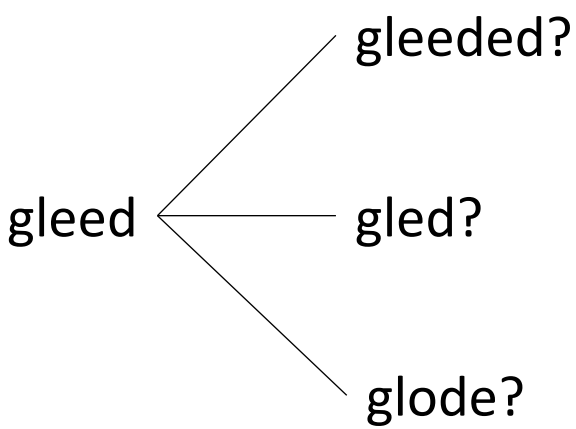
Markedness effects in the history of Samoan thematic consonants

Jennifer Kuo, Cornell University

AFLA 30, Lund University

Background

- Paradigms with conflicting data patterns can be difficult to learn



PRESENT	PAST
laugh	laugh**ed**
dance	danc**ed**
jump	jump**ed**
bleed [blid]	bl**e**d [bl**ɛ**d]
ride [raɪd]	r**o**de [r**oʊ**d]
go	**went**

Background

- Paradigms with conflicting data patterns can be difficult to learn
- ….resulting in learning errors.

  go → **goed*

- **Reanalysis**: when such errors are adopted into speech community, resulting in a type of language change.

help/h**a**lp → help/help**ed** (c1300)
dive, div**ed** → dive, d**o**ve (c1800)

PRESENT	PAST
laugh	laugh**ed**
dance	danc**ed**
jump	jump**ed**
bleed [blid]	bl**e**d [bl**ɛ**d]
ride [raɪd]	r**o**de [r**oʊ**d]
go	**went**

Background

- **Question:** How do learners decide the direction of reanalysis?

  Frequency-matching: apply patterns at the proportion in which they occur in the paradigm (e.g. Eddington 1996; Coleman and Pierrehumbert 1997; Zuraw 2000; Ernestus & Baayen 2003; Albright & Hayes 2003; Hayes et al. 2009; Jun & Lee 2007, etc)

  → Existing **models of reanalysis** are frequency-matching

  Ex: Minimal Generalization Learner (Albright & Hayes 2003)

  Analogical models (Nosofsky 1990; Hare & Elman 1995)

Background

- **Question:** How do learners decide the direction of reanalysis?

Frequency-matching: apply patterns at the proportion in which they occur in the paradigm (e.g. Eddington 1996; Coleman and Pierrehumbert 1997; Zuraw 2000; Ernestus & Baayen 2003; Albright & Hayes 2003; Hayes et al. 2009; Jun & Lee 2007, etc)

+

Markedness bias: preference for surface forms that are less marked (Jarosz 2006)

…where markedness effects are restricted to those already **active** in language-specific stem phonotactics

Why “active” markedness?

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

  English ex: *[**ʃs**ak] in roots

  *[dɪ**ʃ-s**] ‘dish + PL’ in morphophonology

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

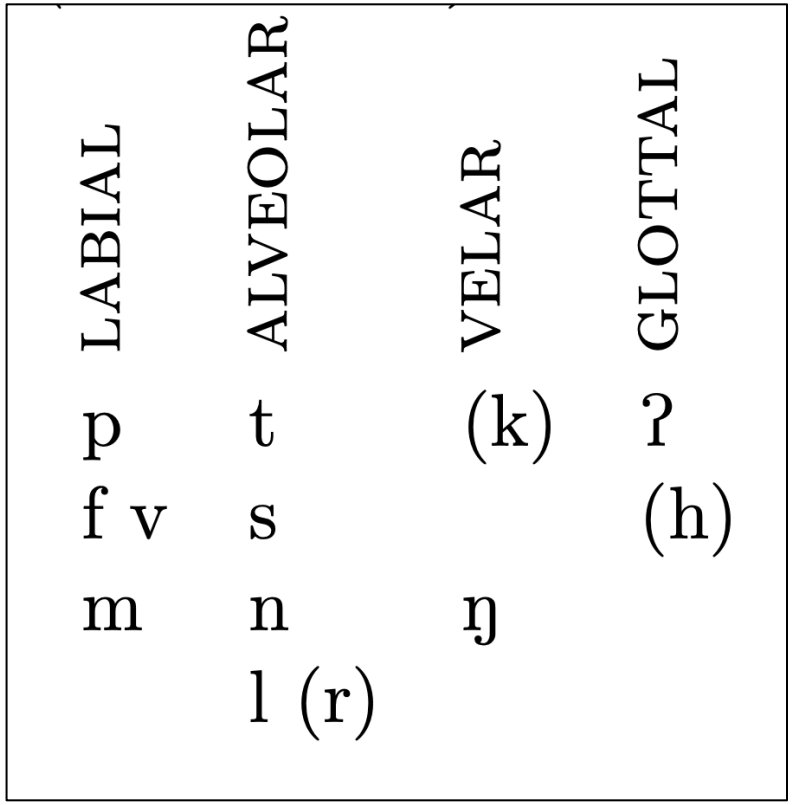
Goals of the talk

- Show that in Samoan, reanalysis is driven by both:
  - Frequency-matching
  - the avoidance of marked outputs
- Results are confirmed using a constraint-based model of reanalysis.

Background: Samoan

- Oceanic, Polynesian
- (C)V(V) syllable structure
- Vowels: /a, e, i, o, u/ (contrastive in length)

Consonant Inventory



Background: Samoan

- **Thematic consonant alternations (Ø~C)**: under suffixation, a consonant of unpredictable quality may surface
- **Example**: ergative suffix allomorphy

ERG.	STEM	SUFFIXED	GLOSS
a	lele	lele-a	‘to take’
ia	fana	fana-ia	‘to shoot’
ina	iloa	iloa-ina	'to see, perceive’
tia	pulu	pulu-**t**ia	‘to plug up’
sia	laka	laka-**s**ia	‘to step over’
ŋia	tutu	tutu-**ŋ**ia	'to light a fire’
fia	utu	utu-**f**ia	‘to draw water’
mia	inu	inu-**m**ia	‘to drink’
lia	tautau	tautau-**l**ia	‘to hang up’
na	aŋi	aŋi-**n**a	‘to blow’

Background: Samoan

/a, ia, ina/: vowel-initial allomorphs

- Oceanic language of the Polynesian subgroup
- **Thematic consonant alternations (Ø~C)**: under suffixation, a consonant of unpredictable quality may surface

