n* | | *ŋ* | | |
| Approximants | | *r* [ɾ] | *y* [j] | *w* | | |
| Laterals | | l | | | | |

# Vowel alternations in Seediq

- Stress is always penultimate
    - written with acute accent on stressed vowel
    - e.g. [**pé.**mux]
- Extensive stress-driven vowel alternations

# Vowel alternations in Seediq (Yang 1976)

- Pretonic vowel reduction: before the stressed syllable, all vowels become [u]*

| UR | stem | suffixed | gloss |
|---|---|---|---|
| /gedaŋ/ | gédaŋ | gudáŋ-an | ‘die’ |
| /biciq/ | bíciq | bucíq-an | ‘decrease’ |
| /barah/ | bárah | buráh-an | ‘rare’ |

**Sample derivation**

| | |
|---|---|
| UR | /gedaŋ-an/ |
| Stress | gedáŋan |
| pretonic V→[u] | gudáŋan |
| SR | [gudáŋan] |

*simplifying a bit here; feel free to ask me in the Q&A!

# Vowel alternations in Seediq

- Post-tonic vowel reduction: after the stressed syllable, /e/ and /o/ become become [u].

| UR | stem | suffixed | gloss |
|---|---|---|---|
| /remux/ | rémux | rumúx-an | ‘enter’ |
| /pemex/ | pémux | puméx-an | ‘hold’ |
| /kodoŋ/ | kóduŋ | kudóŋ-an | ‘hook’ |

**Sample derivation**

| | |
|---|---|
| UR | /pemex/ |
| Stress | pémex |
| pret. V→[u] | -- |
| post.  /e,o/→[u] | pémux |
| SR | [pémux] |

- In other words, post-tonic [u] can alternate with [e] or [o]

# Vowel alternations in Seediq

- As a result of these two processes, surface forms within a paradigm can look very different.
- Some more examples…

| stem | suffixed | gloss |
|---|---|---|
| h**á**ŋ**u**c | h**u**ŋ**é**d-an | ‘cook, boil’ |
| m**á**l**u** | m**u**l**é**(j)-an | ‘able to’ |
| dóʔ**u**s | doʔ**ó**s-an | ‘refine’ (metal)’ |

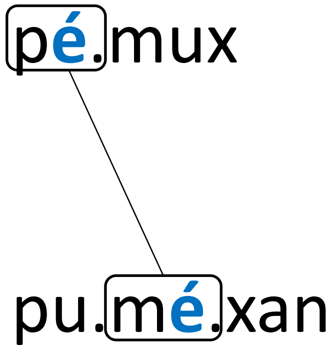
# Prosodic correspondence in Seediq

- For stems which undergo post-tonic VR, there is a strong tendency for stressed vowels of stem and suffixed forms to match.

| **stem** | **suffixed** | **gloss** | |
|---|---|---|---|
| pémux | puméx-an | ‘hold’ | ☺ **Stressed Vs match** |
| kóduŋ | kudóŋ-an | ‘hook’ | |
| háŋuc | huŋéd-an | ‘cook, boil’ | ☹ **Stressed Vs mismatch** |
| rémux | rumúx-an | ‘enter’ | |

# Prosodic correspondence in Seediq

- Evidence for **prosodic correspondence** (i.e. pressure for stressed syllables within a paradigm to be similar to e/o)



pé.mux

pu.mé.xan ‘hold’

# Outline of talk

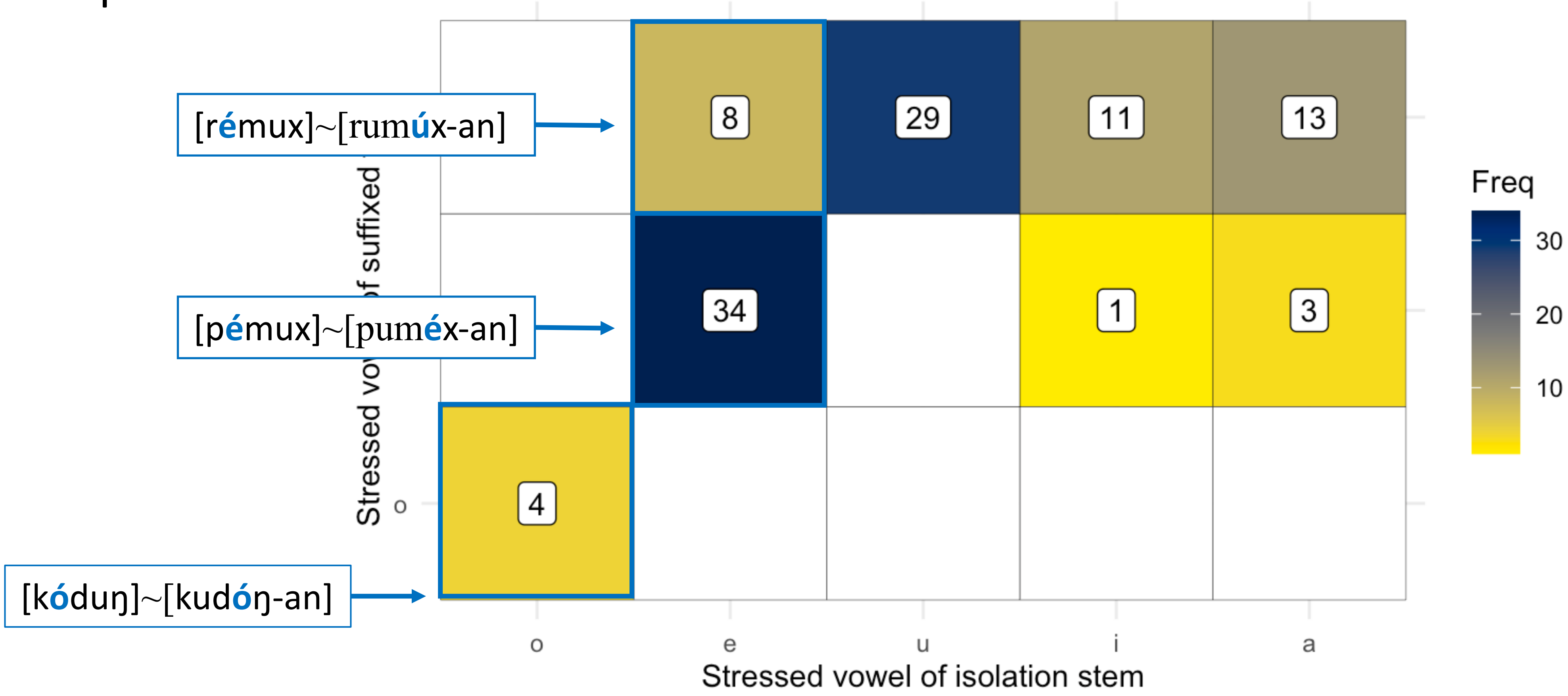
| Intro | **Corpus** | Experiment | Modeling |
|---|---|---|---|
| Descriptive facts of Seediq | Evidence for a gradient prosodic corr. effect | Results of a wug test supporting these findings | A model of how speakers can learn and extend pros. corr |

# Data

- 341 verbal paradigms (stem-suffix pairs)
  - Taiwan Aboriginal e-Dictionary (Council of Indigenous Peoples 2020)
  - fieldwork with three Seediq speakers (ages 69-78), carried out in Puli Township, Nantou, Taiwan.

# Vowel matching in Seediq

**Figure:** Stressed vowel in stem vs. suffixed form, in words that undergo post-tonic VR.



[rémux]~[rumúx-an]

[pémux]~[pumé͡x-an]

[kóduŋ]~[kudóŋ-an]

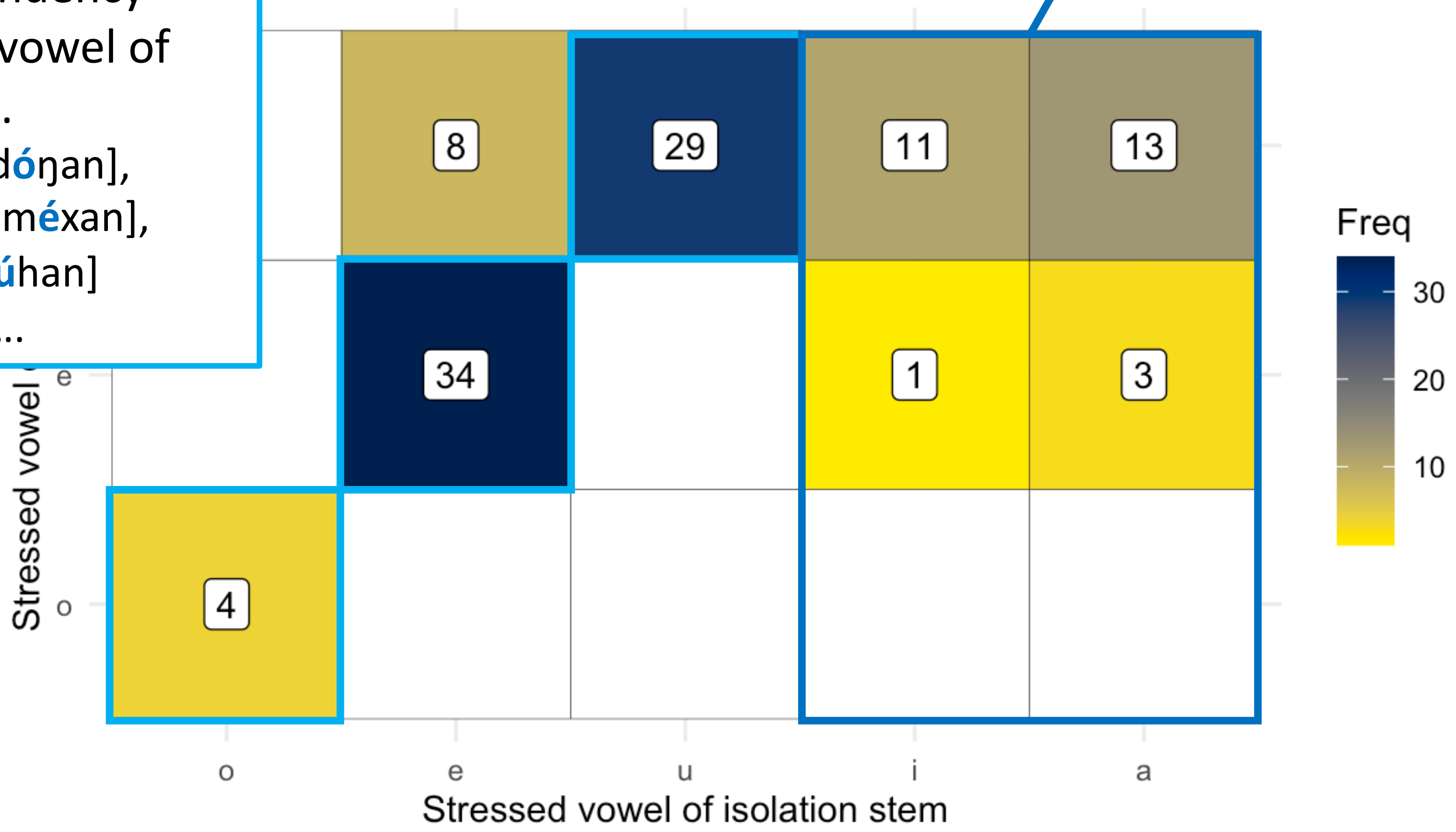
# Vowel matching in Seediq

Otherwise, there is a preference for **non-alternation**.

e.g. [gátuk]~[gutúkan],
[híluŋ]~[hulúŋan]

**Vowel matching** tendency when the stressed vowel of the stem is [e, o, u].

e.g. [k**ó**duŋ~kud**ó**ŋan],
[p**é**mux~pum**é**xan],
[p**ú**luh~pul**ú**han]
....



# Another way of looking at the data…

[pémux]~[puméx-an]

[kóduŋ]~[kudóŋ-an]

[búnuh]~[bunúh-an]

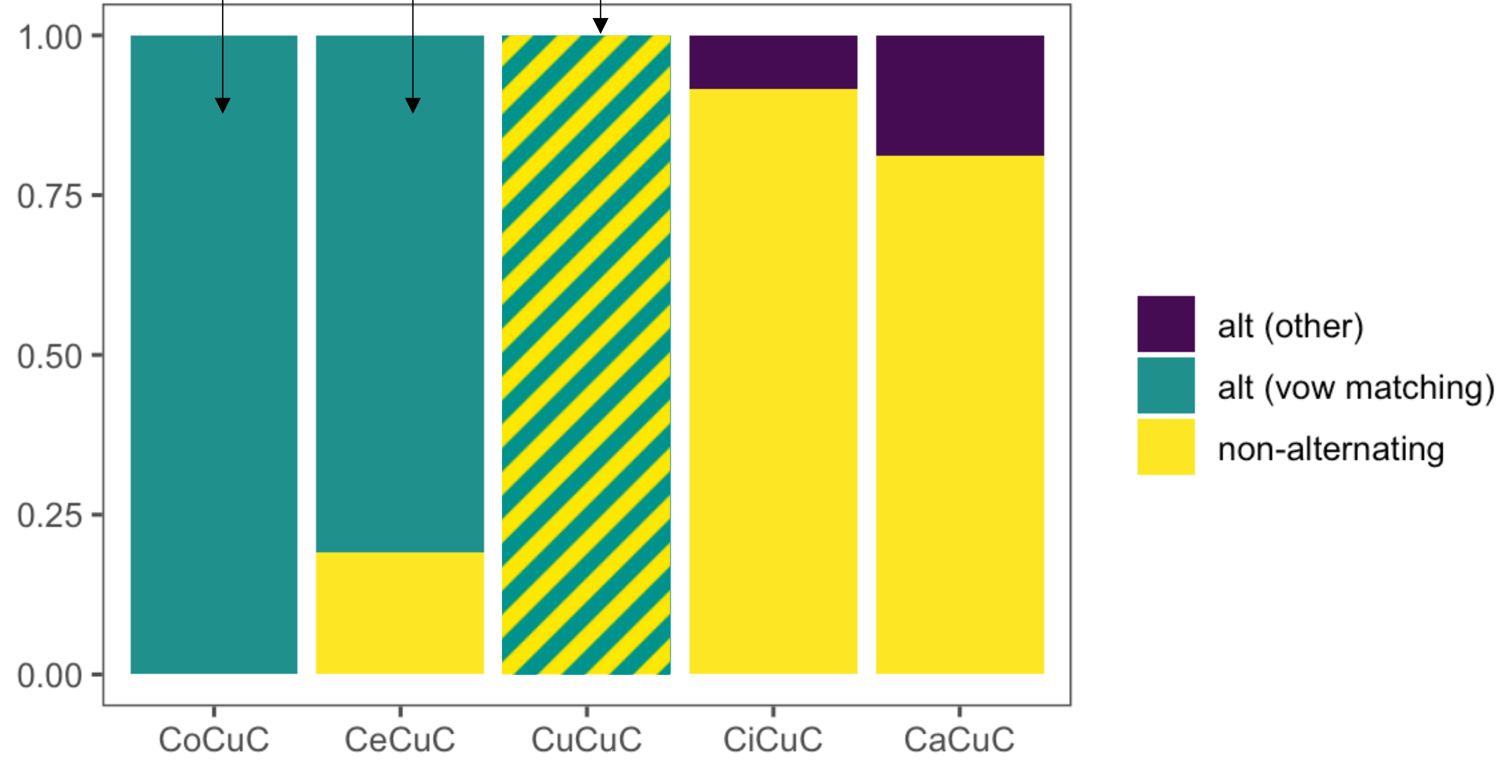


**Figure**: Proportion of alternating suffixed forms for CVCuC stems

# Psychological reality of prosodic correspondence

- So far, it seems like vowel matching exists as a **gradient tendency** in the lexicon.
- But it is psychologically real?

