

# Phonological influences on lexicalized compound formation in Māori

SIMON TODD, JEREMY NEEDLE,  
JEN HAY & JEANETTE KING

# Big picture

Language is constantly changing

Recent work: **evolutionary perspective** on the lexicon

- Words compete for survival

“Survival of the fittest”

- What survives → what is “**linguistically fit**”

We study lexicalized compounds in Māori

- **What gets to lexicalize, and why?**

# Why compounds?

## Competition

- *blackboard* ~ *chalkboard*

## Lexical creation

- From known pieces
- Wide-ranging possibilities

## Combine influences parallel to syntax / morphology

- Component ordering (cf. syntactic alternations)
- Component choice (cf. allomorphy)



# Why Māori?

Small phoneme inventory

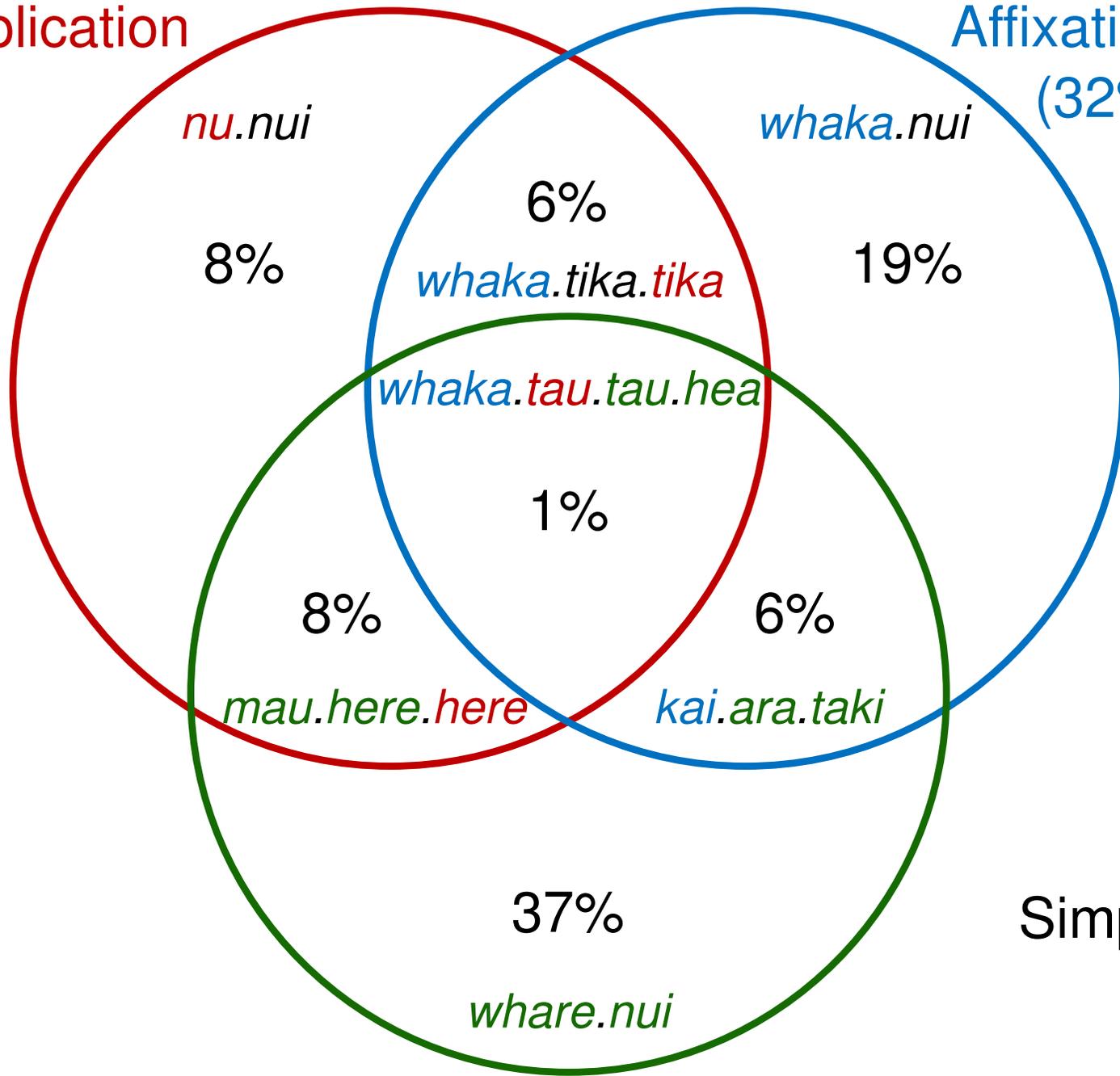
+ Preference for short morphemes

+ Limited morphological processes

→ Extremely high prevalence of compounds

Reduplication  
(23%)