ERG.	STEM	SUFFIXED	GLOSS
a	lele	lele-a	‘to take’
ia	fana	fana-ia	‘to shoot’
ina	iloa	iloa-ina	'to see, perceive’
tia	pulu	pulu-**t**ia	‘to plug up’
sia	laka	laka-**s**ia	‘to step over’
ŋia	tutu	tutu-**ŋ**ia	'to light a fire’
fia	utu	utu-**f**ia	‘to draw water’
mia	inu	inu-**m**ia	‘to drink’
lia	tautau	tautau-**l**ia	‘to hang up’
na	aŋi	aŋi-**n**a	‘to blow’

Background: Samoan

- Oceanic language of the Polynesian subgroup
- **Thematic consonant alternations (Ø~C)**: under suffixation, a consonant of unpredictable quality may surface

/Cia/: consonant-initial allomorphs

ERG.	STEM	SUFFIXED	GLOSS
a	lele	lele-a	‘to take’
ia	fana	fana-ia	‘to shoot’
ina	iloa	iloa-ina	'to see, perceive’
tia	pulu	pulu-**t**ia	‘to plug up’
sia	laka	laka-**s**ia	‘to step over’
ŋia	tutu	tutu-**ŋ**ia	'to light a fire’
fia	utu	utu-**f**ia	‘to draw water’
mia	inu	inu-**m**ia	‘to drink’
lia	tautau	tautau-**l**ia	‘to hang up’
na	aŋi	aŋi-**n**a	‘to blow’

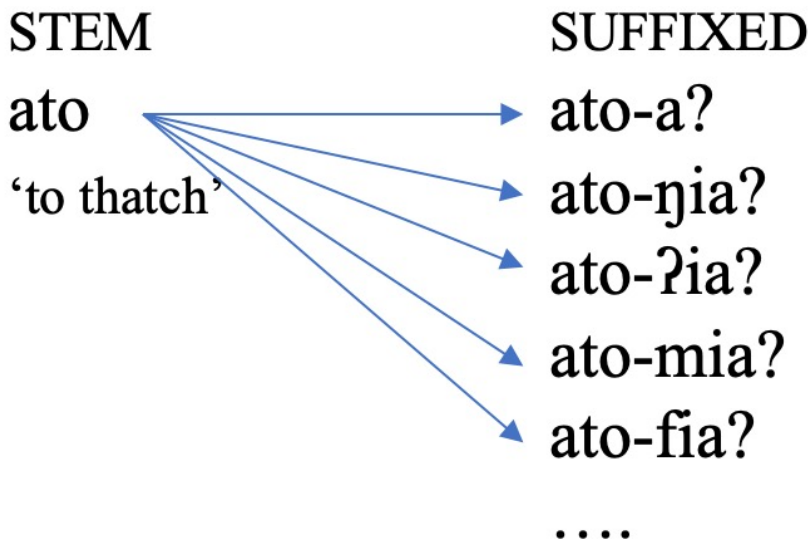
Development of Ø~C alternations

	‘TO DRINK’		'TO PLUG UP’		‘TO RUB’	
	STEM	SUFFIXED	STEM	SUFFIXED	STEM	SUFFIXED
Proto-Oceanic (POC)	inu**m**	inu**m**+ia	pulu**t**	pulu**t**+ia	pulu	pulu+ia
i→Ø/__a	-		-			pulua
Proto-Polynesian (PPn)						
C→Ø/__#	inu	inu**m**ia	pulu	pulu**t**ia	-	
Samoan	inu	inu+**m**ia	pulu	pulu+**t**ia	pulu	pulu+a

Deletion of final consonants in PPn made suffix allomorphy unpredictable

Background: reanalyses in Samoan

- Conflicting patterns make learning difficult



…Leading to reanalyses

POC	Expected	Actual	Reanalysis
*qato**p**	ato-**f**ia	ato-**a**	f→Ø
*ako**t**	aʔo-**t**ia	aʔo-**ina**	t→Ø
*quli**n**	uli-**n**a	uli-**ŋ**ia	n→ŋ

Question: What are the factors driving reanalysis of thematic consonants in Samoan?

Preview of results

- Reanalysis is generally towards the more frequent allomorphs
  - →But **phonotactically marked** outputs are more likely to be reanalyzed
  - →How so? Samoan roots are subject to transvocalic OCP-place effects

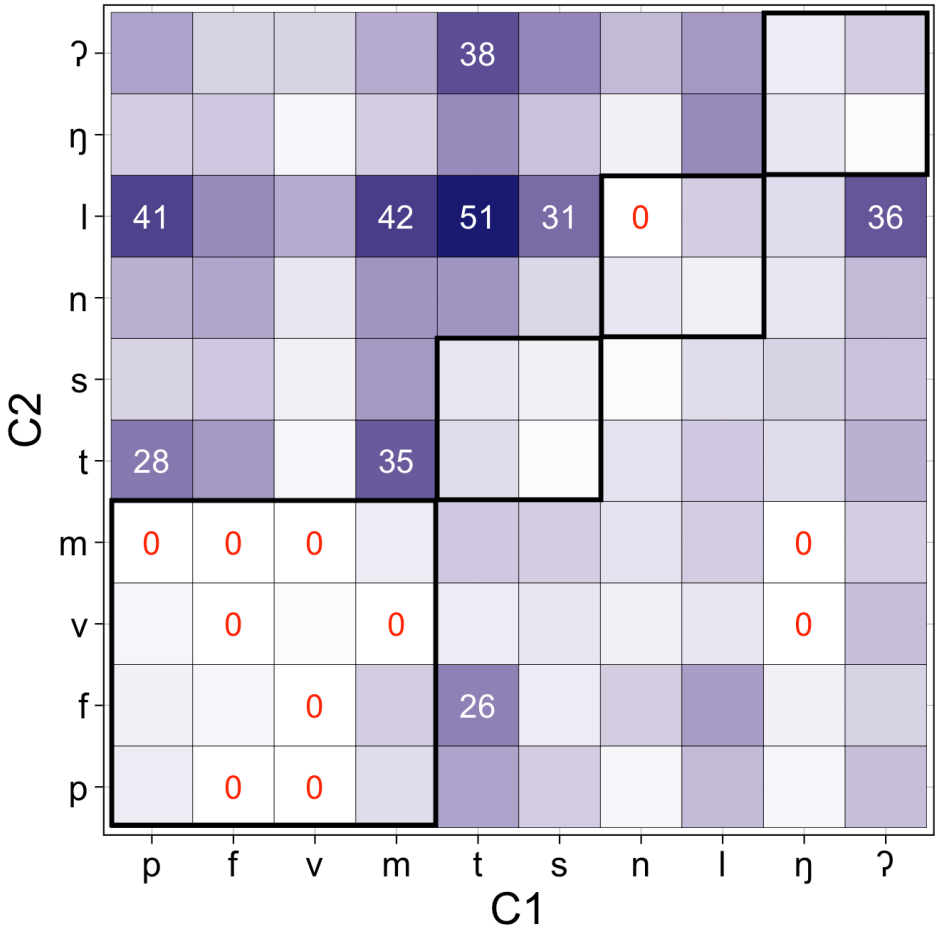
    e.g. *[**p**u**f**a] (***p…f**)