# Outline of talk

**Intro**

Descriptive facts of Seediq

**Corpus**

Evidence for a gradient prosodic corr. effect

**Experiment**

Results of a wug test supporting these findings

**Modeling**

A model of how speakers can learn and extend pros. corr

# Method: wug test (Berko 1958)

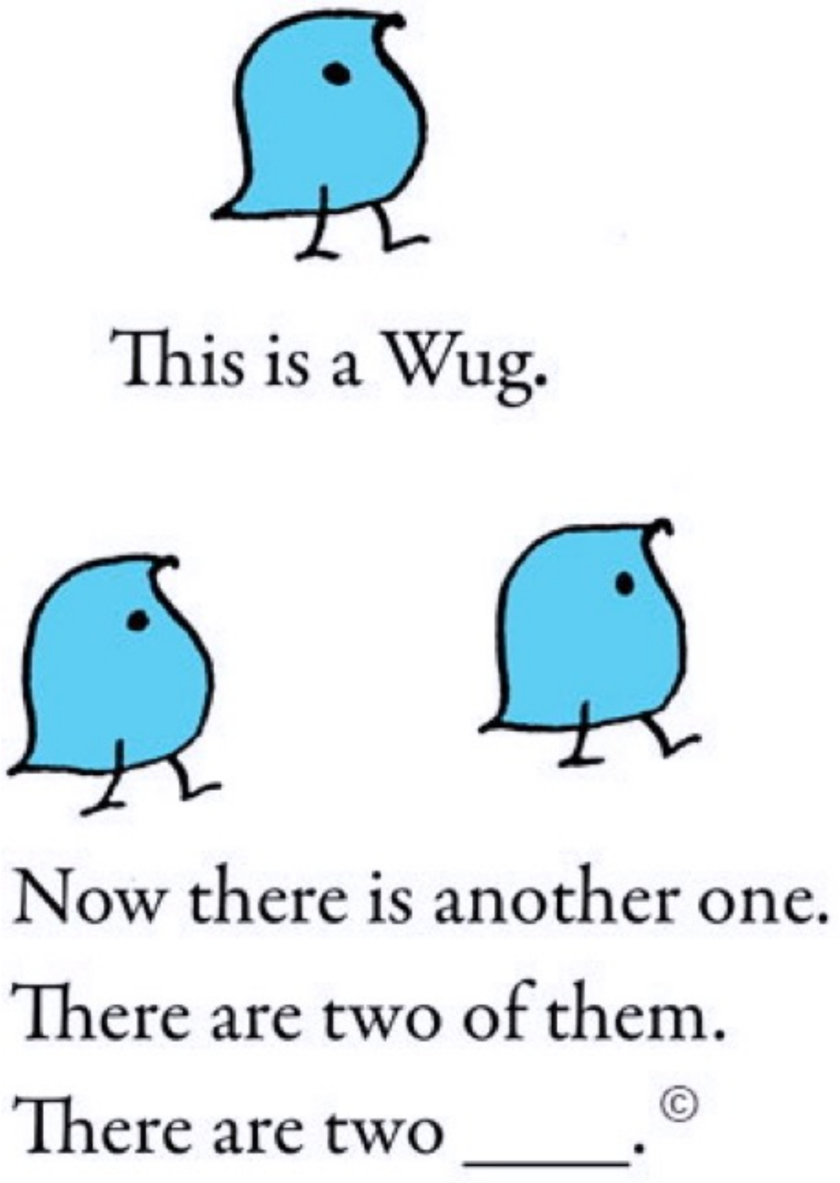
- Tests whether speakers have generalized productive grammars from the lexicon.
- Present participants with nonce words of their native language
- …and ask them to apply a morphological rule (e.g. plural formation)

This is a Wug.

# Method: wug test (Berko 1958)

- Tests whether speakers have generalized productive grammars from the lexicon.
- Present participants with ~~nonce words~~ of their native language
- …and ask them to apply a morphological rule (e.g. plural formation)

**‘gapped’ stems**, i.e. ones with no known suffixed forms



This is a Wug.

Now there is another one.

There are two of them.

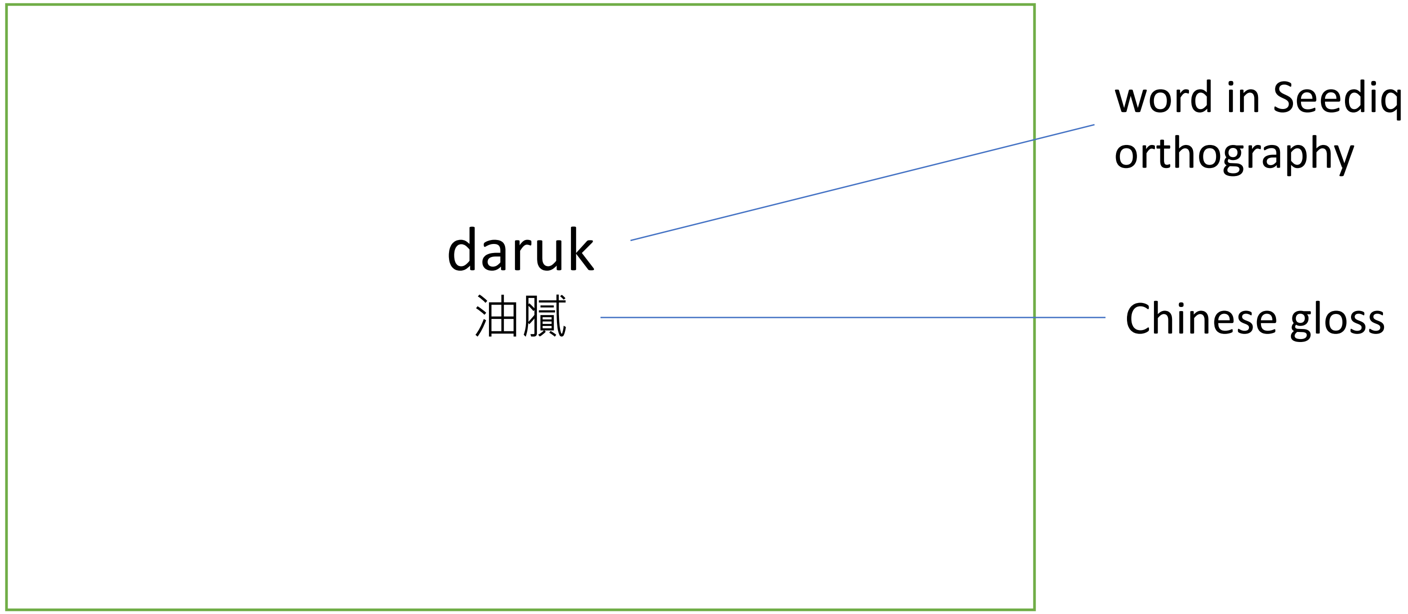
There are two _____.

©

Photo courtesy of Jean Berko Gleason

# Methods *cont.*

- **Participants**: adult native speakers (N=10, 7F, ages 45-76).
- **Procedure**: Speakers were shown test items, and asked to produce them with two suffixes: /-an/ ‘LF’ & /-i/ ‘PF.IMP’



daruk

油膩

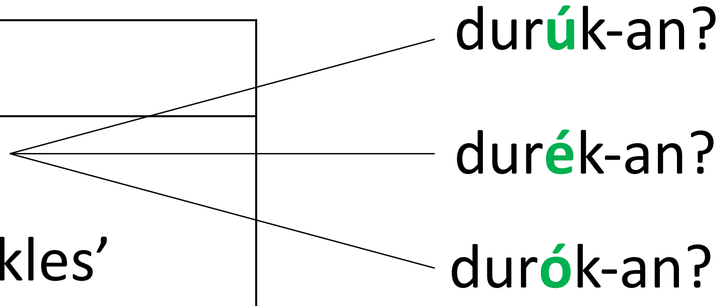
word in Seediq orthography

Chinese gloss

# Stimuli

- ‘gapped’ stems of the form $CV_1CV_2C$; $V_1 = \{a,e,u\}$, $V_2=\{a,u\}$

| $V_1$ | $V_2$ | Example |
|---|---|---|
| a | u | dáruk ‘oil’ |
| e | u | kéruŋ ‘wrinkles’ |
| u | u | cúguk ‘type of plant’ |



durúk-an?

durék-an?

durók-an?

# Stimuli

- ‘gapped’ stems of the form $CV_1CV_2C$; $V_1$ = {a,e,u}, $V_2$={a,u}
- 72 items (6x8 test items + 24 fillers)

| $V_1$ | $V_2$ | Example |
|---|---|---|
| a | u | dáruk ‘oil(y)’ |
| e | u | kéruŋ ‘wrinkles’ |
| u | u | cúguk ‘type of plant’ |
| a | a | sábak ‘dregs, pulp’ |
| e | a | réhak ‘seed’ |
| u | a | súwak ‘yawn’ |

**Control:** [a] should be non-alternating.

subák-an

~~subék-an~~

~~subók-an~~

# Predictions

Possible outcomes:

- **No pattern internalized:** no vowel alternations.

| $V_1$ | $V_2$ | Example | Outcomes: non-alternation |
|---|---|---|---|
| a | u | dáruk | durúk-an |
| e | u | kéruŋ | kurúŋ-an |
| u | u | cúguk | cugúk-an |
| a | a | sábak | subák-an |
| e | a | réhak | ruhák-an |
| u | a | súwak | suwák-an |

# Predictions

Vowel matching alternation

Possible outcomes:

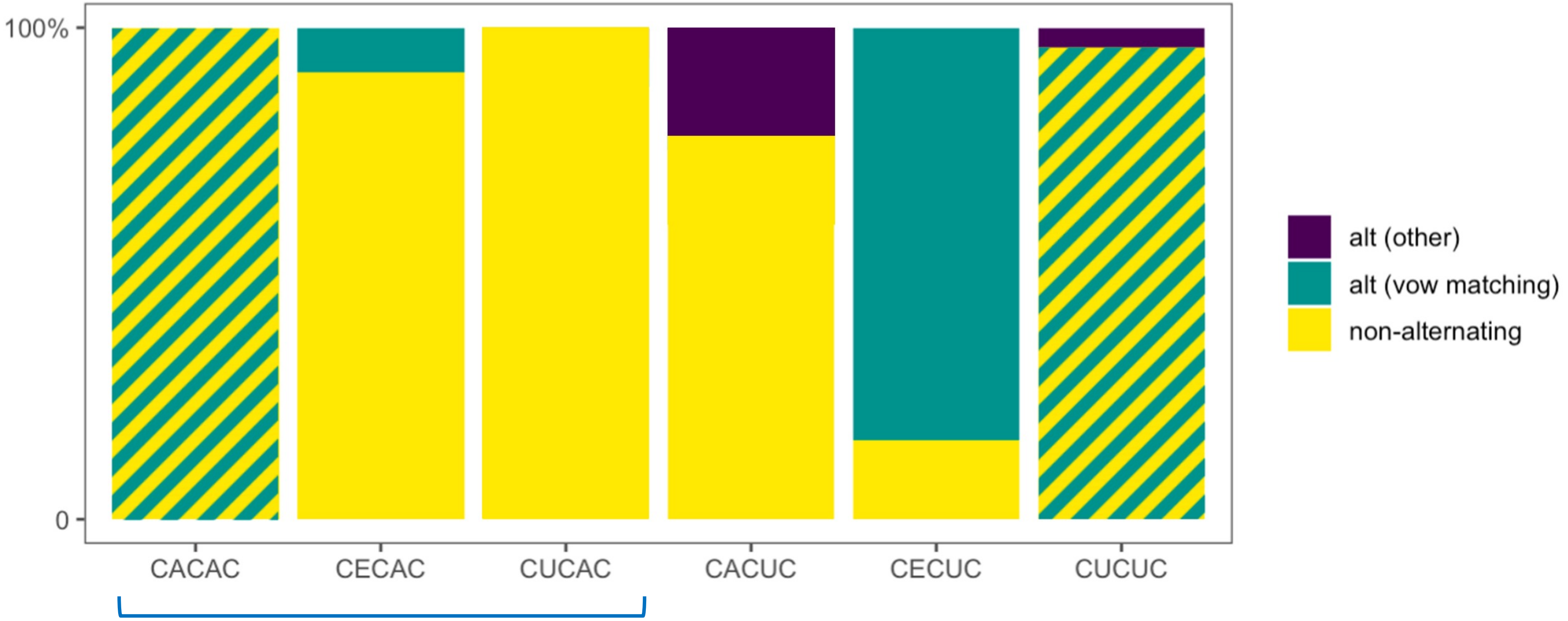
- **No pattern internalized**: no vowel alternations.
- **Frequency-matching**: apply alternations in a way that matches their rate in the lexicon.

