Markedness bias in reanalysis: an iterated learning model of Samoan thematic consonant alternations

Jennifer Kuo, UCLA

HisPhonCog 2023, Hanyang University

Background

- Paradigms with conflicting data patterns can be difficult to learn, resulting in learning errors.
- **Reanalysis**: when such errors are adopted into speech community, resulting in a type of language change.

Background: Samoan example

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

ERG.	STEM	SUFFIXED	GLOSS
a	rere	rere-a	‘to take’
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 example

/a, 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	rere	rere-a	‘to take’
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 example

- 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	rere	rere-a	‘to take’
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’

/Cia/: consonant-initial allomorphs

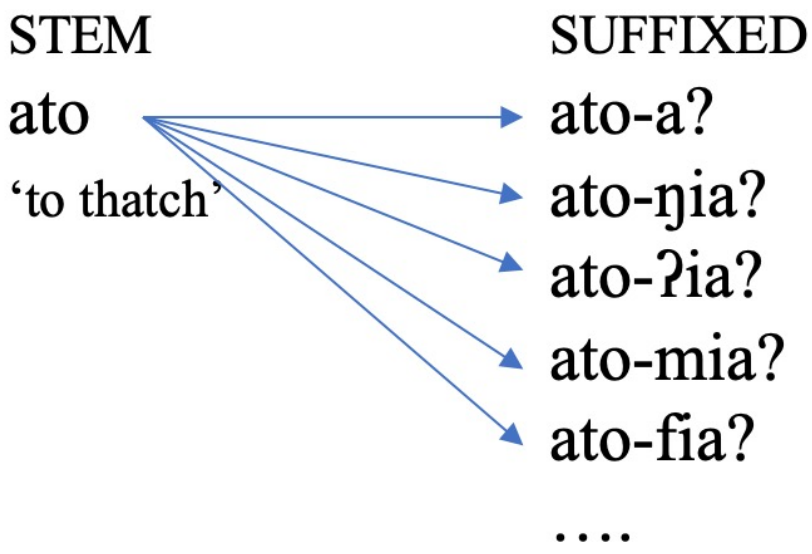
Development of Ø~C alternations

	‘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
Proto-Polynesian (PPn)						
C→Ø/__#	inu	inu**m**ia	pulu	pulu**t**ia	-	
i→Ø/__a	-		-		pulu	pulu=a
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
*qatop	ato-**fia**	ato-**a**	f→∅

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: a bias for outputs that are less marked (Jarosz 2006)

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

Active markedness in reanalysis

- 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)

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

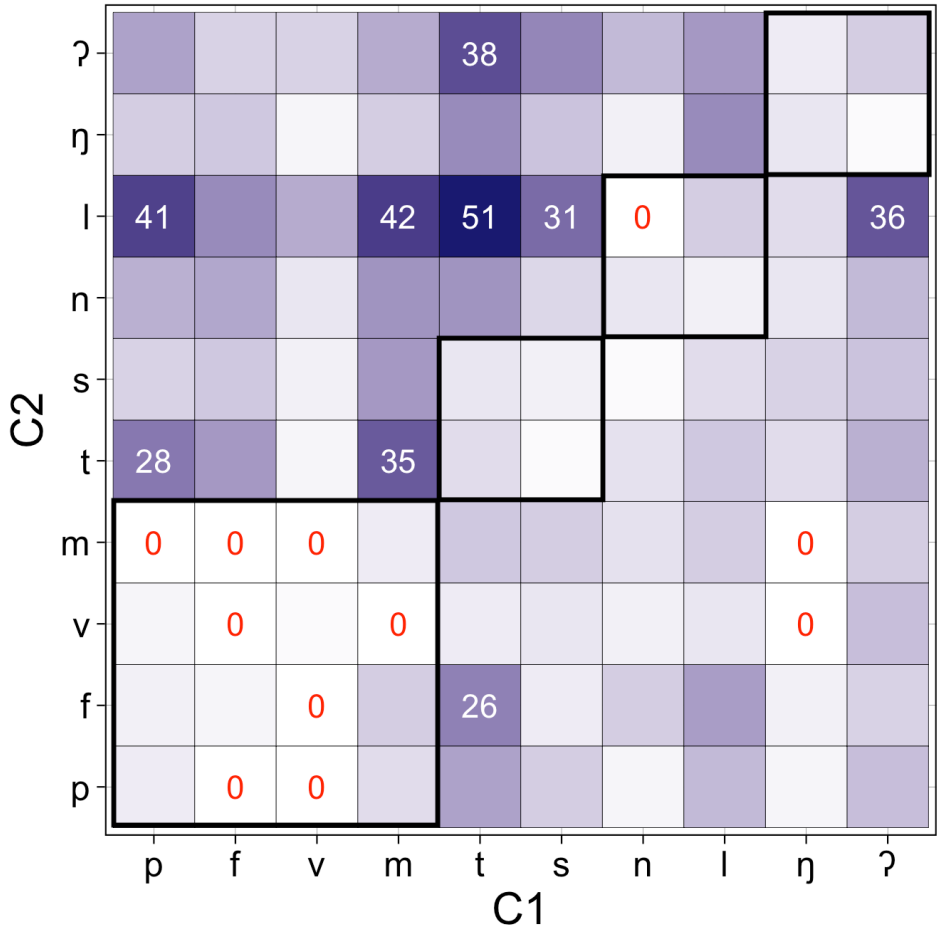
    e.g. *[**p**u**f**a] (***p…f**)