OCP-place effects in Samoan

- **Data:** 1498 roots (Alderete & Bradshaw 2013, originally from Milner 1966)

CONSTRAINT	EXAMPLE
OCP-LABIAL	**m**a**p**a, **m**a**f**u
OCP-CORONAL-SONORANT	**n**a**n**u, **l**a**n**u
OCP-DORSAL	**ŋ**a**ʔ**o, **ʔ**a**ʔ**e

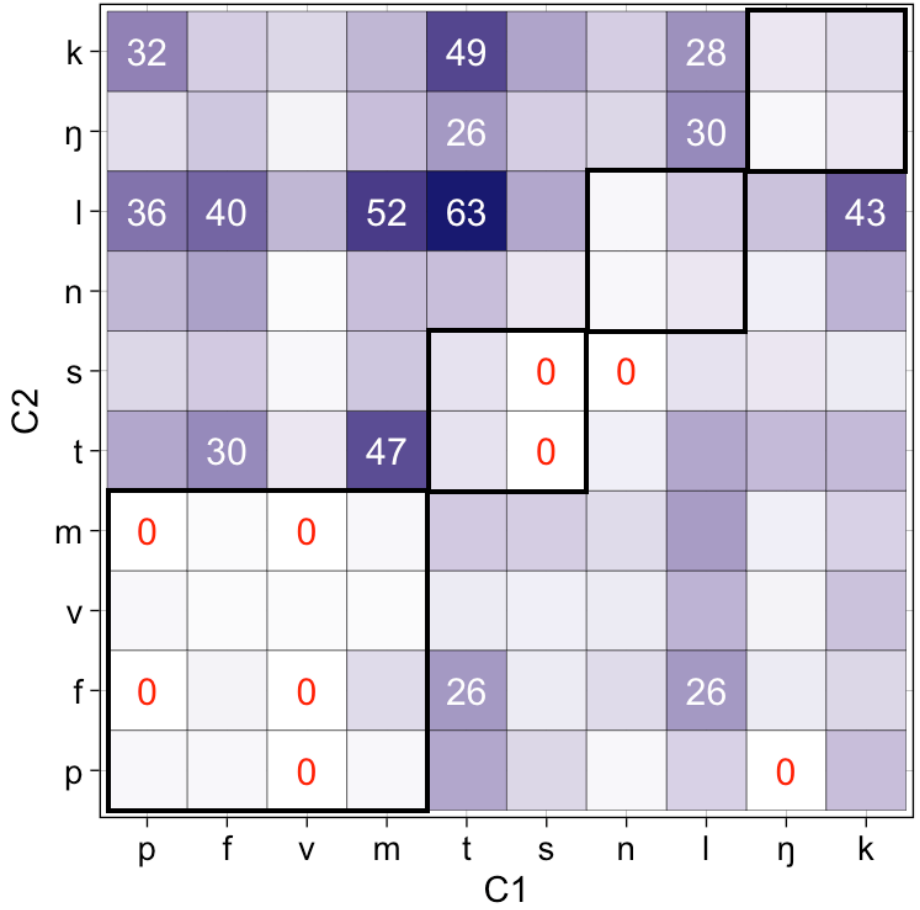
- Active OCP-place constraints based on a Maxent phonotactic model (Wilson & Obdeyn 2009)
- PPn *k > Samoan [ʔ], but [ŋ] and [ʔ] still behave as a natural class.



OCP-place in Proto-Polynesian

- **Data:** Protoforms from the Polynesian Lexicon Project (POLLEX; Greenhill & Clark 2011).
- Same patterns found in PPn

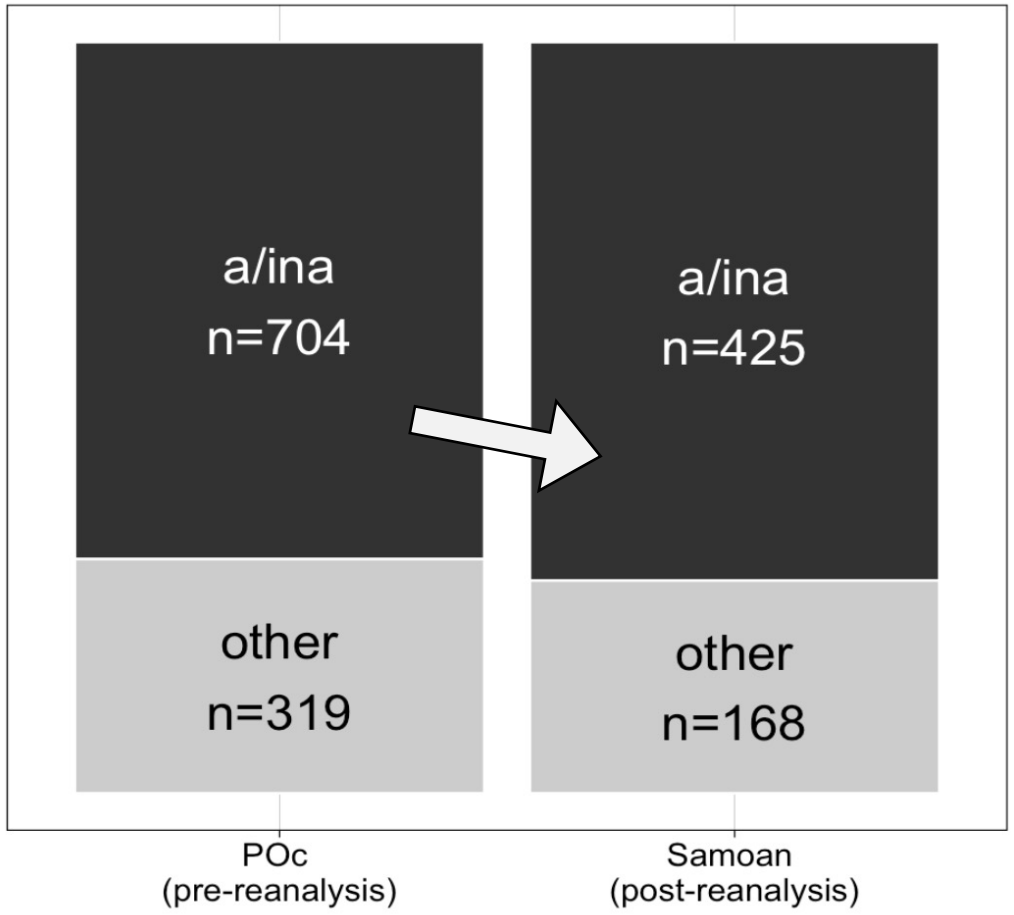
CONSTRAINT	EXAMPLE
OCP-LABIAL	*mapa, *mafu
OCP-CORONAL-SONORANT	*nanu, *lanu
OCP-DORSAL	*ŋako, *kake