For more examples of frequency-matching: Zuraw 2000, Ernestus & Baayen 2003, Hayes & Londe 2006; Zuraw 2010

| $V_1$ | $V_2$ | Example | Outcomes:<br>freq. matching |
|---|---|---|---|
| a | u | dáruk | mostly durúk-an |
| e | u | kéruŋ | mostly kuréŋ-an |
| u | u | cúguk | always cugúk-an |
| a | a | sábak | subák-an |
| e | a | réhak | ruhák-an |
| u | a | súwak | suwák-an |

# Frequency matching predictions

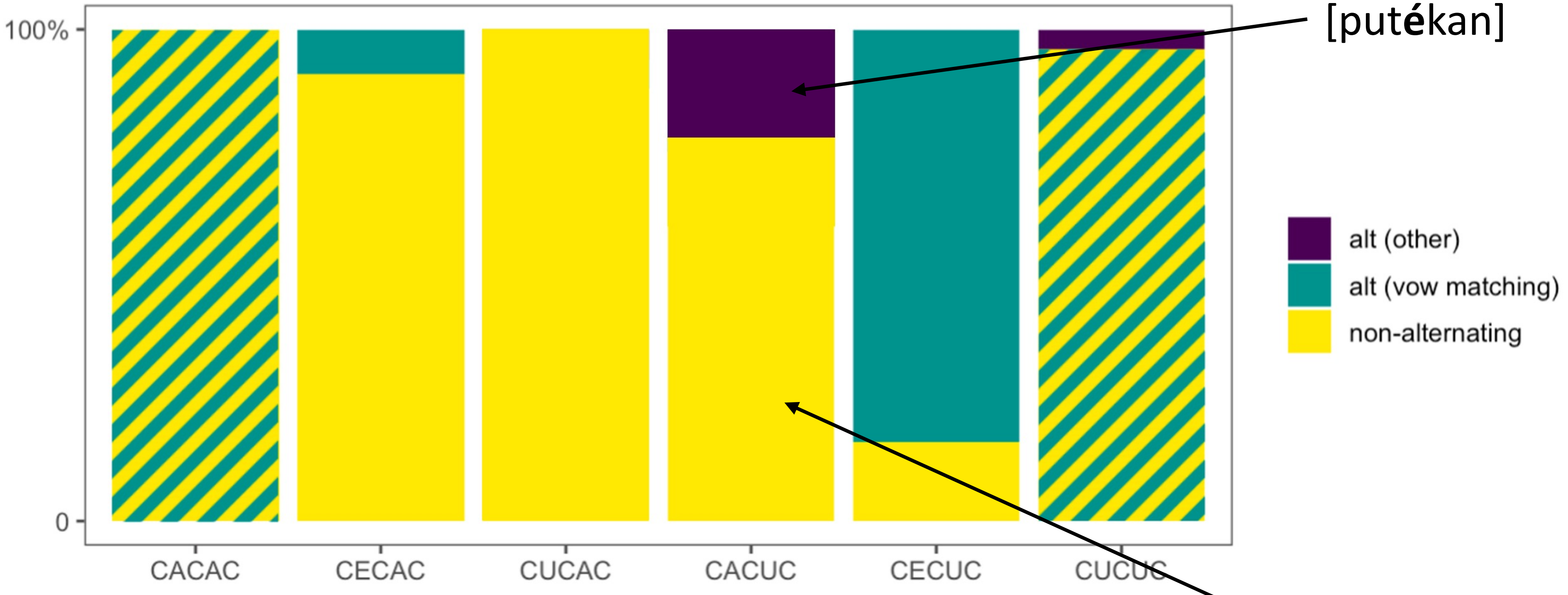
**Figure**: proportion of vowel alternation types in the lexicon



Post-tonic [a] should be non-alternating

# Frequency matching predictions

**Figure**: proportion of vowel alternation types in the lexicon



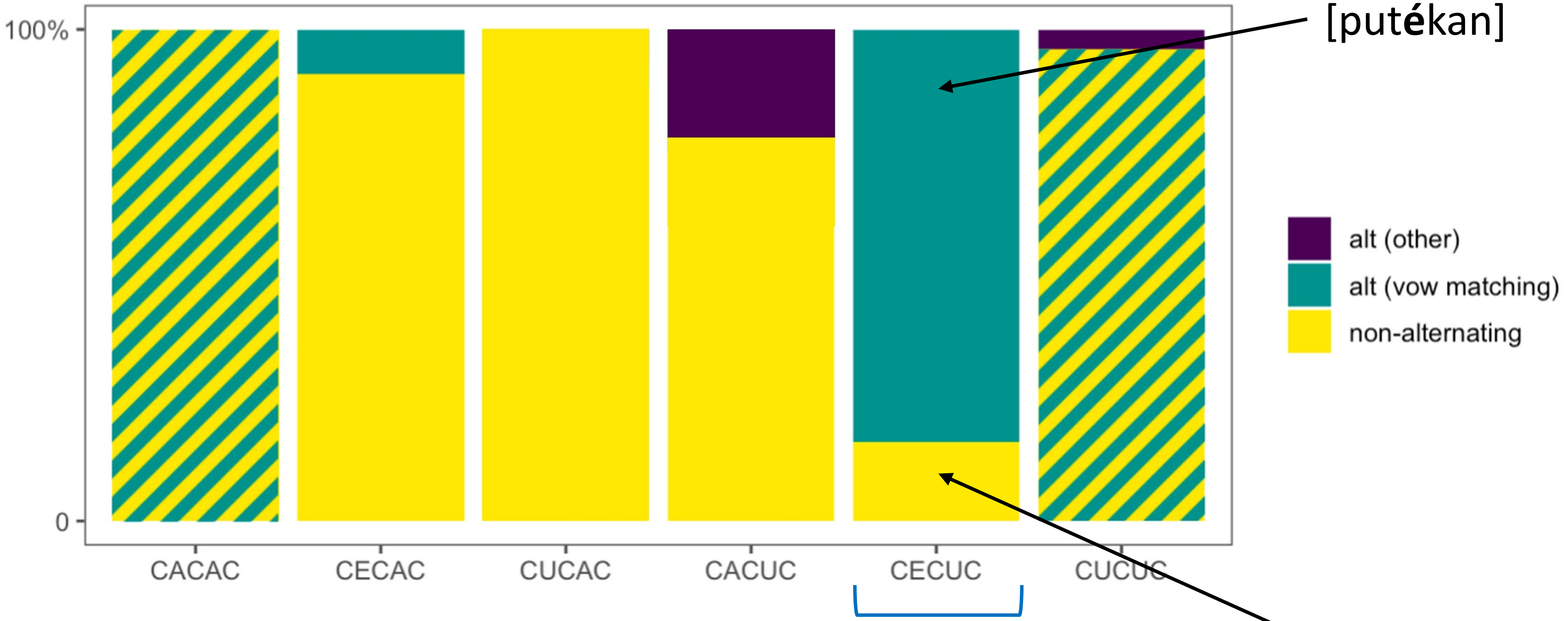
[put**é**kan]

[put**ú**kan]

For words like [p**á**tuk], the suffixed form should mostly be [put**ú**kan], but sometimes [put**é**kan]

# Frequency matching predictions

**Figure**: proportion of vowel alternation types in the lexicon



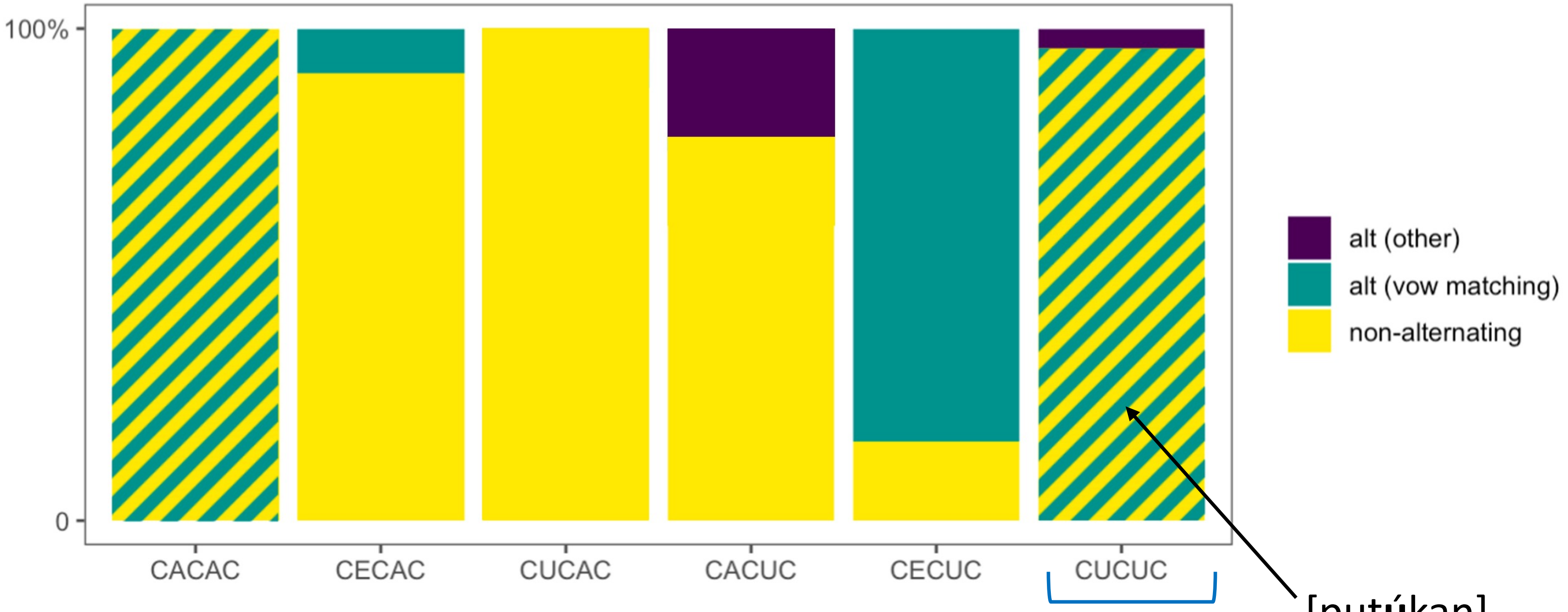
[put**é**kan]

[put**ú**kan]

For words like [p**é**tuk], the suffixed form  should mostly be [put**é**kan], but sometimes [put**ú**kan]

# Frequency matching predictions

**Figure**: proportion of vowel alternation types in the lexicon



[put**ú**kan]

For words like [p**ú**tuk], the suffixed form  should mostly/always be be [put**ú**kan]

# Predictions

not observed in the lexicon

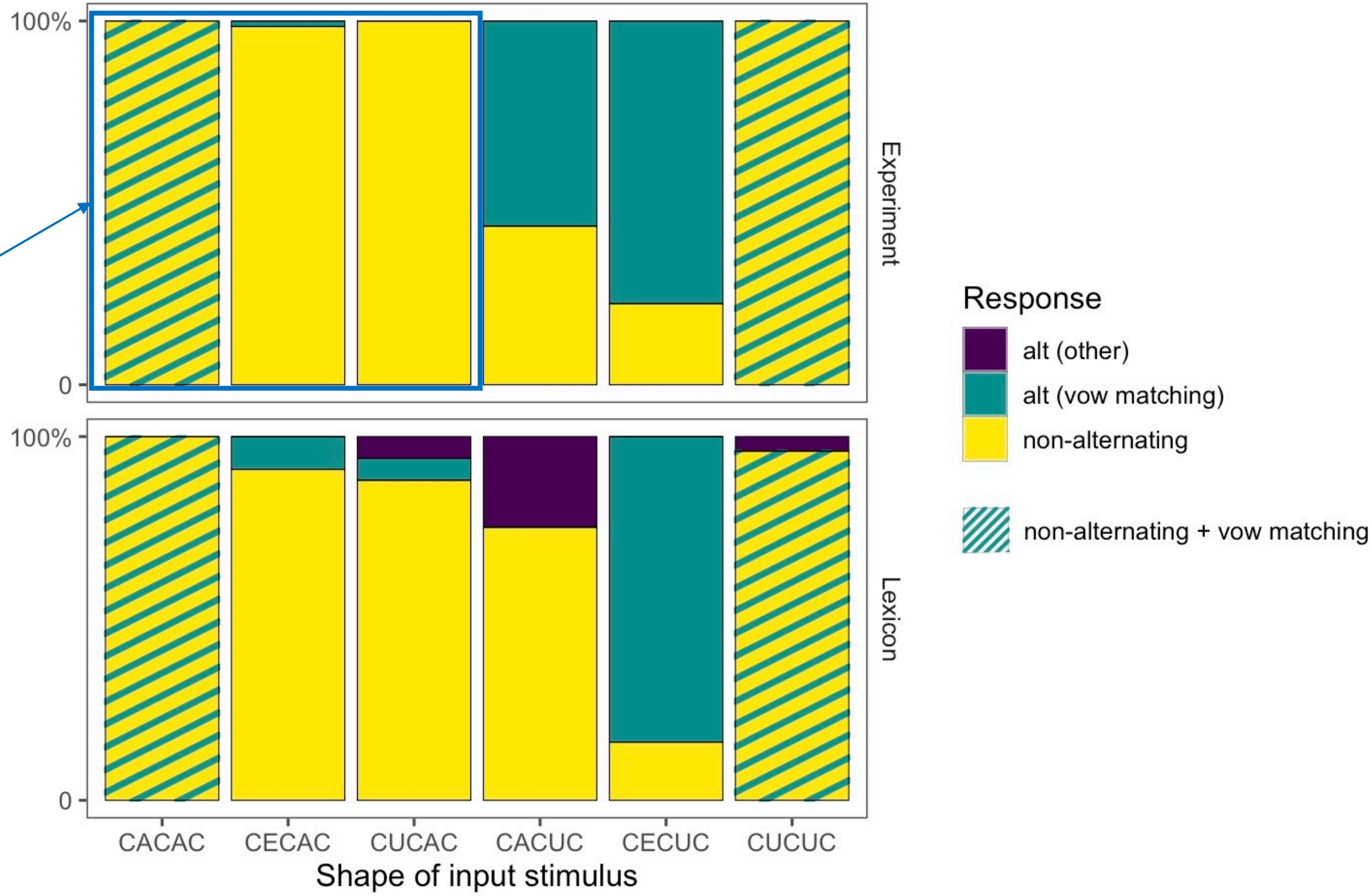
Possible outcomes:

- **No pattern internalized**: no vowel alternations.
- **Frequency-matching**: apply alternations in a way that matches their rate in the lexicon.
- **Overlearning**: apply vowel matching alternations more than predicted by the lexicon.