Affixation  
(32%)



Simplex: 16%

Compounding  
(52%)

# Why Māori?

Extremely high prevalence of compounds (~50%)

Revitalization: **new word creation**

- compounding = “principle productive means”  
(Bauer, 1993)

# Why Māori?

Extremely high prevalence of compounds (~50%)

Revitalization: new word creation

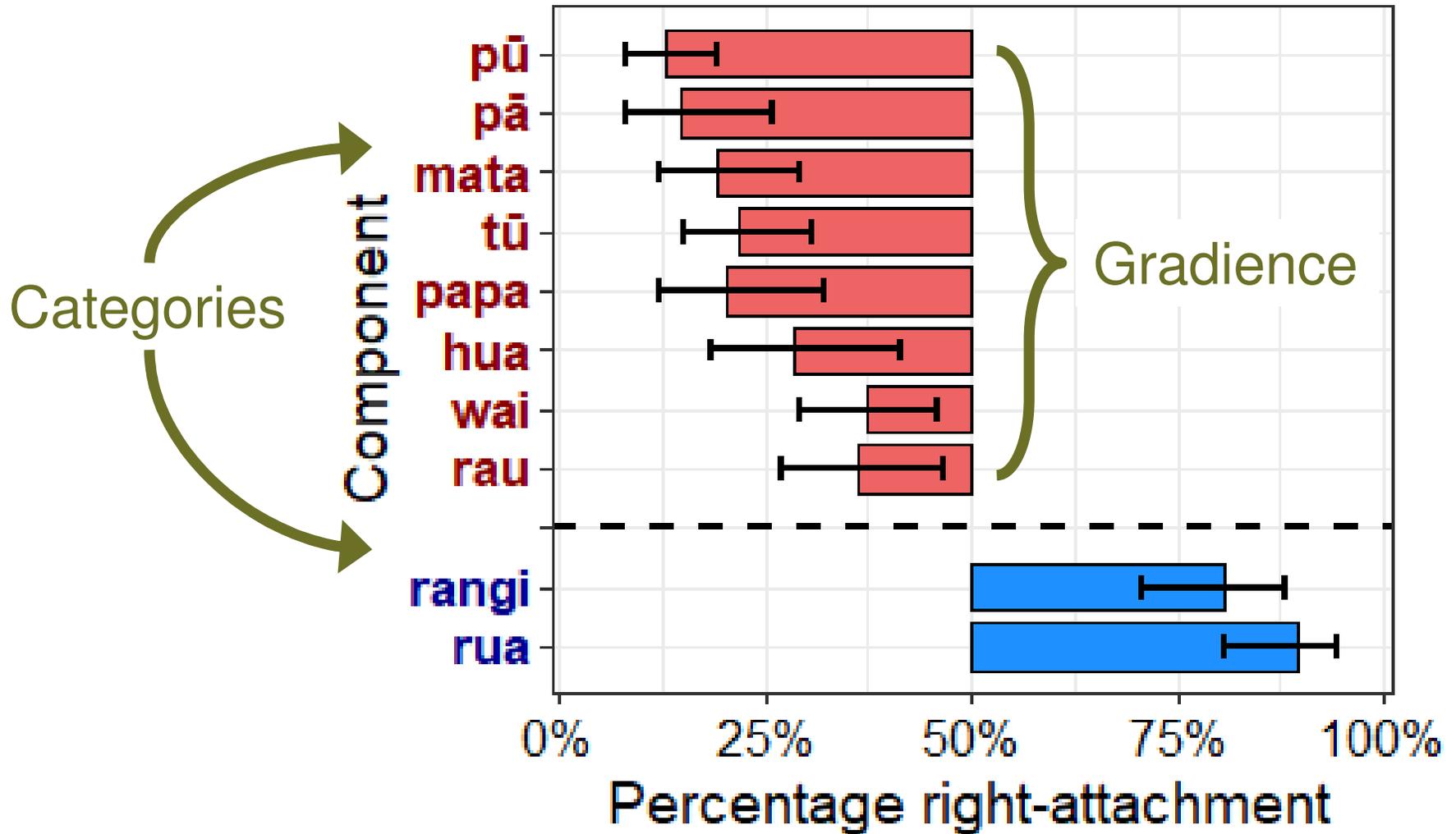
POS flexibility → compound structure ≠ syntax