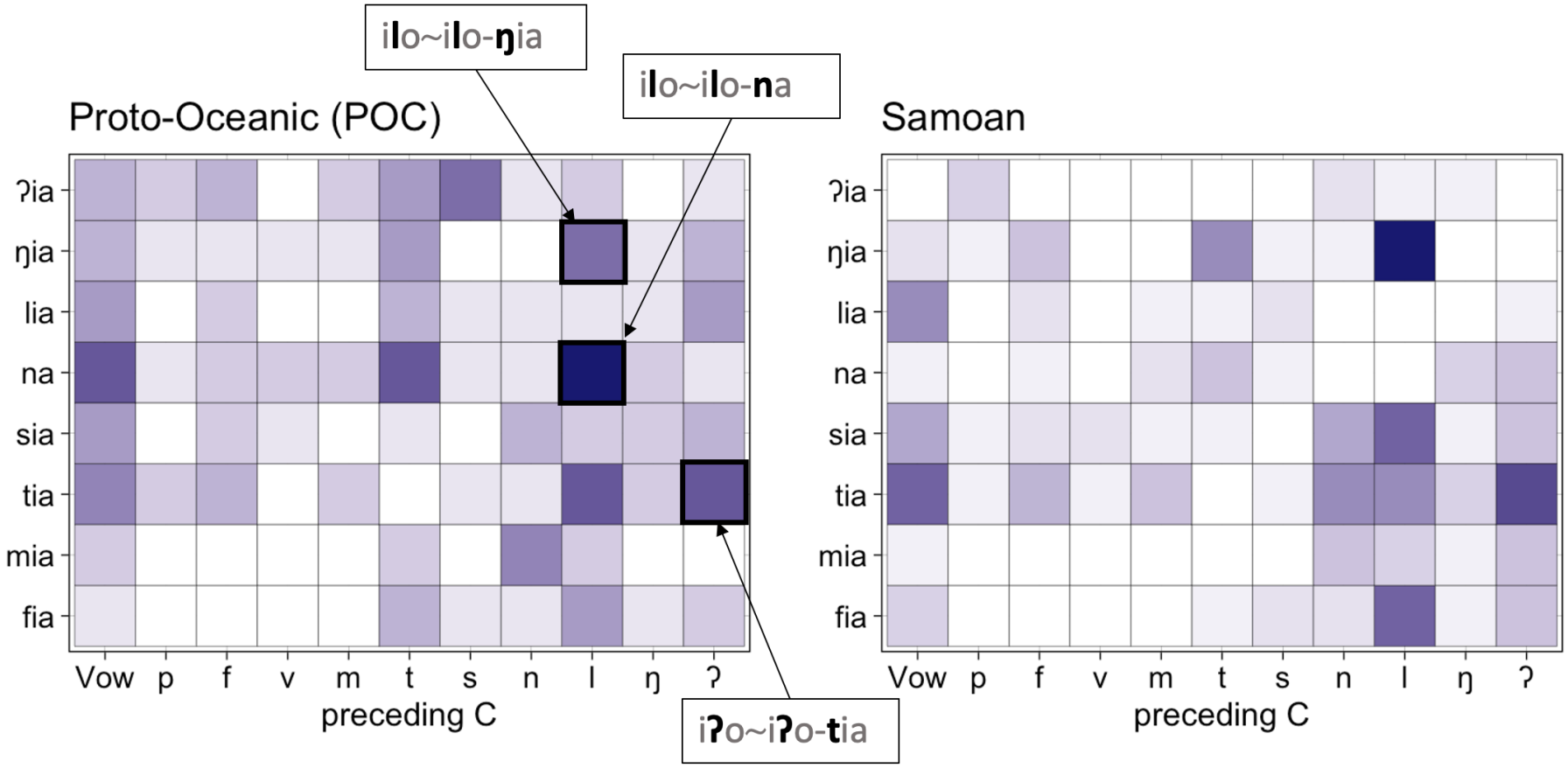
Reanalysis in Samoan

- **Data:** comparison of POC and Samoan
  - POC: 1023 protoforms, Austronesian Comparative Dictionary (Blust, Trussel, & Smith 2020)
  - Samoan: 583 stem-suffix pairs (Milner 1966)