| $V_1$ | $V_2$ | Example | Outcomes:<br>vow matching |
|---|---|---|---|
| a | u | dáruk | durák-an?? |
| e | u | kéruŋ | kuréŋ-an |
| u | u | cúguk | cugúk-an |
| a | a | sábak | subák-an |
| e | a | réhak | ruhák-an |
| u | a | súwak | suwák-an |

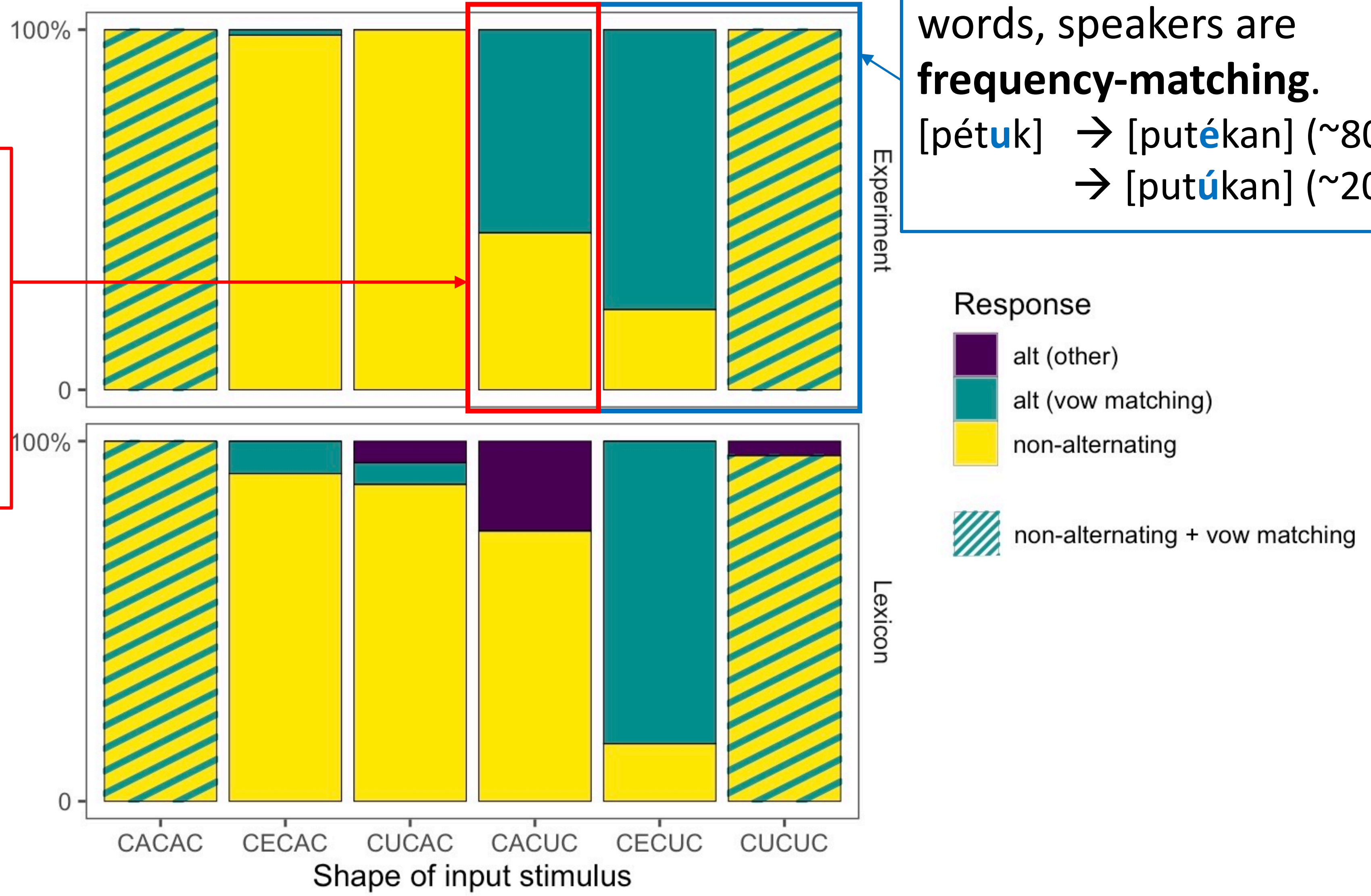
# Results

As expected, post-tonic [a] is non-alternating. [pútak]→[putákan], never *[putúkan]



# Results

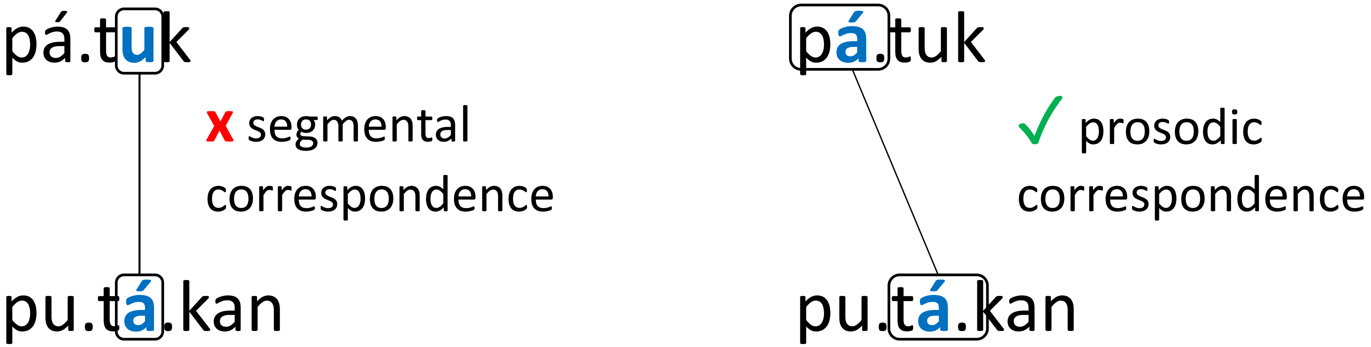
For CaCuC words, speakers are applying a new **vowel-matching** alternation that is ***not*** observed in the lexicon.
e.g. [pátuk] → [putákan]



For CeCuC and CuCuC words, speakers are **frequency-matching**.
[pétuk] → [putékan] (~80%)
→ [putúkan] (~20%)

# Interim summary

- **Vowel matching** is present in Seediq both as
  - trend in the lexicon
  - an active principle in wug tests
- Evidence for **prosodic correspondence** (pressure for stressed syllables within a paradigm to be similar)
- In fact, prosodic correspondence overrides segmental correspondence



pá.tuk

✗ segmental correspondence

pu.tá.kan

pá.tuk

✓ prosodic correspondence

pu.tá.kan

# Interim summary, *cont.*

- Unresolved issue: how do we model the learning of vowel matching?
  - Lexicon: vowel matching on [pétus], [pótus], [pútus]
  - Learned pattern: vowel matching *overgeneralized* to [pátus]
- Difficult, as learning models are generally frequency-matching

# Outline of talk

**Intro**

Descriptive facts of Seediq

**Corpus**

Evidence for a gradient prosodic corr. effect

**Experiment**

Results of a wug test supporting these findings

**Modeling**

A model of how speakers can learn and extend pros. corr

# Proposal: generality bias

- People are biased to learn more general patterns (Moreton & Pater 2012)

  “**Vowels match**” vs. “**Vowels match**, *if they are mid vowels*”

- **Modeling:** test this hypothesis
  - **Goal:** model that when trained on the lexicon, can predict the experimental results
  - **Preview**: generality bias improves model predictions

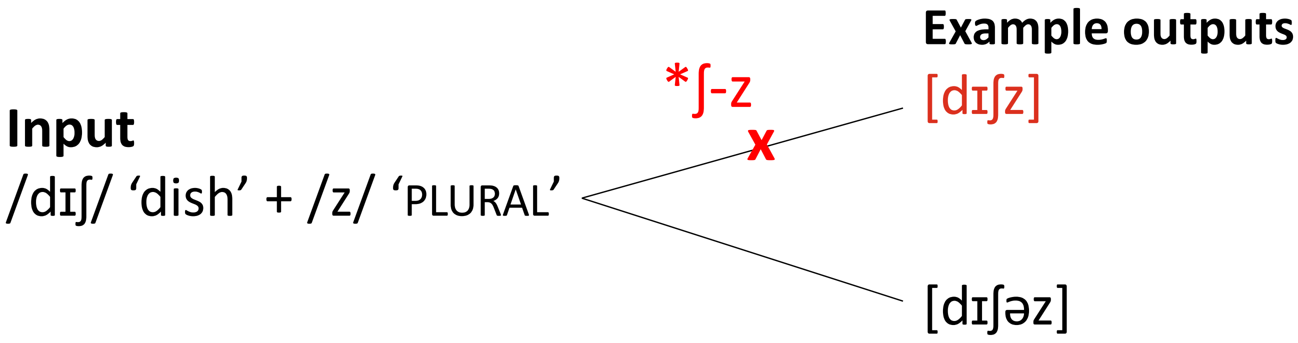
# Elements of the model

- **A probabilistic phonological grammar**
- Ability to incorporate generality bias

# 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” [ʃ] followed by [z] is not allowed (*ʃ-z)



**Input**

/dɪʃ/ ‘dish’ + /z/ ‘PLURAL’

*ʃ-z

x

**Example outputs**

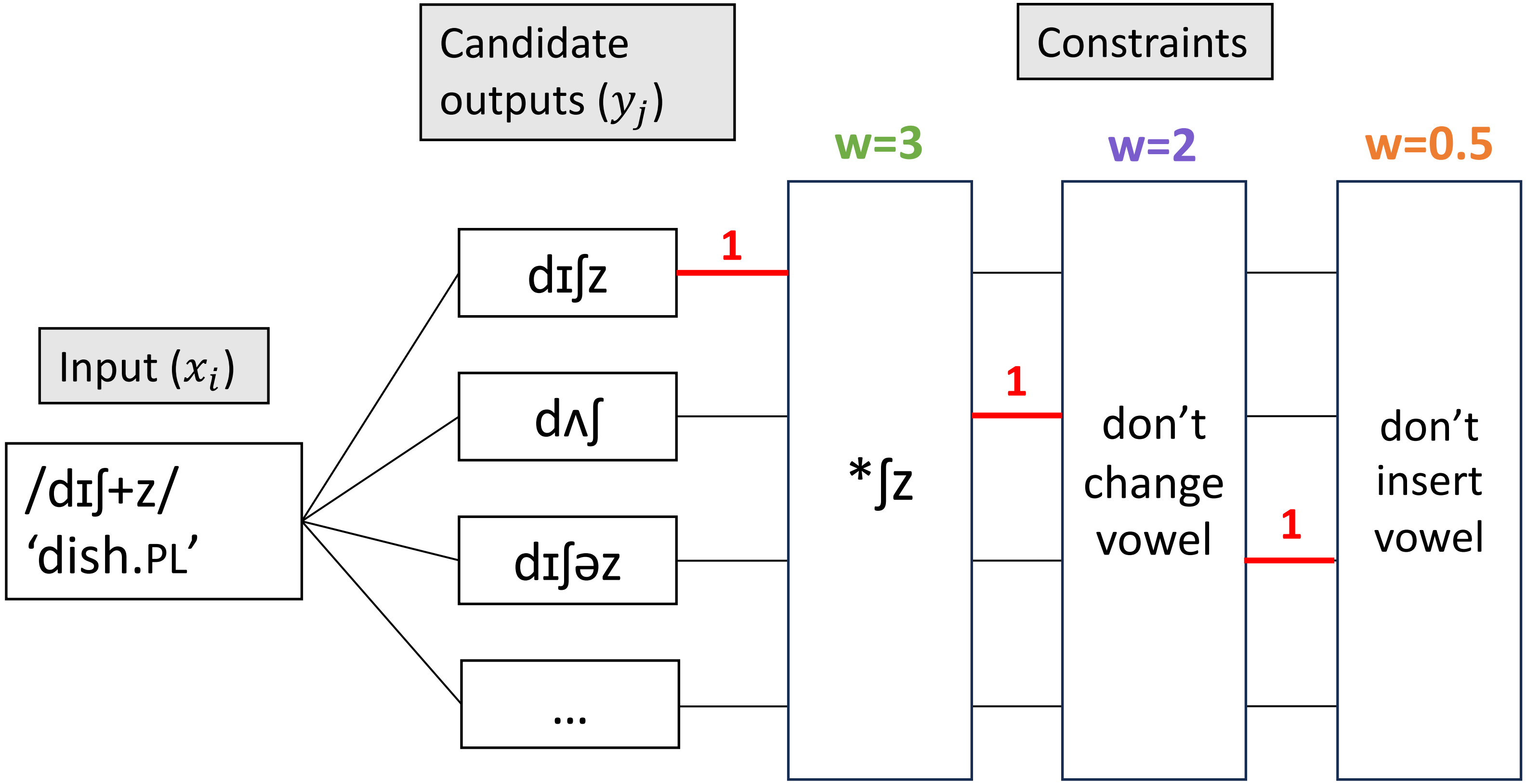
[dɪʃz]

[dɪʃəz]

# Phonological grammar

- The grammar also needs to be probabilistic
    - Maximum Entropy Harmonic Grammar (Smolensky 1986; Goldwater & Johnson, 2003)
    - probabilistic version of Optimality Theory
    - = multinomial logistic regression

# Phonological grammar



Candidate outputs ($y_j$)

Constraints

Input ($x_i$)