<i>mate</i>	~ noun	‘death’
	~ verb	‘to die’
	~ adjective	‘dead’

Same component can attach on either side

<i>matepā</i> ‘hazard’	<i>waimate</i> ‘stagnant’
------------------------	---------------------------

# Ordering preferences: 10 top components



# Research questions

1. Assuming components may be flexibly ordered, what influences the order that lexicalizes?

*mā* ('white') + *whero* ('red') = *māwhero* ('pink')

**Why not *wheromā*?**

2. What influences the selection of components?

*tea* ~ 'white' also. **Why not *teawhero*?**

Our focus: phonological factors

# Why phonological factors?

Known influences at the word and phrase levels

→ Potential to be codified in compounds

- Stress: seeks out long vowels, left-aligned

*kekē*

*kēkē*

*keke*

→ Idea: components with long vowels come first

# Why phonological factors?

Known influences at the word and phrase levels

→ Potential to be codified in compounds

- Stress: seeks out long vowels, left-aligned
- End-weight: light phrases before heavy when possible

*Ka puta katoa ki waho ngā tāngata o te pā...*

T/A exit all to outside the.PL people of the.SG pā...

‘All the people of the pā came outside...’ (Bauer, 1993; ex. 981)

→ **Idea:** light components before heavy

# Why phonological factors?

Known influences at the word and phrase levels

→ Potential to be codified in compounds

- Stress: seeks out long vowels, left-aligned
- End-weight: light phrases before heavy when possible
- Vowel harmony and consonant disharmony in words

(Rácz et al., 2016)



→ **Idea:** cross-component phonotactics influence order

# Data: source

19,595 words from *Te Aka* dictionary

Exclude:

- loanwords
- passives
- closed-class words

→ 12,441 words with unknown morphological structure

Which are compounds? Need human annotation

# Data: annotations

2 young adult fluent speakers of Māori

Words split into random lists of ~1k (**written**)

## Task:

- Identify the **parts** (if any) that make up the word
- Based on what “most Māori speakers would think”
- Separate parts by .
- Can indicate when words / parts are unknown

# Data: compounds

Focus: binary compounds

Extract words that:

- are known to both speakers
- are decomposed identically by both speakers
- have exactly 2 parts
- do not contain a known affix
- do not follow a known reduplication template

→ 2,711 clear binary compounds

# Analysis

Four kinds of factors

- Possible POS of component
  - Type frequency of component
  - Phonological properties of component
  - Cross-component phonotactics
- } Controls  
(~headedness)

Consider cases where components **differ** wrt factors

**Statistics:** mixed-effects logistic regression

- Randomly scramble order; predict unscrambled
- All factors + random intercept for component