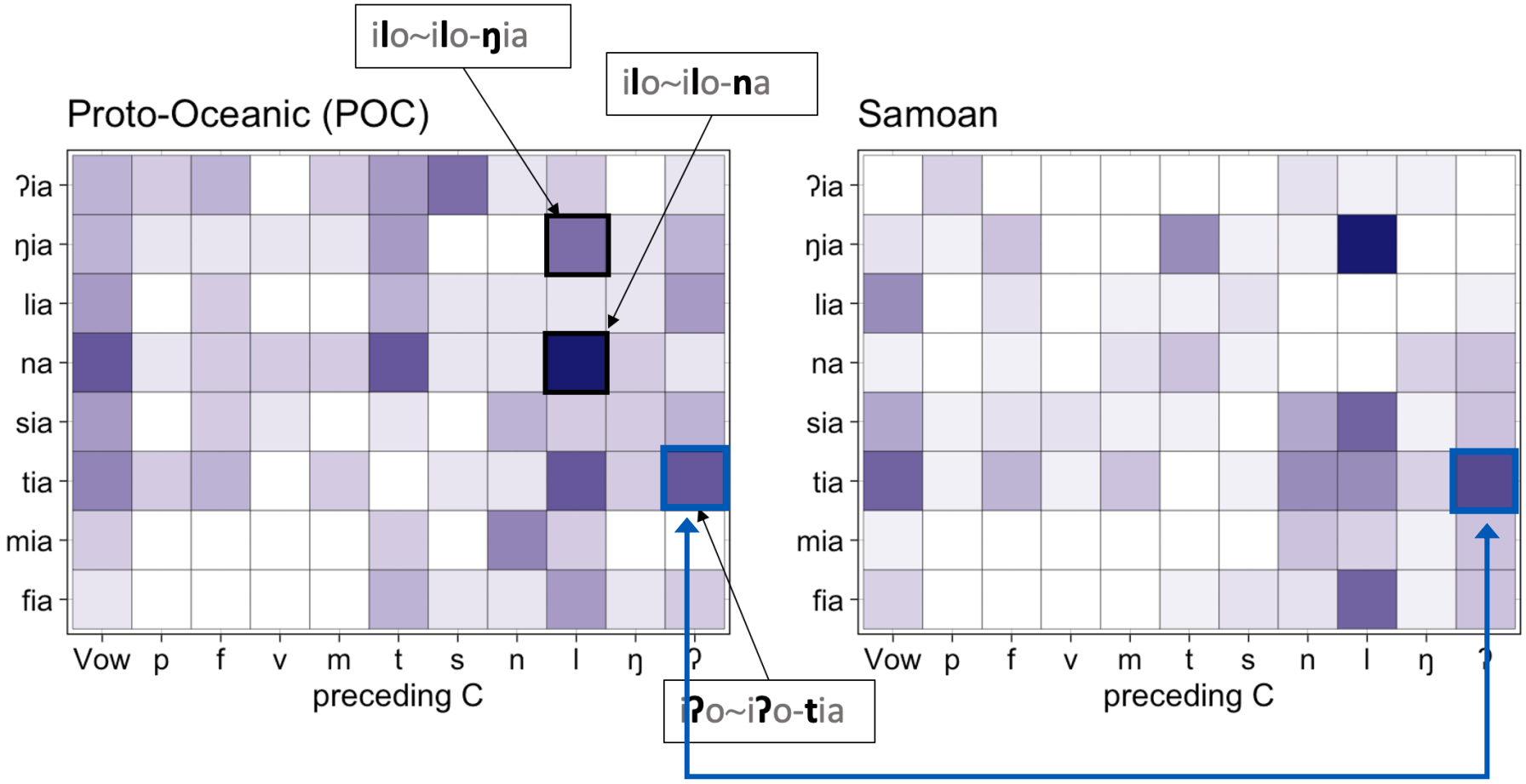
Reanalysis is generally towards the **most frequent allomorphs**