/dɪʃ+z/
‘dish.PL’

dɪʃz

dʌʃ

dɪʃəz

…

w=3

*ʃz

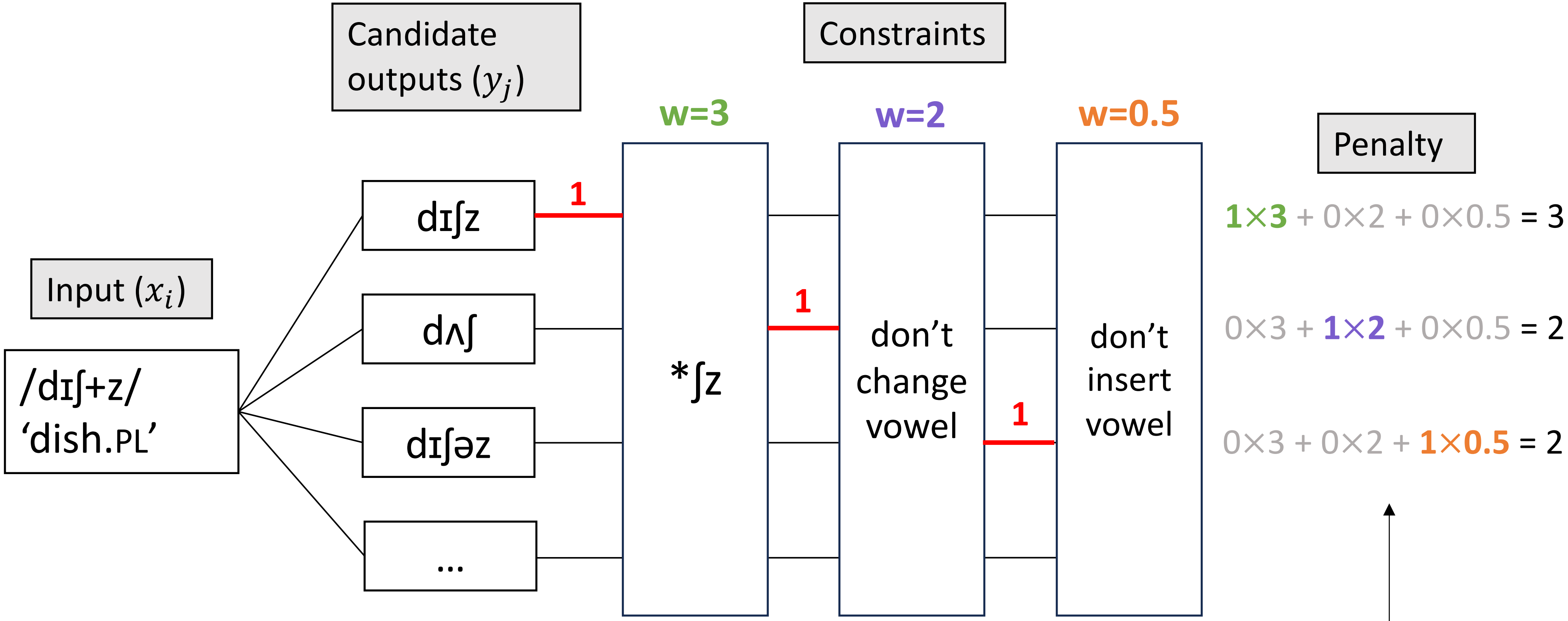
w=2

don’t change vowel

w=0.5

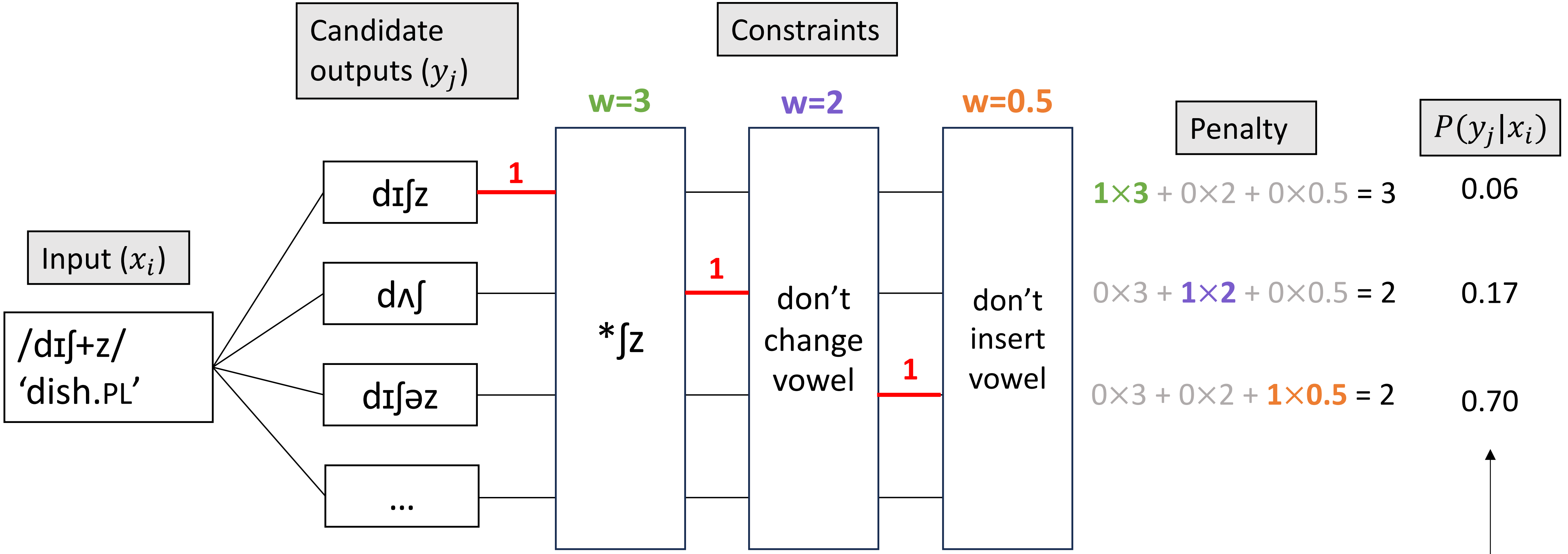
don’t insert vowel

# Phonological grammar



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

# Phonological grammar



Candidate outputs ($y_j$)

Constraints

Input ($x_i$)

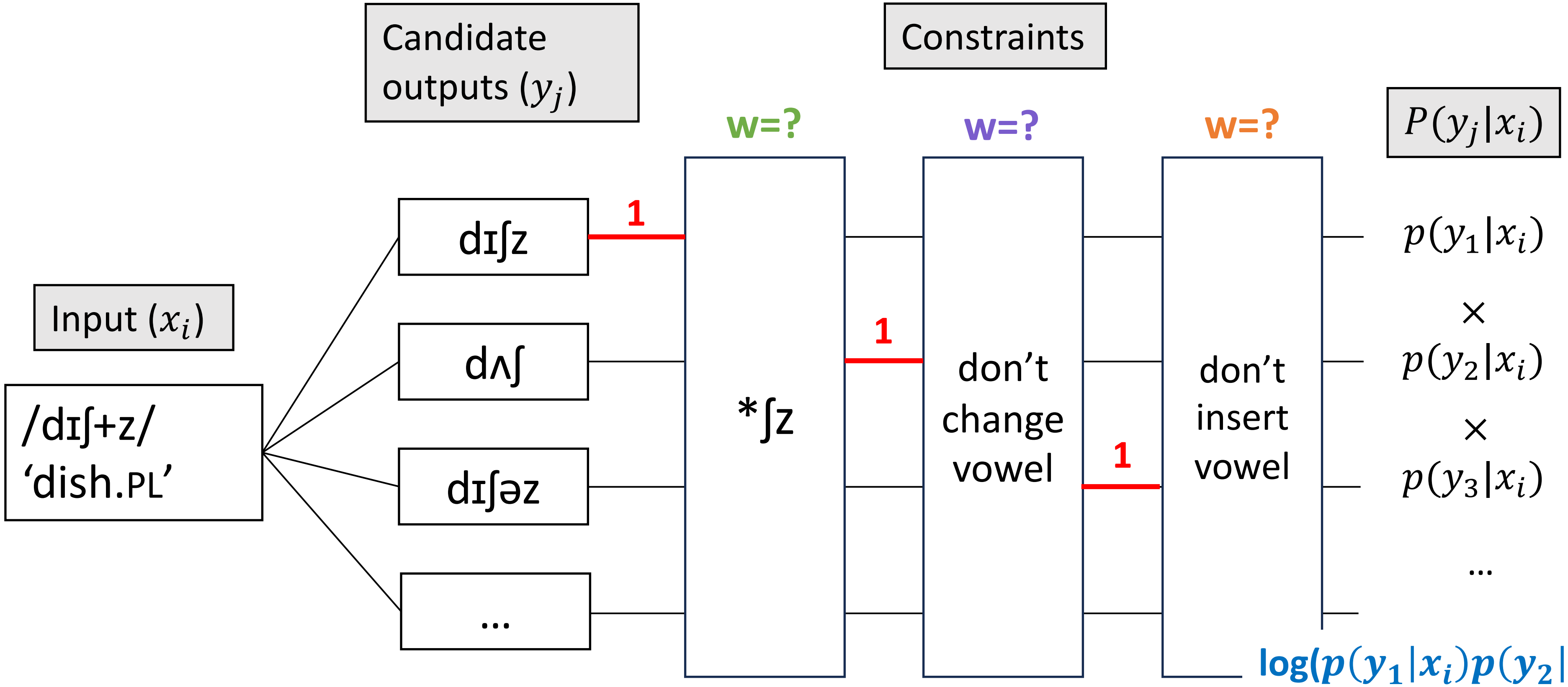
/dɪʃ+z/
‘dish.PL’

| Penalty | $P(y_j \mid x_i)$ |
|---|---|
| 1×3 + 0×2 + 0×0.5 = 3 | 0.06 |
| 0×3 + 1×2 + 0×0.5 = 2 | 0.17 |
| 0×3 + 0×2 + 1×0.5 = 2 | 0.70 |

**Takeaway:** Violating highly weighted constraints → lower probability

**Convert to probabilities**
probability of candidate $y_j$
given input $x_i$

# Phonological grammar

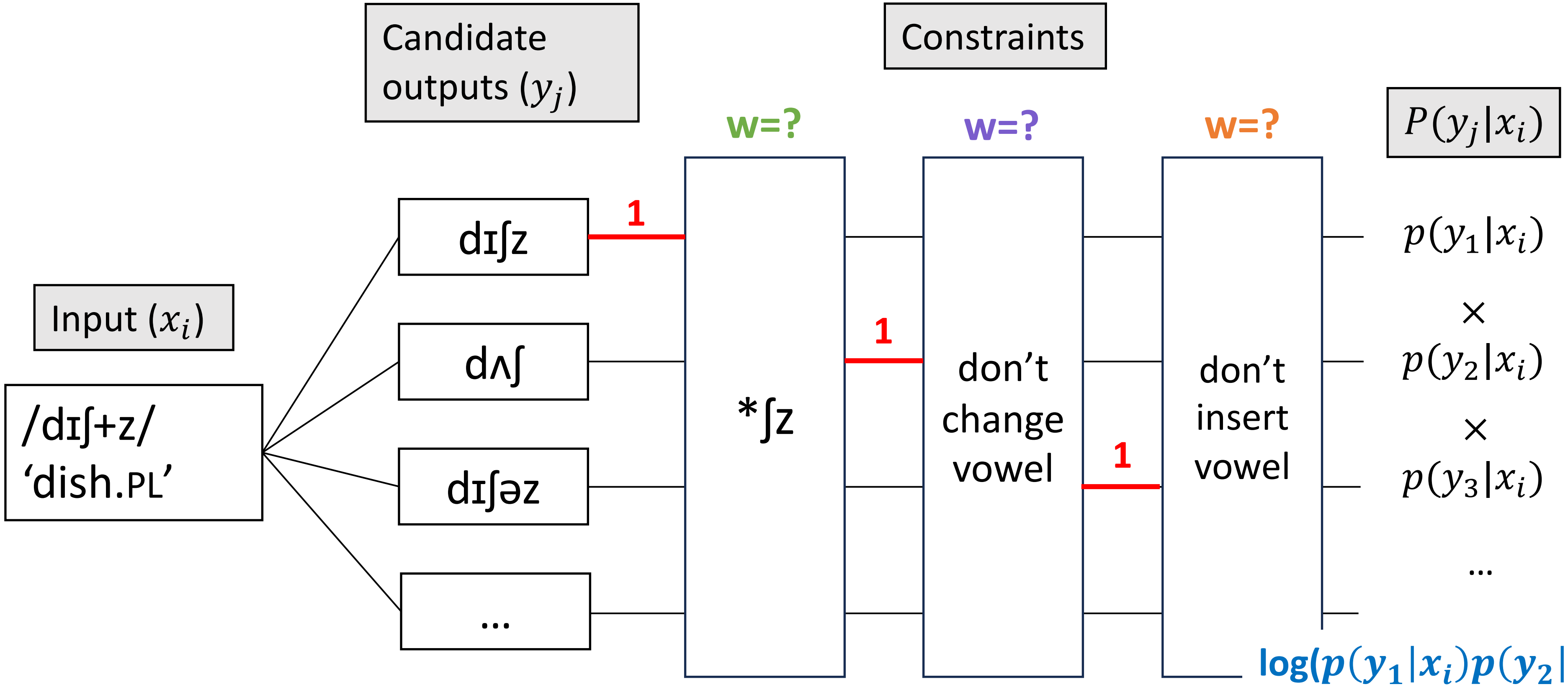


$$\log(p(y_1|x_i)p(y_2|x_i)\dots p(y_n|x_i))$$

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

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

# Phonological grammar



Candidate outputs ($y_j$)

Constraints

w=?

w=?

w=?

$P(y_j|x_i)$

Input ($x_i$)

/dɪʃ+z/
‘dish.PL’

dɪʃz

dʌʃ

dɪʃəz

…

*ʃz

don’t change vowel

don’t insert vowel

$p(y_1|x_i)$

$\times$

$p(y_2|x_i)$

$\times$

$p(y_3|x_i)$

…

$$\log(p(y_1|x_i)p(y_2|x_i)\ldots p(y_n|x_i))$$

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

The model at this point is **frequency-matching** (can match the lexicon)

Now let’s apply this to Seediq!

# Phonological grammar: constraints

***Specific*** vowel matching constraint

MATCHV-MID if the stressed syllable of the the base is **a mid vowel**, the stressed syllables of the base and output must correspond to each other and share the same vowel. (base = unsuffixed stem form)

MATCHV the stressed syllables of the base and output must correspond to each other and share the same vowel.

IDENT-OO-V if two vowels correspond segmentally, they must be the same

(simplifying a bit, and ignoring some complications…)

# Phonological grammar: constraints

MATCHV-MID if the stressed syllable of the input is **a mid vowel**, the stressed syllables of the input and output must correspond to each other and share the same vowel.

***General*** vowel matching constraint

MATCHV the stressed syllables of the base and output must correspond to each other and share the same vowel.

IDENT-OO-V if two vowels correspond segmentally, they must be the same

(simplifying a bit, and ignoring some complications…)

# Phonological grammar: constraints

MATCHV-MID — if the stressed syllable of the input is **a mid vowel**, the stressed syllables of the input and output must correspond to each other and share the same vowel.