# Results: controls

Consistent with expectations from headedness

- Generally: head left, modifier right

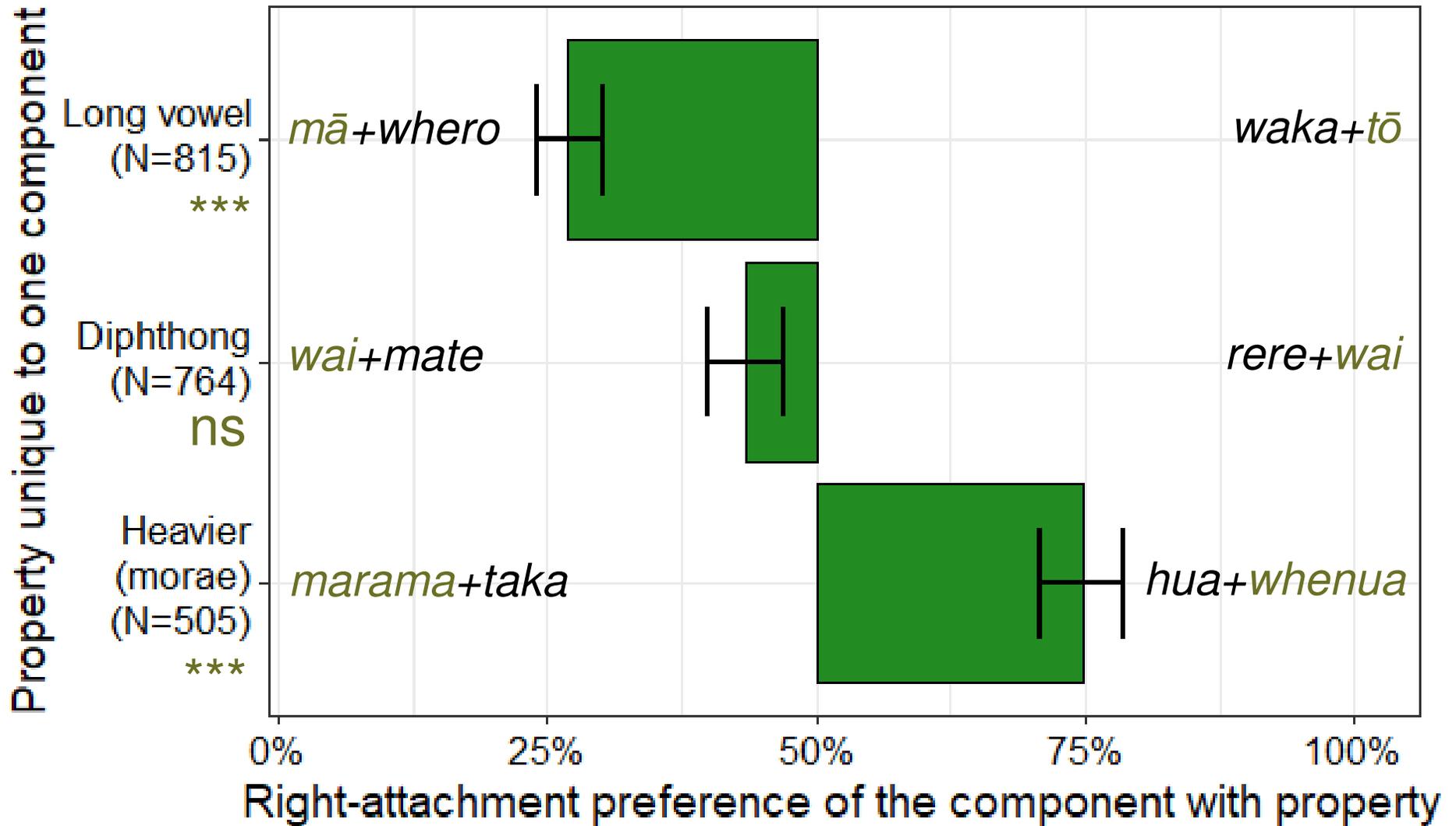
Possible to use component as a modifier

→ stronger right-attachment preference