Cia allomorphs by identity of preceding C

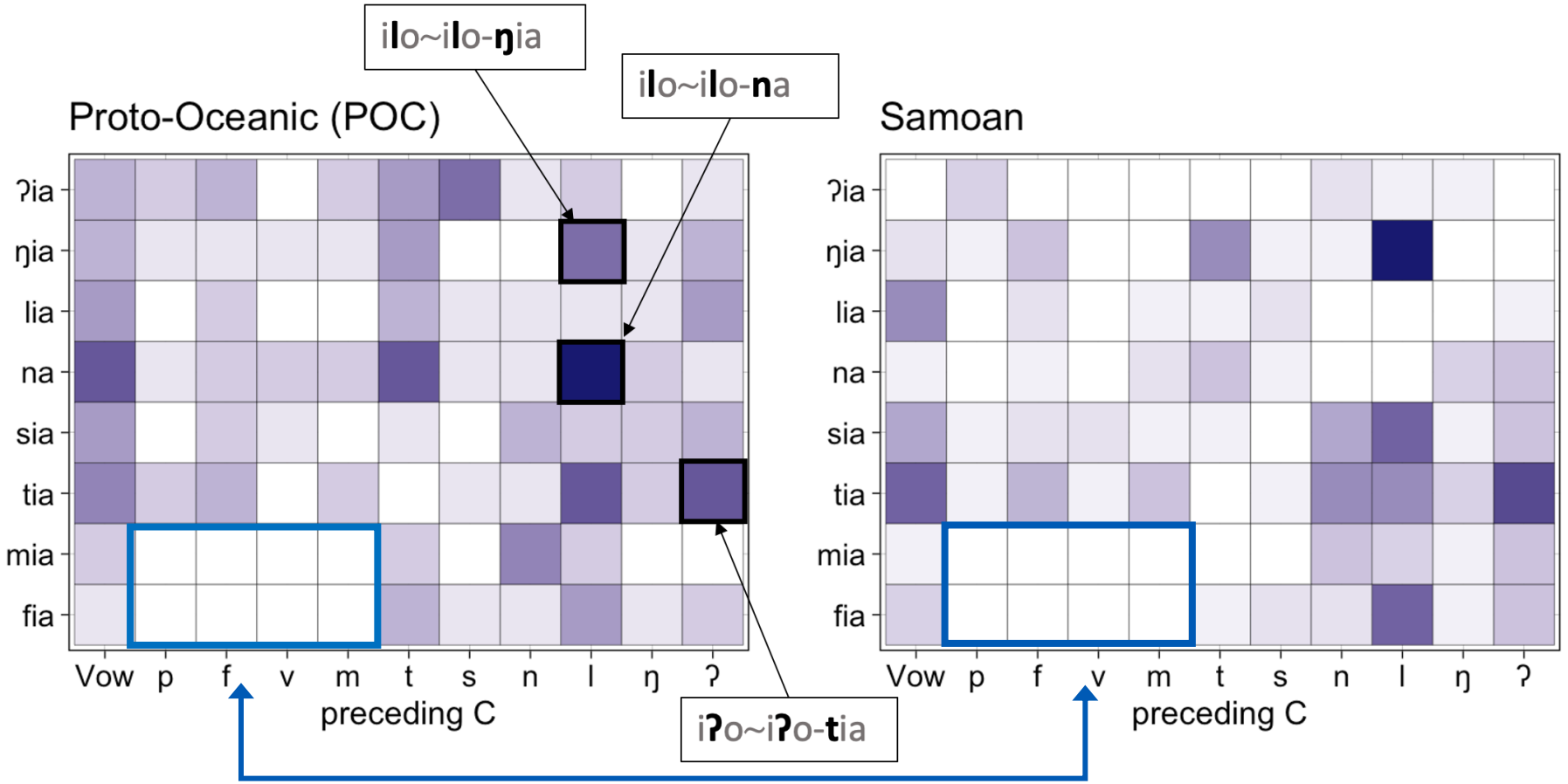


Cia allomorphs by identity of preceding C



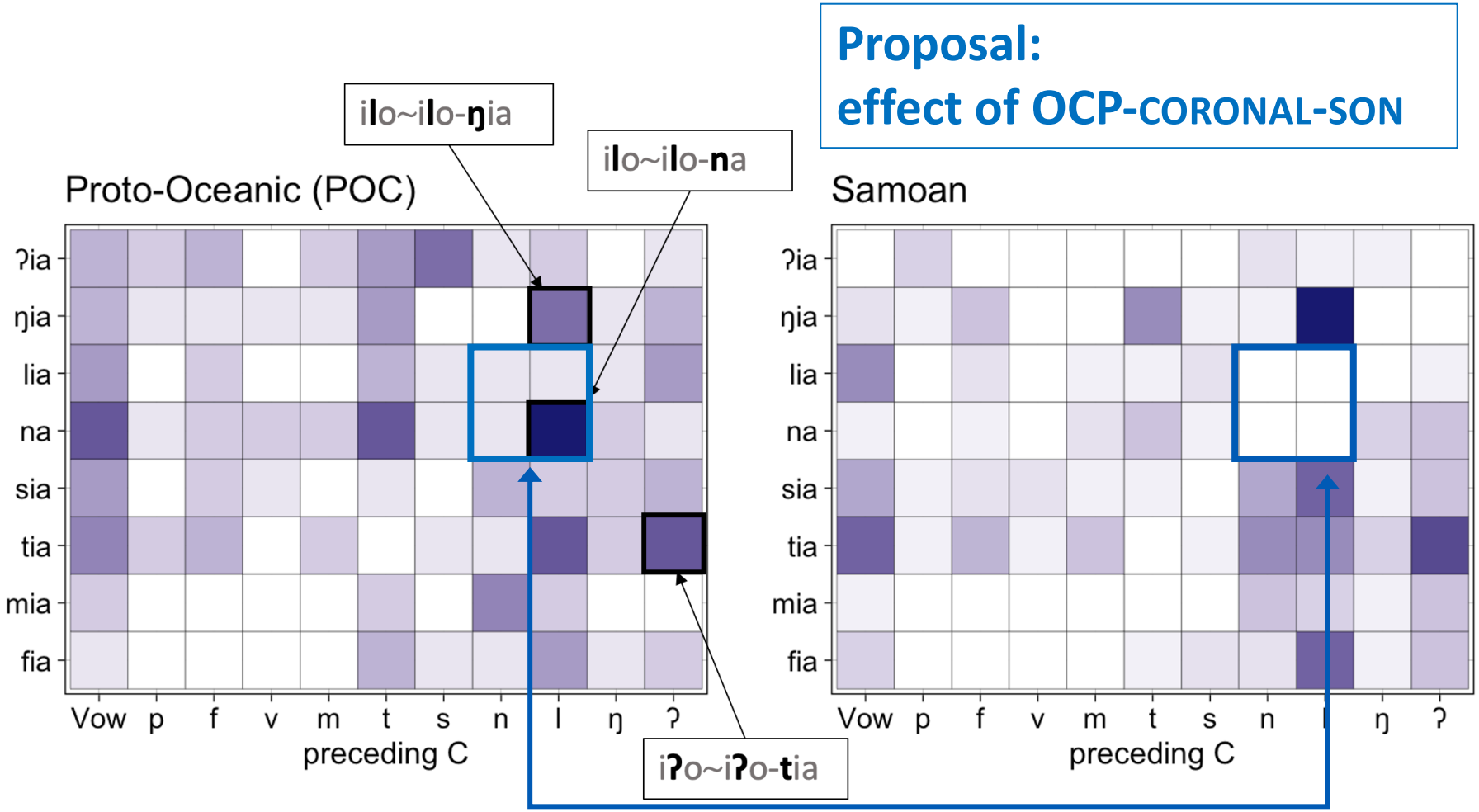
[iʔo-tia] is frequent in both POC and Samoan.

Cia allomorphs by identity of preceding C



[i{p,f,v,m}o-mia] is not observed in both POC and Samoan.

Cia allomorphs by identity of preceding C

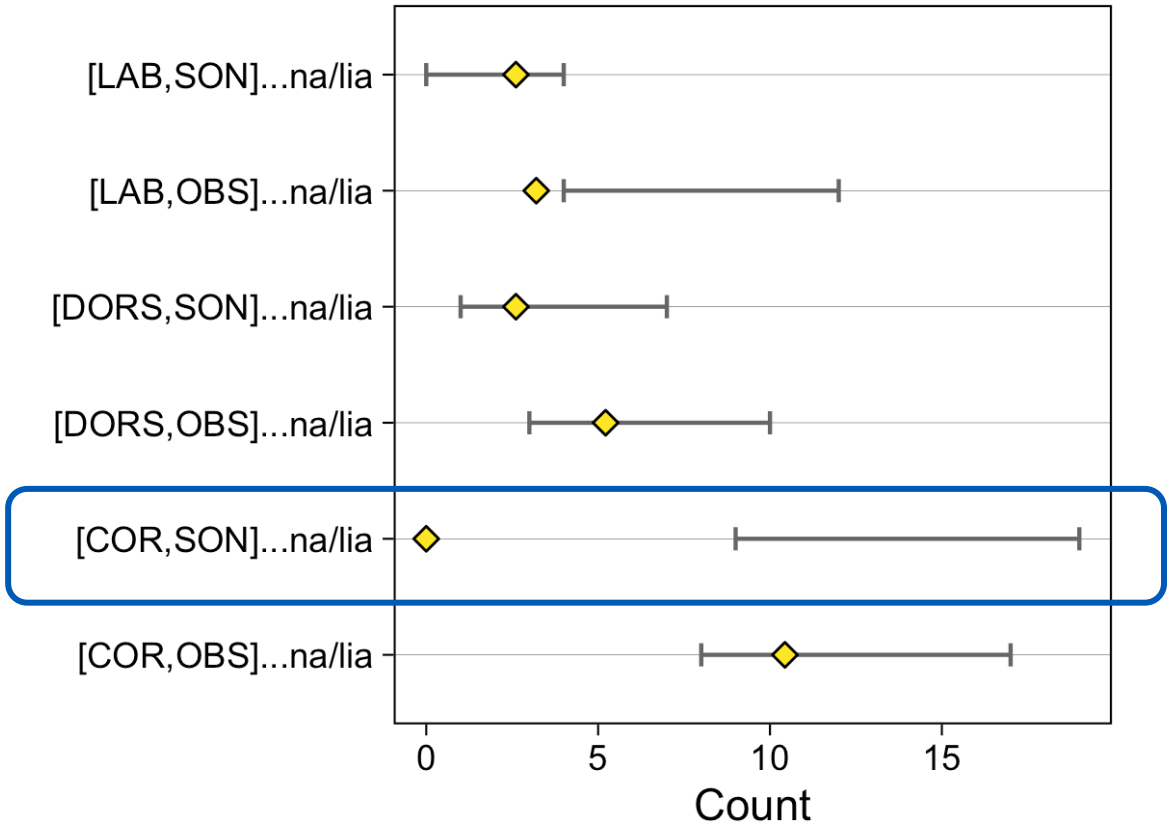


[ilo-na] is underrepresented in Samoan, given the POC distribution.