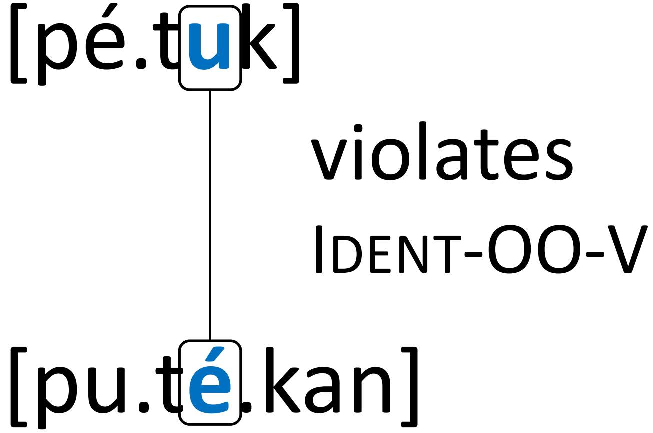
MATCHV — the stressed syllables of the base and output must correspond to each other and share the same vowel.

penalizes changes between ***segmentally*** corresponding segments

IDENT-OO-V — if two vowels correspond segmentally, they must be the same

(simplifying a bit, and ignoring some complications…)

# Phonological grammar: constraints



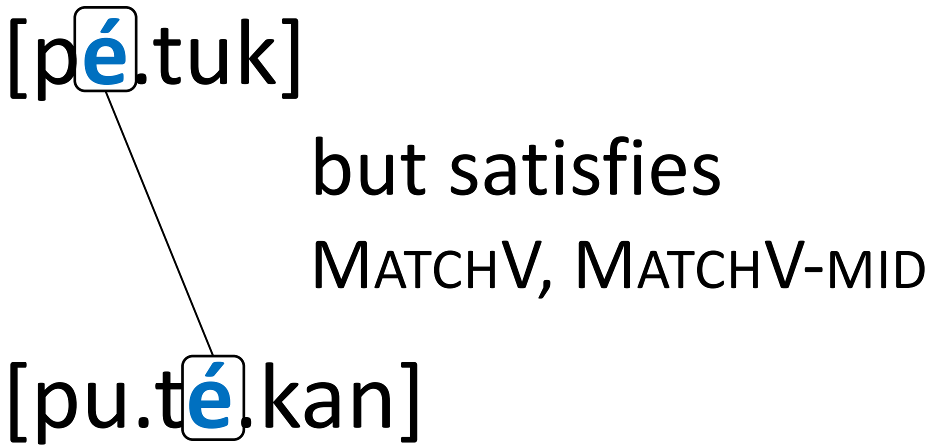
[pé.tuk]

violates

IDENT-OO-V

[pu.té.kan]

# Phonological grammar: constraints



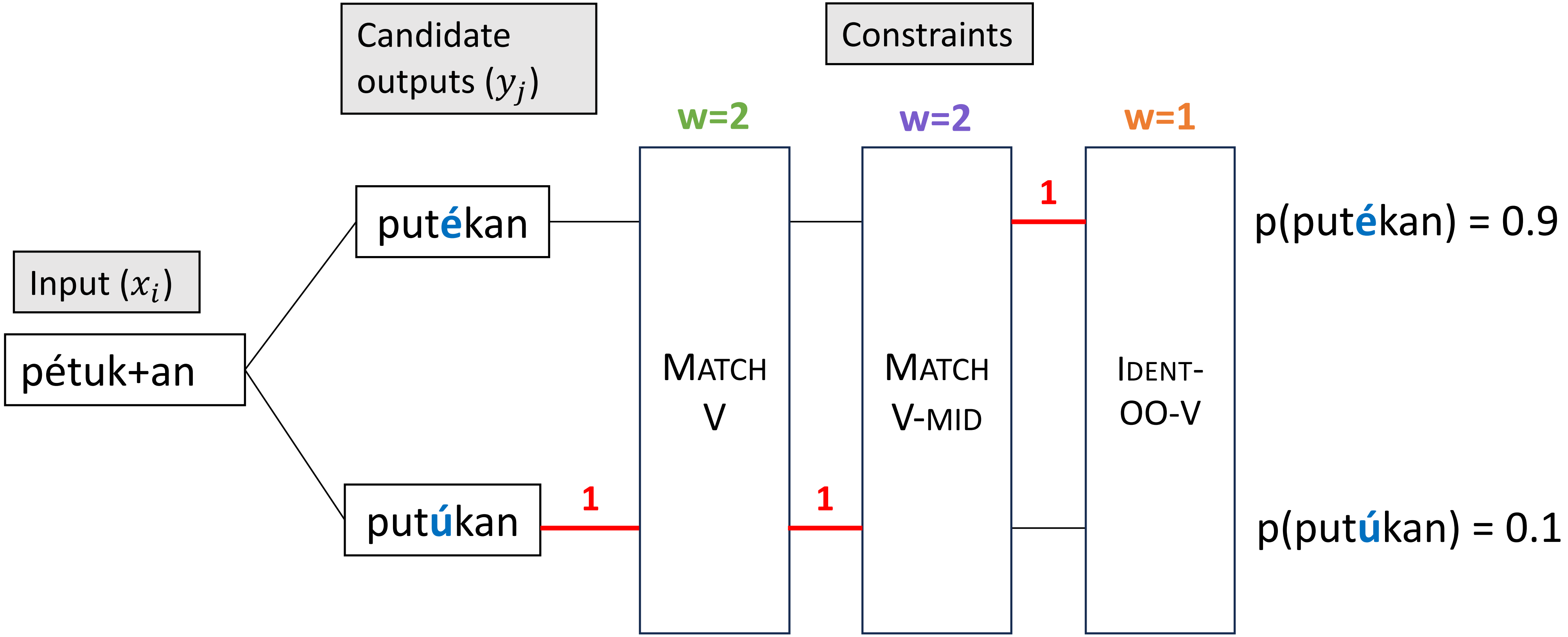
[pé.tuk]

[pu.té.kan]

but satisfies

MATCHV, MATCHV-MID

# A Seediq example (simplified)



Candidate outputs ($y_j$)

Constraints

Input ($x_i$)

pétuk+an

putékan

putúkan

w=2 Match V

w=2 Match V-mid

w=1 Ident-OO-V

1

1

1

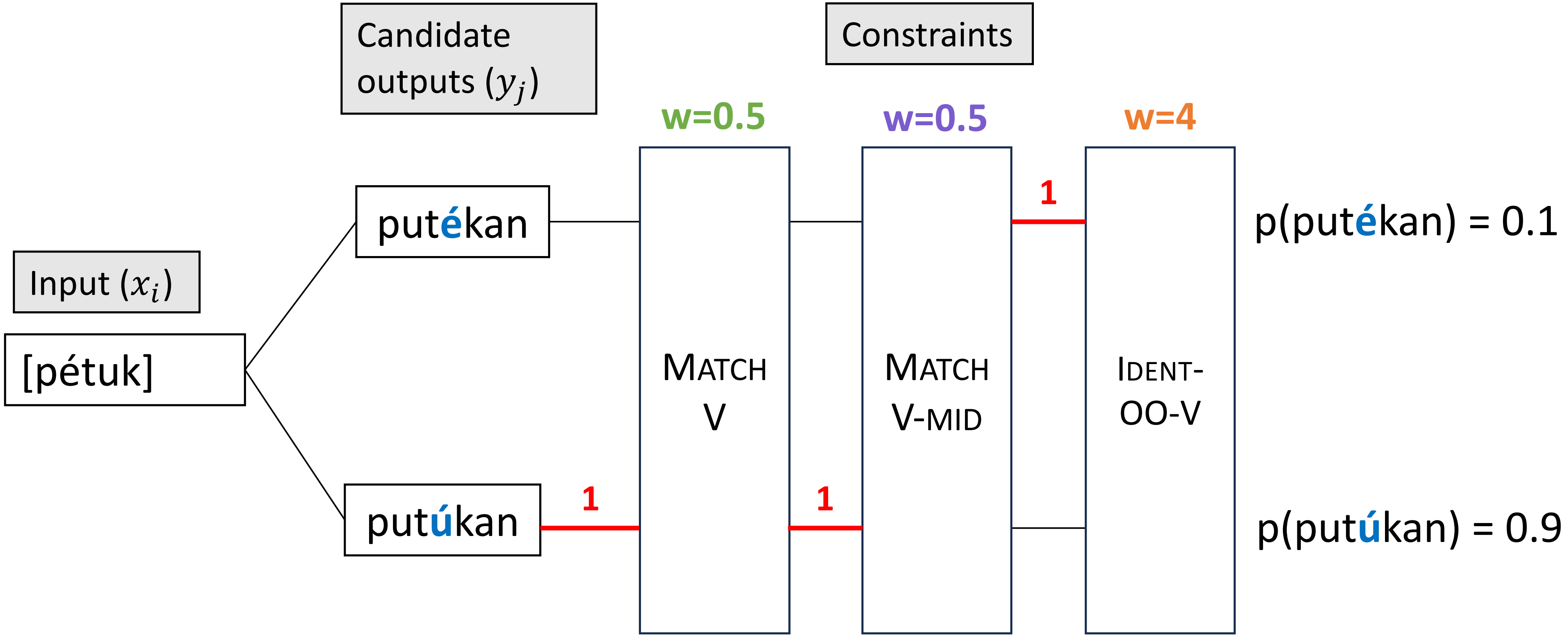
p(putékan) = 0.9

p(putúkan) = 0.1

**If w(MatchV, MatchV-mid) > w(Ident-V), the grammar will prefer** [putékan]

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

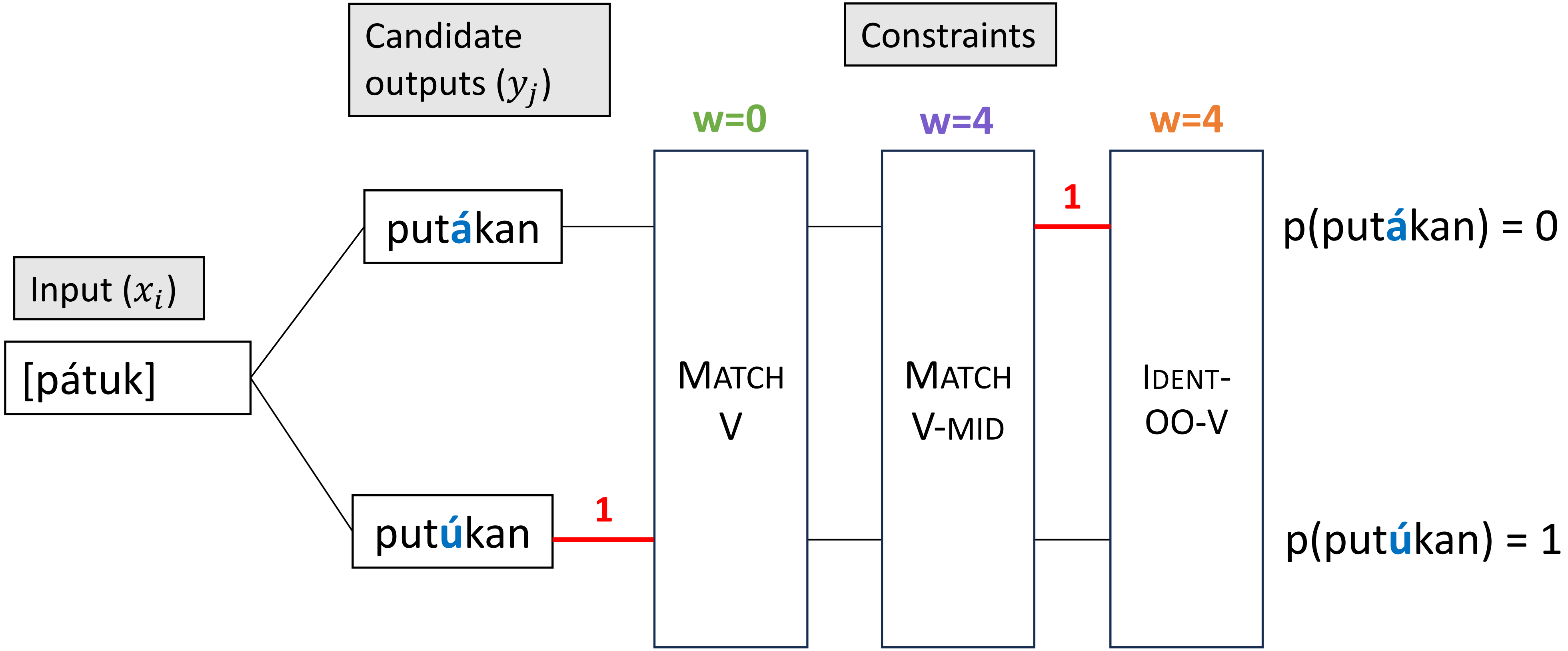
# A Seediq example (simplified)



If w(MatchV, MatchV-mid) > w(IDENT-V), the grammar will prefer [putékan]