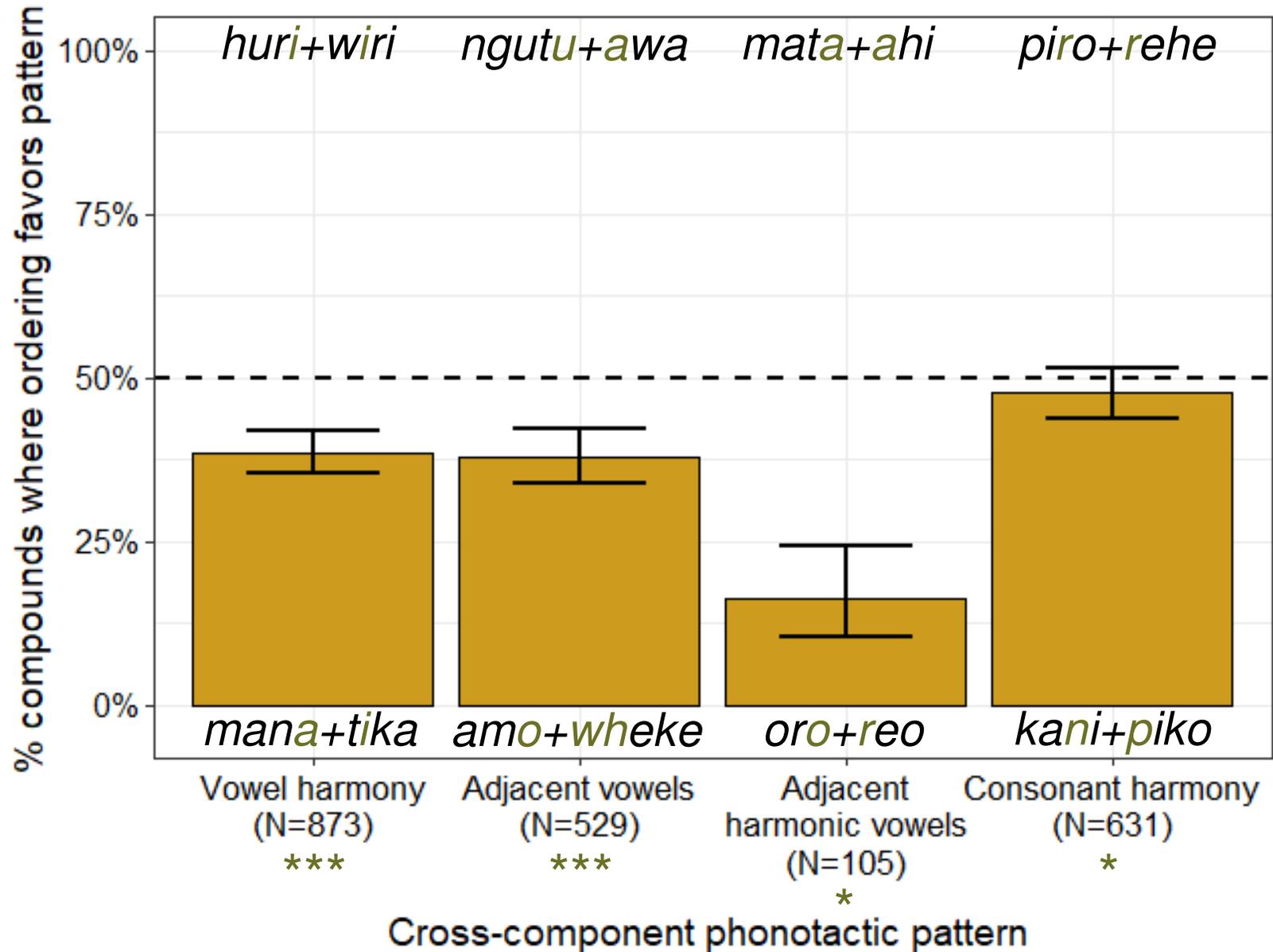
Component occurs in many compounds

→ stronger left-attachment preference

# Results: component phonology



# Results: cross-component phonotactics



# A closer look at vowel harmony

Current result:

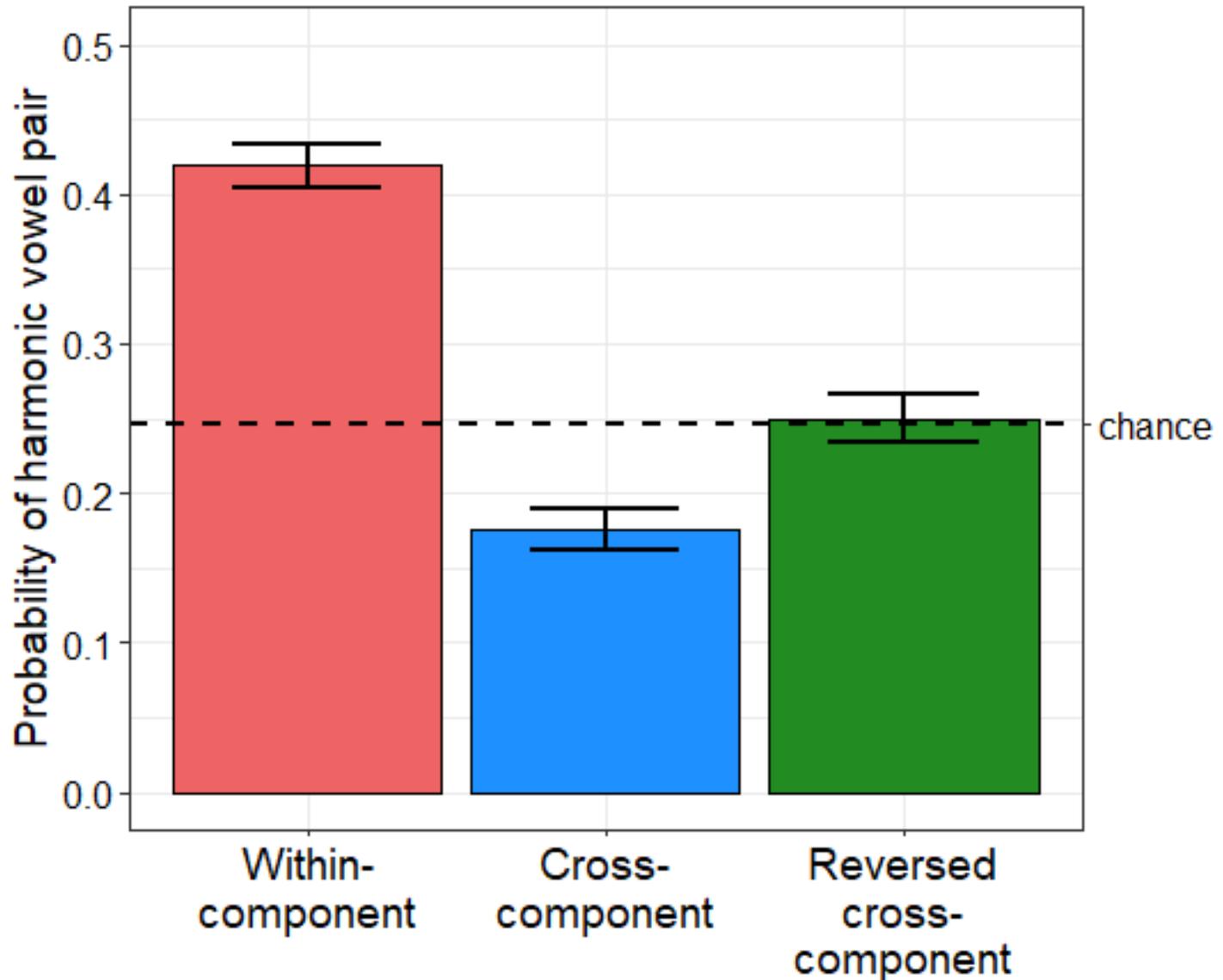
- Compound orderings tend to **avoid** vowel harmony

In spite of previous result (Rácz et al., 2016):

- Words tend to **favor** vowel harmony

**Possibility:** anti-harmony ~ divergence from norm  
→ morpheme boundary cue

# A closer look at vowel harmony



# General discussion

We have focused on RQ1: component ordering

- What about RQ2: component selection?

*Assumption:*

compound “fitness” ~

conceptual appropriateness of components

× strength of component ordering preference

Given similarly appropriate component candidates

→ selection follows influences on ordering preferences

# General discussion

How do the phonological influences play out?

A. When choosing between variants already in use

- 
- Subtle, gradual, at population level
  - Compounds may be generated blind to phonology
  - Phonology filters candidates in usage

Z. When coining new compounds

- Less subtle, instantaneous, individualized
- Compound generation is sensitive to phonology

# Summing up

Lexicon enshrines what is “linguistically fit”

- Victors of **competitions** between potential wordforms

In Māori compounds, phonology influences fitness

- **Codification** of influences on words / phrases
- **Demarcation** of composite nature

Two big-picture takeaways:

- Evolutionarily, **phonology can influence morphology**
- Lexicalized compounds provide **unique insights**

Thank you for listening!

Questions / comments?

[sjtodd@ucsb.edu](mailto:sjtodd@ucsb.edu)