Monte Carlo simulations

- **Method**: randomly resample POC corpus 10,000 times to arrive at chance-level distribution of each suffix by the preceding consonant.
- **Intervals**: chance-level distribution, given the historical POC data.
- ◆ = observed count in Samoan

Suffixed forms that violate OCP-COR-SON ([ilo-**n**a, ino-**n**a, ino-**l**ia, ilo-**l**ia]) are underrepresented.



Modeling reanalysis in Samoan

- **Goal:** Explicit comparison of frequency-matching vs. markedness-biased models
- **Elements of the model:**
  - →**MaxEnt HG:** variant of OT that assigns candidates probabilities (Goldwater and Johnson 2003)
  - →**Bias:** implemented as a Gaussian prior (Wilson 2006; White 2013)
    - OCP-place constraint biased to have higher weight than competing constraints.
  - →**Iterative**: Predictions of one iteration are input to next iteration (de Boer 2000; Kirby 2001; Brighton 2002; Kirby, Griffiths, & Smith 2014, etc.)
  - →**Stem phonotactic grammar** used to derive markedness effects (Hayes and Wilson 2008)

Model reanalysis: inputs

- 500 stems, relative frequencies reflect that of POC protoforms.
- Pooled by…
  - Identity of preceding consonant
  - ignoring effects of final vowel, stem shape, etc.

Note: assumption that thematic consonant underlyingly belongs to the suffix (Hale 1968, 1973)

Example input (+ candidates)

Input	
/inu/+ERGATIVE	inu(in)a
	inu**f**ia
	inu**m**ia
	inu**t**ia
	inu**s**ia
	… etc.

Model reanalysis: constraint set

- **Morpheme exponence constraints** (Russell, 1995; Kager, 1996)
  - demand a particular exponent for a particular morphological category
  - e.g. 'ERG=/tia/'
- **Markedness constraint…**
  - enforcing OCP-place effects

	ERG=/(in)a/	ERG=/na/	ERG=/tia/	OCP-COR-SON		
	3	0.5	0.5	0	$\mathcal{H}$	P
/pili-ERG/						
a. pili-a		1	1		1.00	0.86
b. pili-na	1		1	1	3.50	0.07
c. pili-tia	1	1			3.50	0.07

Model reanalysis: constraint set

- **Morpheme exponence constraints** (Russell, 1995; Kager, 1996)
  - demand a particular exponent for a particular morphological category
  - e.g. 'ERG=/tia/'

- **Markedness constraint…**
  - enforcing OCP-place effects
  - **Active markedness:** can OCP-place effects be derived directly from the phonotactics?

	ERG=/(in)a/	ERG=/na/	ERG=/tia/	OCP-COR-SON	$\mathcal{H}$	P
	3	0.5	0.5	1		
/pili-ERG/						
a. pili-a		1	1		1.00	0.90
b. pili-na	1		1	1	4.50	0.03
c. pili-tia	1	1			3.50	0.07

Incorporating phonotactic markedness

1. Root phonotactics

[inu]
[tapa]
[manu]
[tai]
…

Phonotactic grammar

2. Morphophonology

Candidates:
[ilotia]
[iloŋia]
[llona]
[ilolia]
…

Phonotactic grammar

Violations (H)	
[ilotia]	0.23
[iloŋia]	0
[llona]	1.13
[ilolia]	1.13
…	

	BePhon	**C1**	**C2…**
[ilotia]	0.23	1	0
[iloŋia]	0	0	1
[llona]	1.13	0	0
[ilolia]	1.13	0	0
…			

Phonotactic grammar

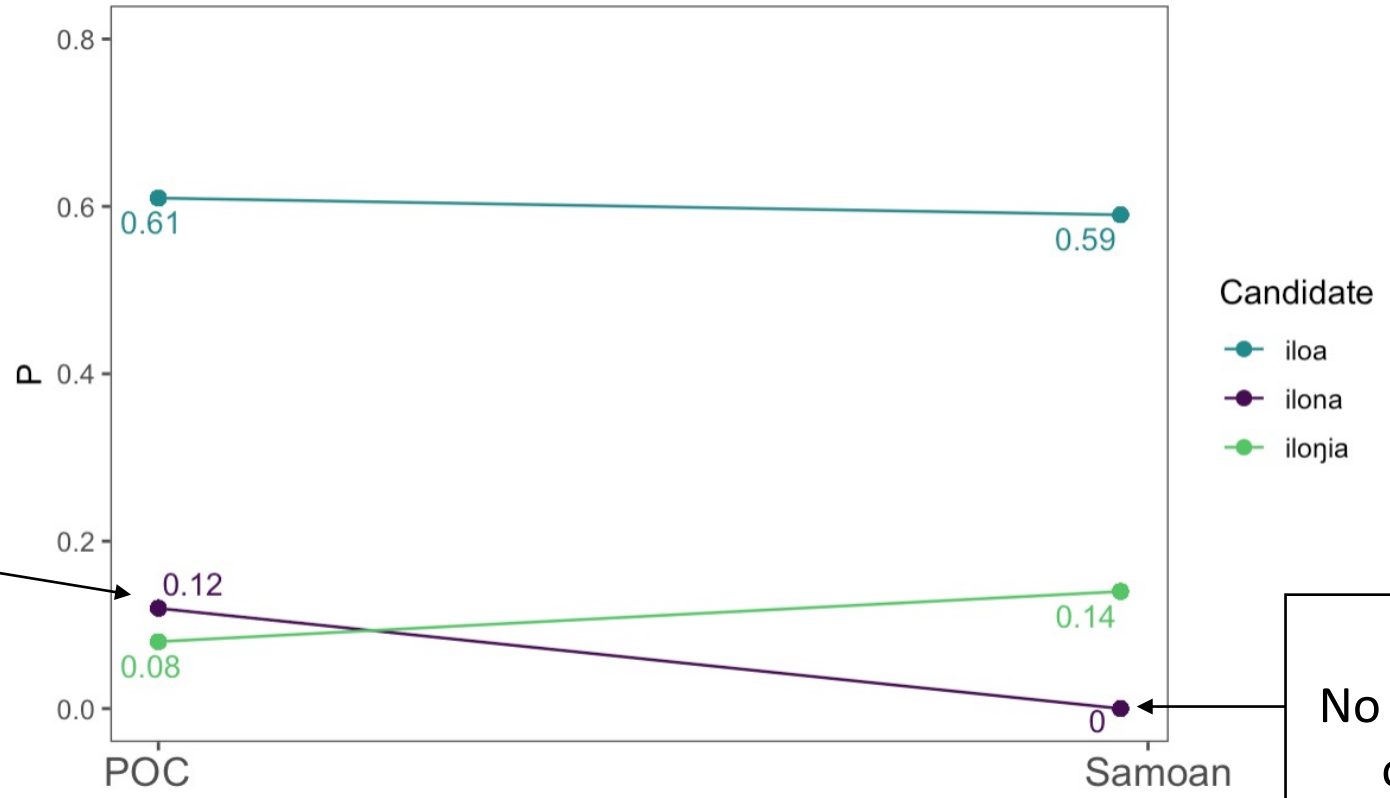
- Two phonotactic grammars:
  - **OCP-PLACE:** constraint set is all possible combinations of OCP-PLACE constraints (OCP-LABIAL, OCP-CORONAL, OCP-DORSAL) with the subsidiary features [sonorant], [voice], and [continuant].
  - **Bigram**: constraint set is all possible consonant bigram combinations
    - *p…p, *p..f, *p…t, etc.