**If w(IDENT-V) > w(MATCHV, MATCHV-MID), the grammar will prefer [putúkan]**

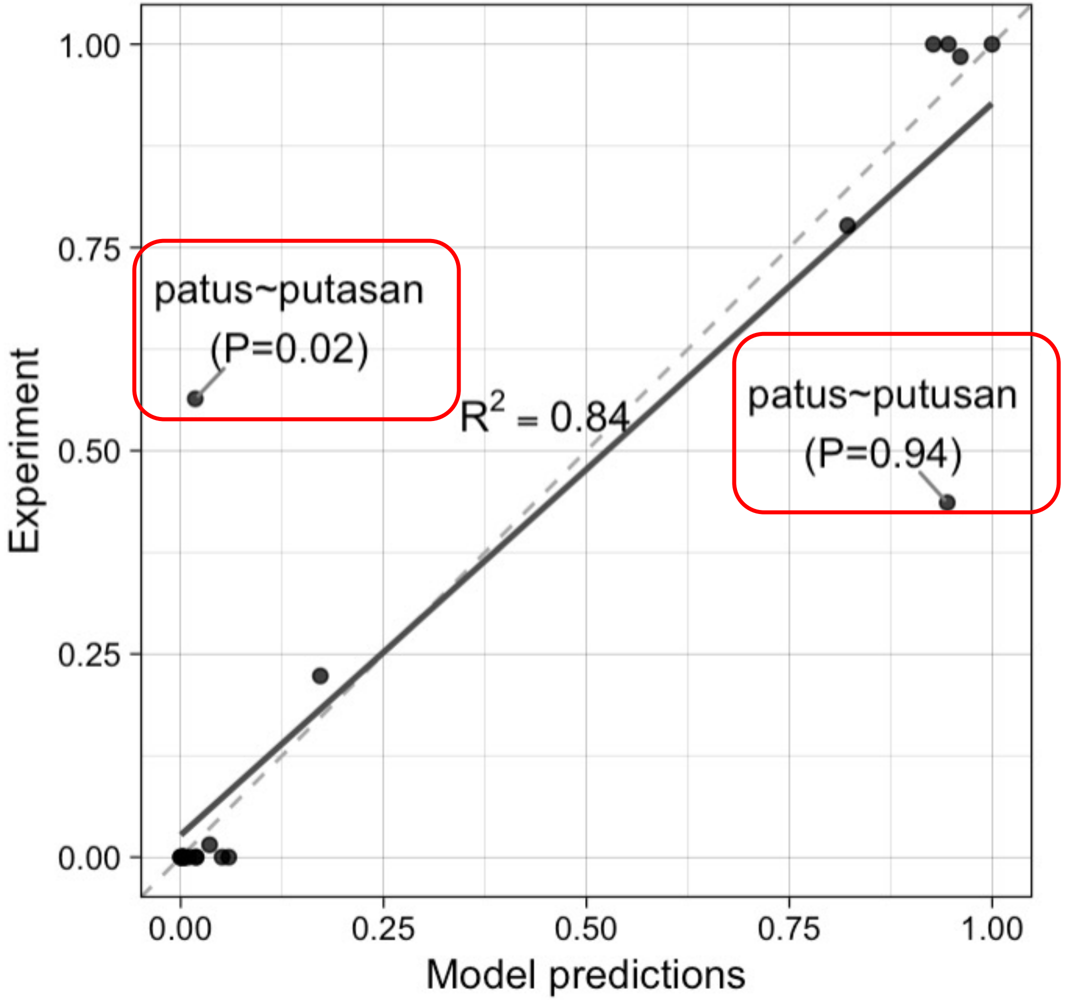
# A Seediq example (simplified)



For inputs like [pátuk], vowel-matching alternations happen only if the weight of MATCHV is high (MATCHV-MID doesn't apply)

# Results-model with no generality bias

- Model is frequency-matching…
  - but underpredicts [pátuk]~[putákan] type responses
- Reason:
  - [patuk]~[putákan] not observed in the lexicon (model input).
  - Model assigns high weight to MATCHV-MID, but **near-zero weight to MATCHV**



# Elements of the model

- A probabilistic phonological grammar
- **Ability to incorporate generality bias**

# Learning biases

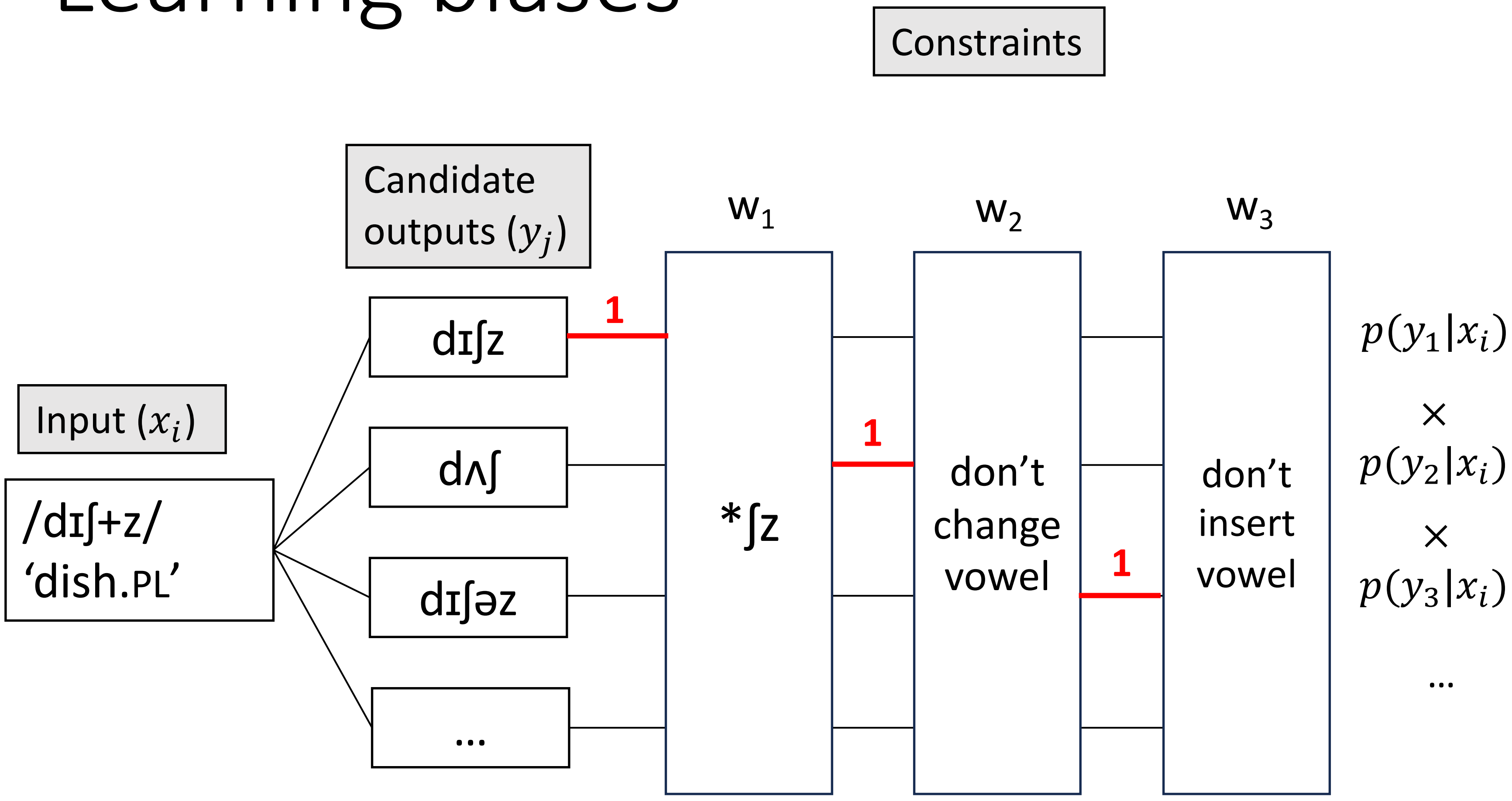
To implement a bias, we can give the model a **Gaussian prior** (Wilson 2006; Martin 2011; White 2013)

- Functionally equivalent to L2 regularization

Each constraint weight *w* is associated with a Gaussian distribution with, **mean (μ)=0** and a **standard deviation (σ)=1**.

$$\frac{(w_m - \mu)^2}{2\sigma^2} \quad \Rightarrow \quad \frac{(w_m - 0)^2}{4}$$

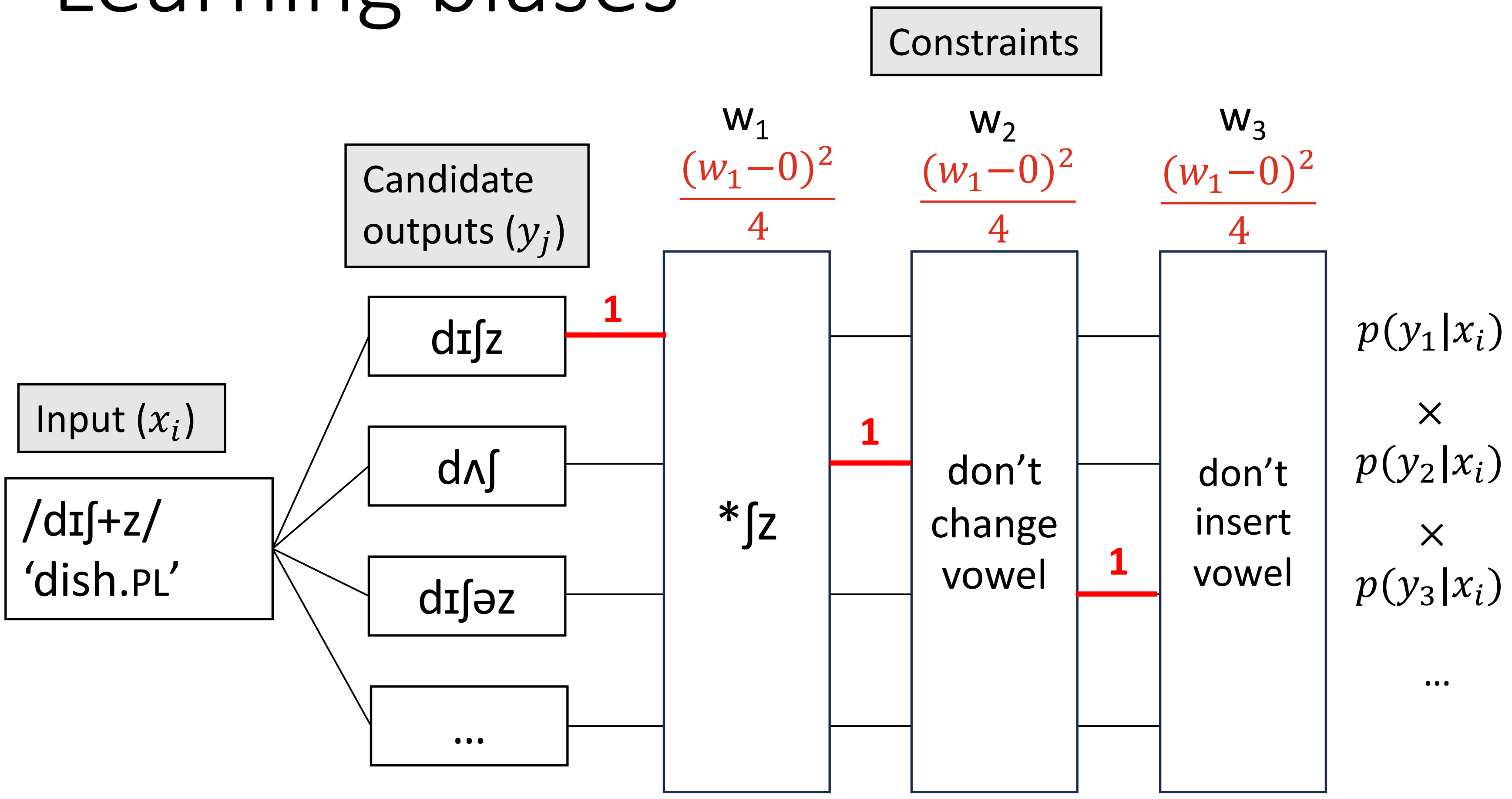
# Learning biases



**old objective function**

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

# Learning biases



**new objective function**

$$\sum_{n=1}^{N} \log\big(P(y_n|x_i)\big) - \boxed{\sum_{m=1}^{M} \frac{(w_m-0)^2}{4}}$$

The bigger this value, the bigger the penalty.

# Learning biases

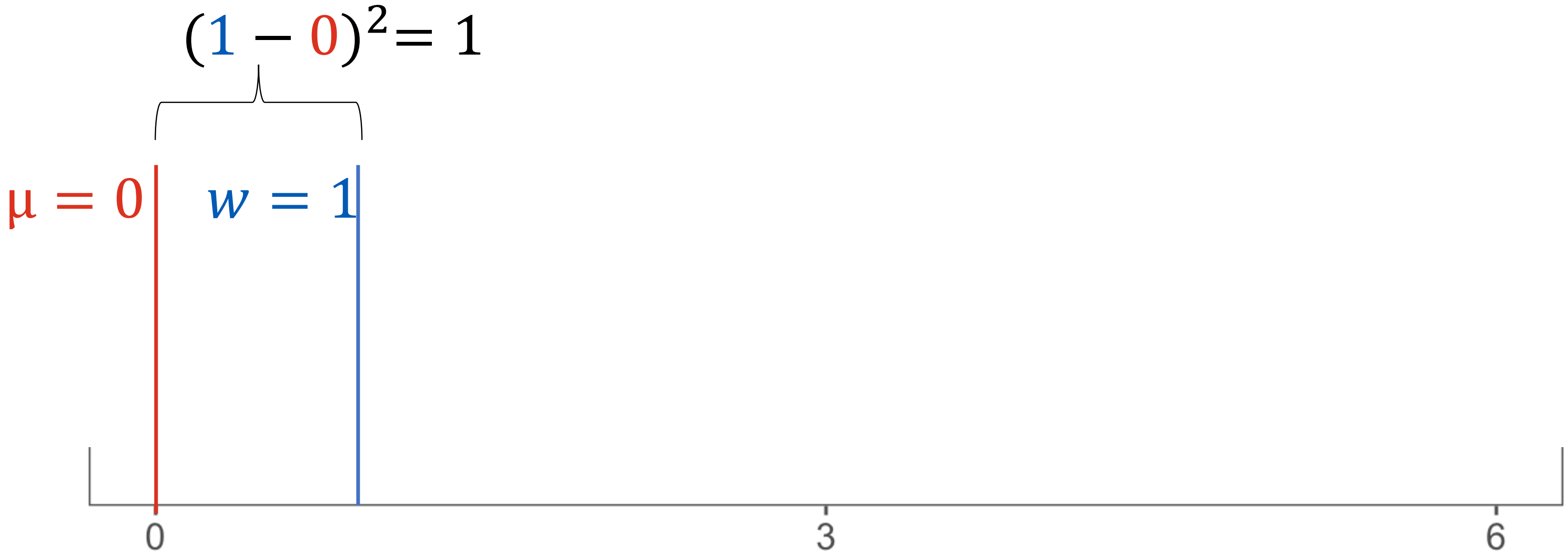
- As a result of this prior, the grammar will prefer to assign constraints uniform, low weights…
  - instead of assigning a lot of weight to one constraint
- Result: weight more evenly spread across **MATCHV** and **MATCHV-MID**