Model results

- **Baseline.** no markedness bias
- **OCP.** markedness constraints derived from OCP grammar, biased to have high weight
- **BIGRAM.** markedness constraints derived from bigram grammar, biased to have high weight

	L	ΔL
Baseline	-2448.81	--
Bigram	-2438.39	10.42
OCP	**-2385.00**	**63.89**

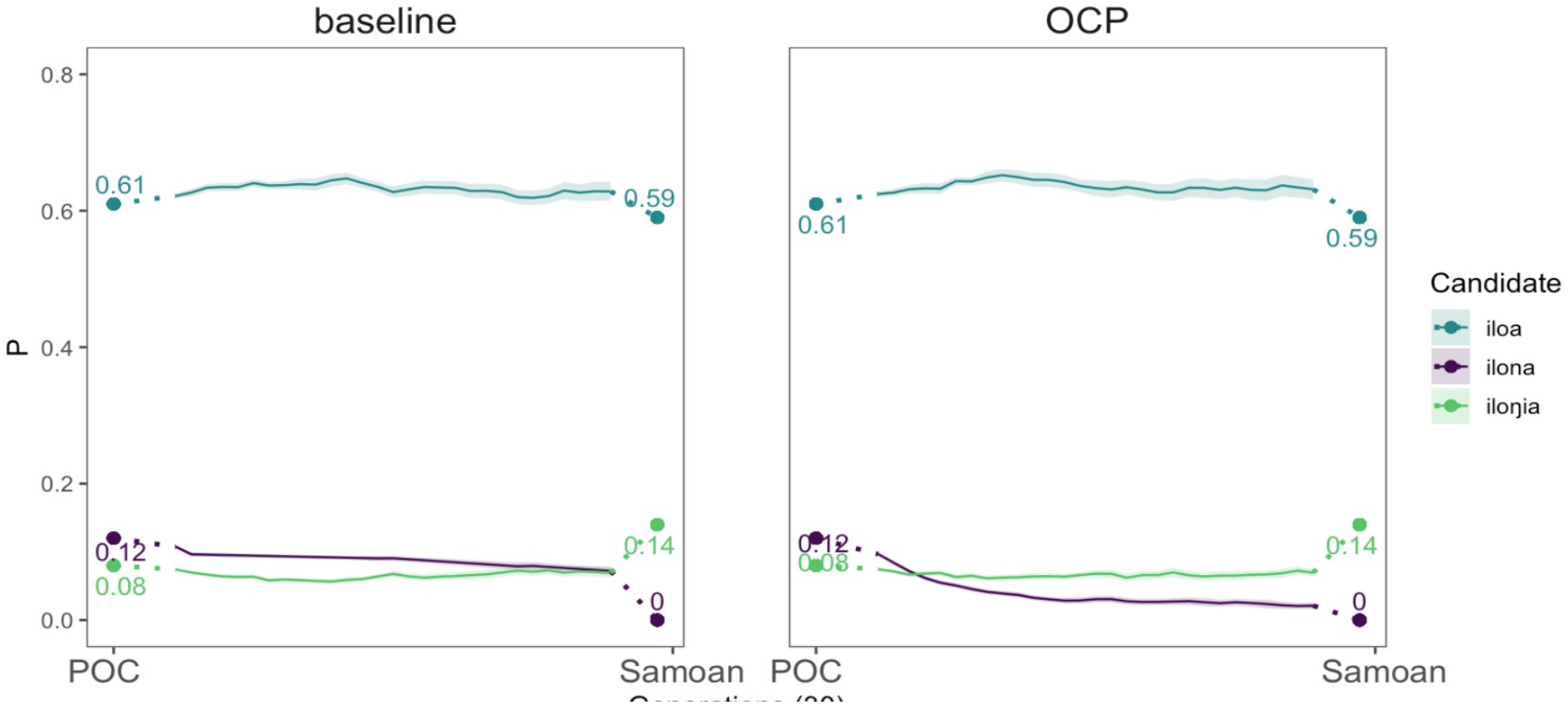
Schematic pattern for /ilo/ type stems



POC:
Relatively high rate of [ilo-na]

Samoan:
No occurrences of [ilo-na]

Model results



Conclusion

- In Samoan, reanalysis shows effects of both frequency-matching and markedness avoidance.
- Not all phonotactic generalizations are picked up by learners.
- Language change can be a **“natural laboratory”** for studying these effects (Kiparsky 1965; 1968; 1978, et seq)

Thank you!

- Thanks to Bruce Hayes, Kie Zuraw, and members of the UCLA Phonology seminar for their many insightful comments.