# low weight = low penalty

$$\text{Prior} = \frac{(w_m - 0)^2}{4}$$

w = constraint weight

μ = “preferred” weight

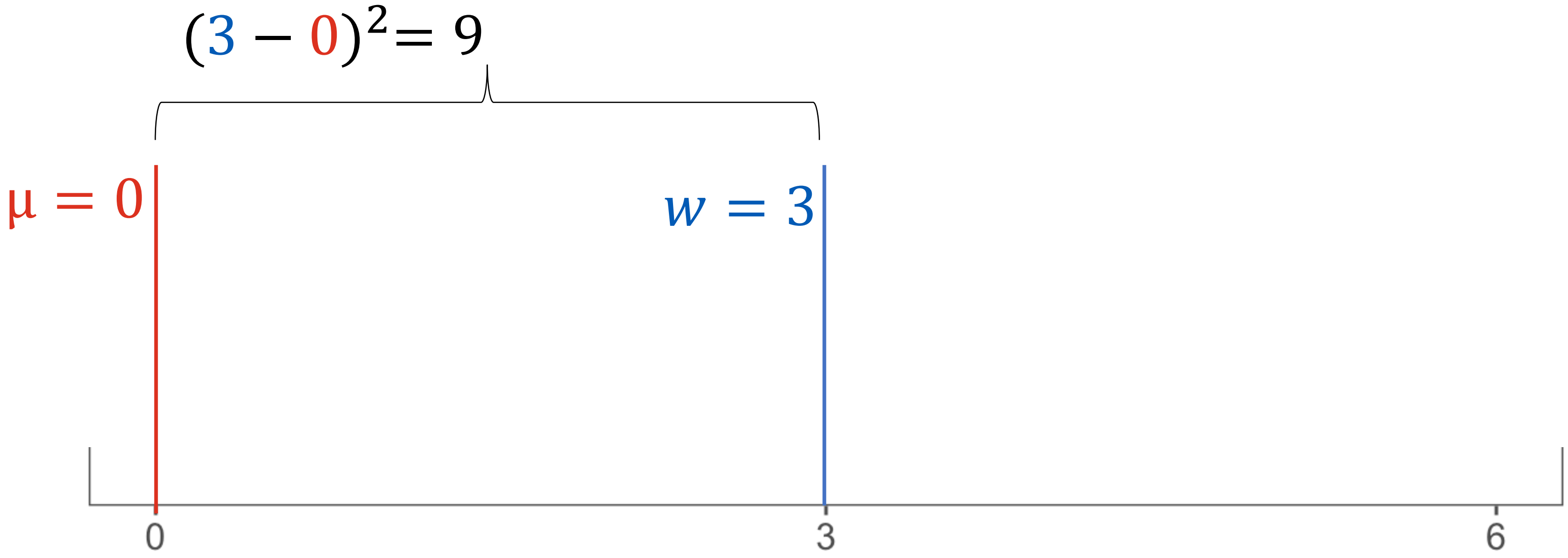


# high weight = exponentially higher penalty

$$\text{Prior} = \frac{(w_m - 0)^2}{4}$$

w = constraint weight

μ = “preferred” weight

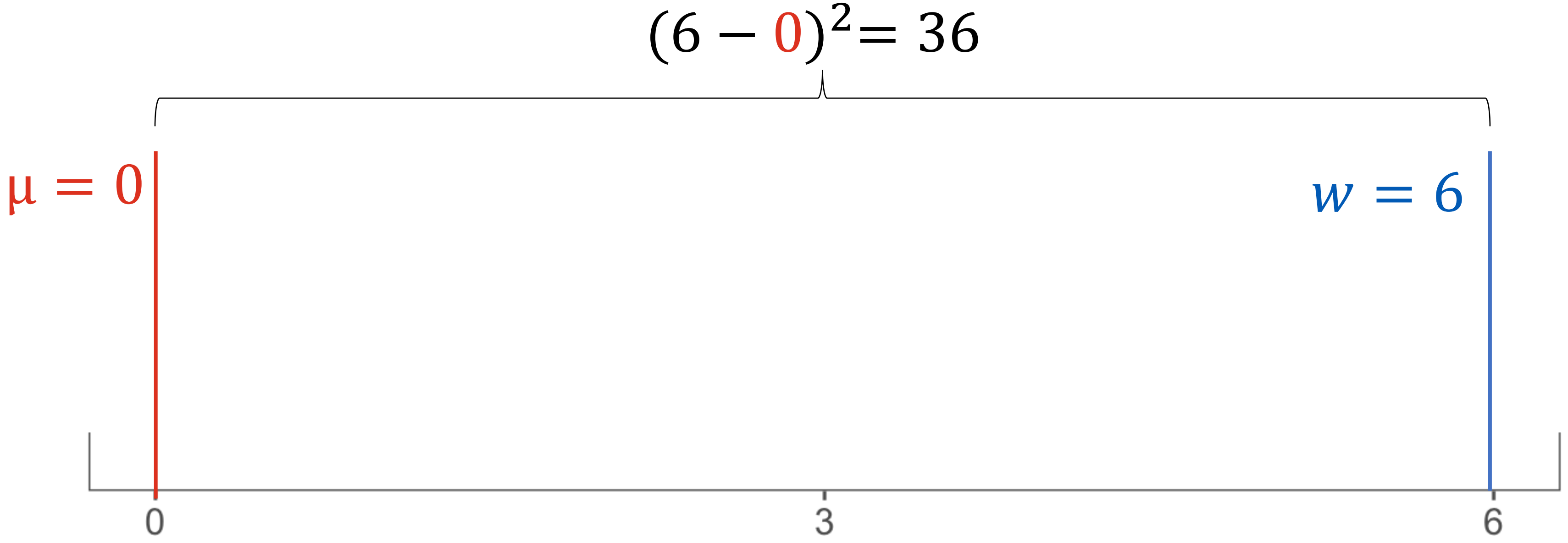


# high weight = exponentially higher penalty

$$\text{Prior} = \frac{(w_m - 0)^2}{4}$$

w = constraint weight

μ = “preferred” weight

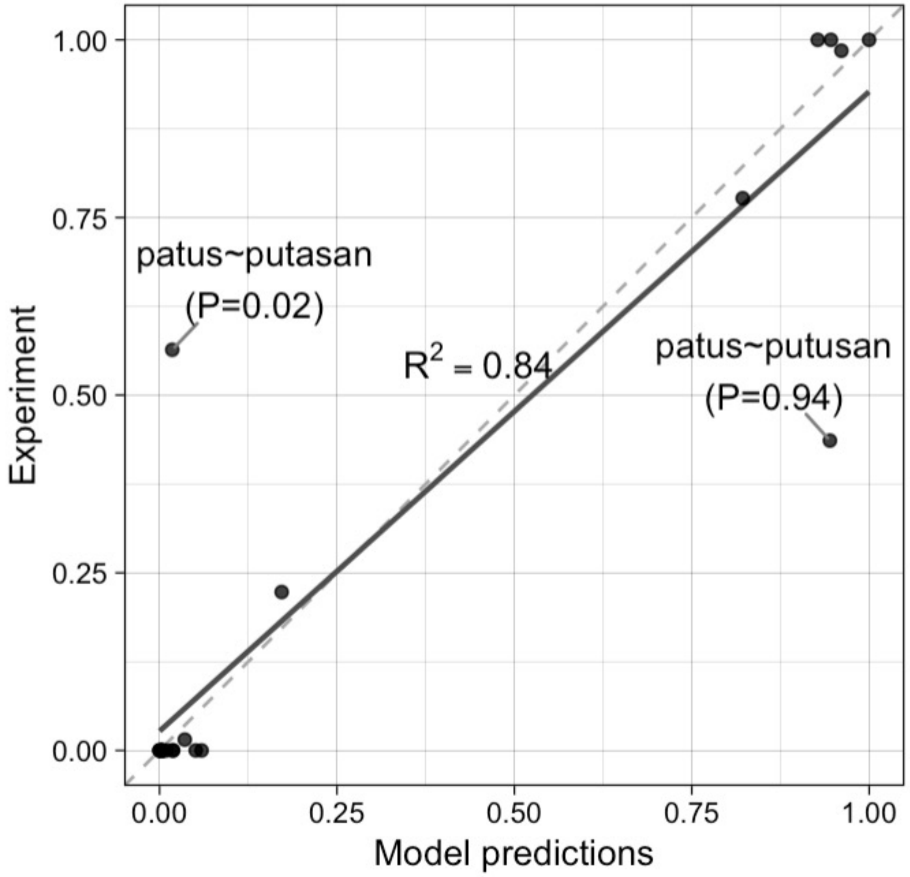


# Elements in a phonological model

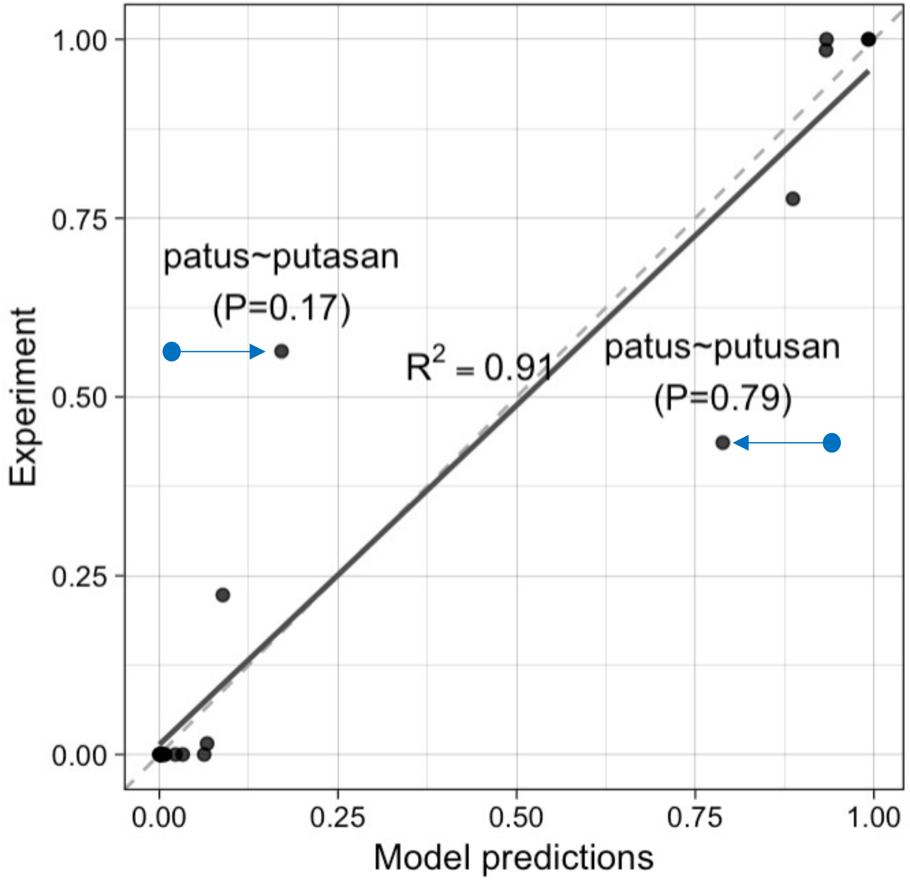
1. A probabilistic phonological grammar ✓
2. Ability to incorporate generality bias ✓

# Results

**Frequency matching** ($R^2$=0.84)



**Generality bias** ($R^2$= 0.91)



# Conclusion

- Seediq has a tendency towards vowel matching
    - where stressed vowels of related surface forms match each other.
    - **psychological reality:** productively applied to new words, and over-generalized beyond what is observed in the lexicon.
- How do we explain speakers' over-learning of vowel matching?
    - modeling results suggest a **generality bias.**

# Conclusion

- Why are there so few documented cases of prosodic correspondence?
  - Missed when we look at just UR→SR mappings
  - gradient pattern
- Importance of…
  - looking at the relations between related surface forms
  - looking at gradient phenomena when addressing issues about phonological representation

# Thank you! to…

Aking Nawi and other Seediq consultants

Bruce Hayes and Kie Zuraw, members of the UCLA Phonology Seminar, two anonymous reviewers of PDA, and

# References

- Albright, Adam & Bruce Hayes. 2003. Rules vs. analogy in english past tenses: A computational/experimental study. *Cognition* 90(2). 119–161. doi:10.1016/S0010-0277(03)00146-X.
- Berko, Jean. 1958. The child’s learning of English morphology. *Word* 14(2-3). 150–177. https://doi.org/10.1080/00437956.1958.11659661.
- Benua, Laura. 1995. Identity effects in morphological truncation. In S. Urbanczyk & JN. Beckman (eds.), *University of Massachusetts occasional papers 18: Papers in Optimality Theory,* 77– 136. Amherst: GLSA.
- Council of Indigenous Peoples. 2020. Online dictionary of aboriginal languages. http://e-dictionary.apc. gov.tw/Index.htm. Accessed: 2020-09-30.
- Crosswhite, Katherine. 1998. Segmental vs. prosodic correspondence in Chamorro. *Phonology* 15(3). 281–316. https://doi.org/10.1017/S0952675799003619.
- Elkins, N., & Kuo, J. (2023). A prominence account of the Northern Mam weight hierarchy. In *Supplemental Proceedings of AMP 2022*.
- Ernestus, M.T., Baayen, R.H., 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79, 5–38.
- Hayes, B., Siptá r, P., Zuraw, K., Londe, Z., 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* , 822–863.
- Kuo, J. (2020). *Evidence for base-driven alternation in Tgdaya Seediq*. UCLA.
- Kuo, J. (2023). Evidence for prosodic correspondence in the vowel alternations of Tgdaya Seediq. *Phonological Data and Analysis* 5(3), p. 1-31. https://doi.org/10.3765/pda.v5art3.77
- Martin, Andrew. 2011. Grammars leak: Modeling how phonotactic generalizations interact within the grammar. Language 87(4). 751–770. https://doi.org/10.1353/lan.2011.0096.
- McCarthy, John J. & Alan S. Prince. 1995. Faithfulness and Reduplicative Identity. In Jill N. Beckman, Laura Walsh Dickey & Suzanne Urbanczyk (eds.), Papers in optimality theory, 249–384. Amherst: GLSA.
- White, James. 2017. Accounting for the learnability of saltation in phonological theory: A maximum entropy model with a p-map bias. *Language* 93(1). 1–36. https://doi.org/10.1353/lan.2017.0001.
- Withgott, Meg. 1983. *Segmental evidence for phonological constituents*, University of Texas, Austin dissertation.
- Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science* 30(5). 945–982. https://doi.org/10.1207/s15516709cog0000 89.

- Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science* 30(5). 945–982. https://doi.org/10.1207/s15516709cog0000 89.

- Yang, Hsiu-fang. 1976. The phonological structure of the paran dialect of Sediq. *Bulletin of the Institute of History and Philology Academia Sinica* 47(4). 611–706.
- Zuraw, K., 2000. *Patterned exceptions in phonology*. Ph.D. thesis. University of California, Los Angeles.