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

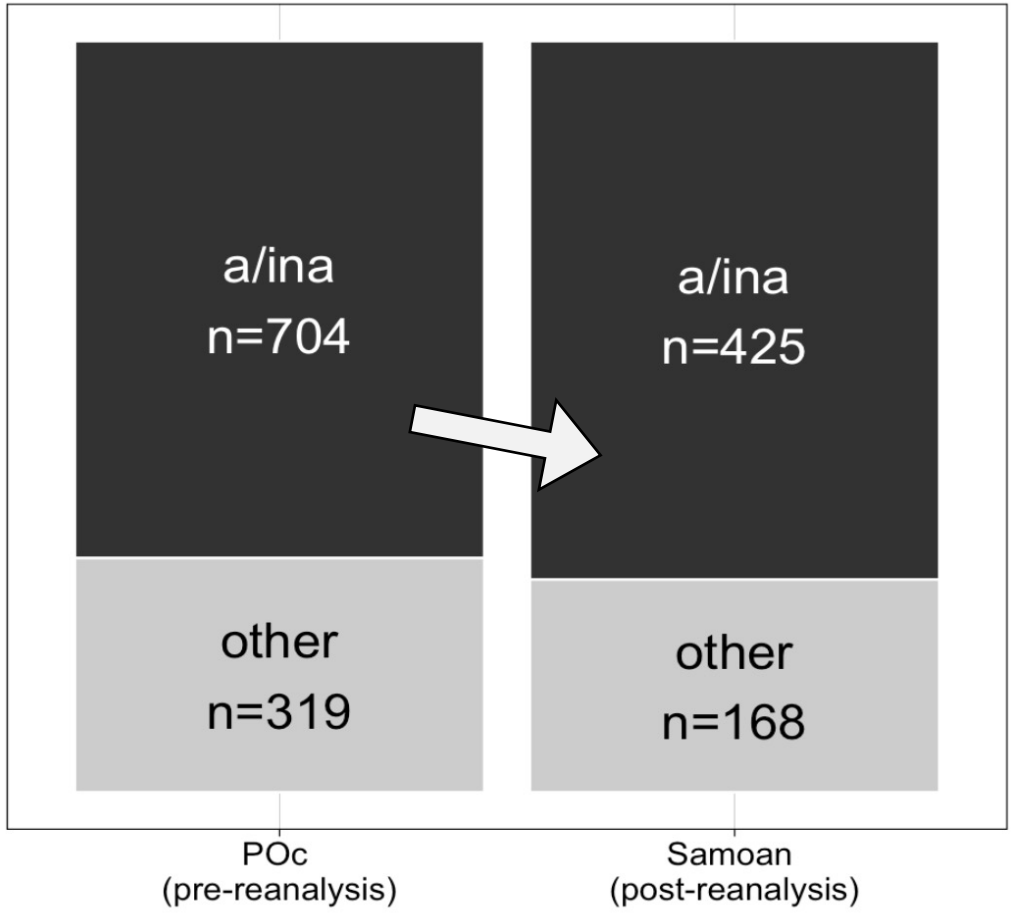
- PPn *k > Samoan [ʔ], but [ŋ] and [ʔ] still behave as a natural class.
- Same patterns found in Proto-Polynesian



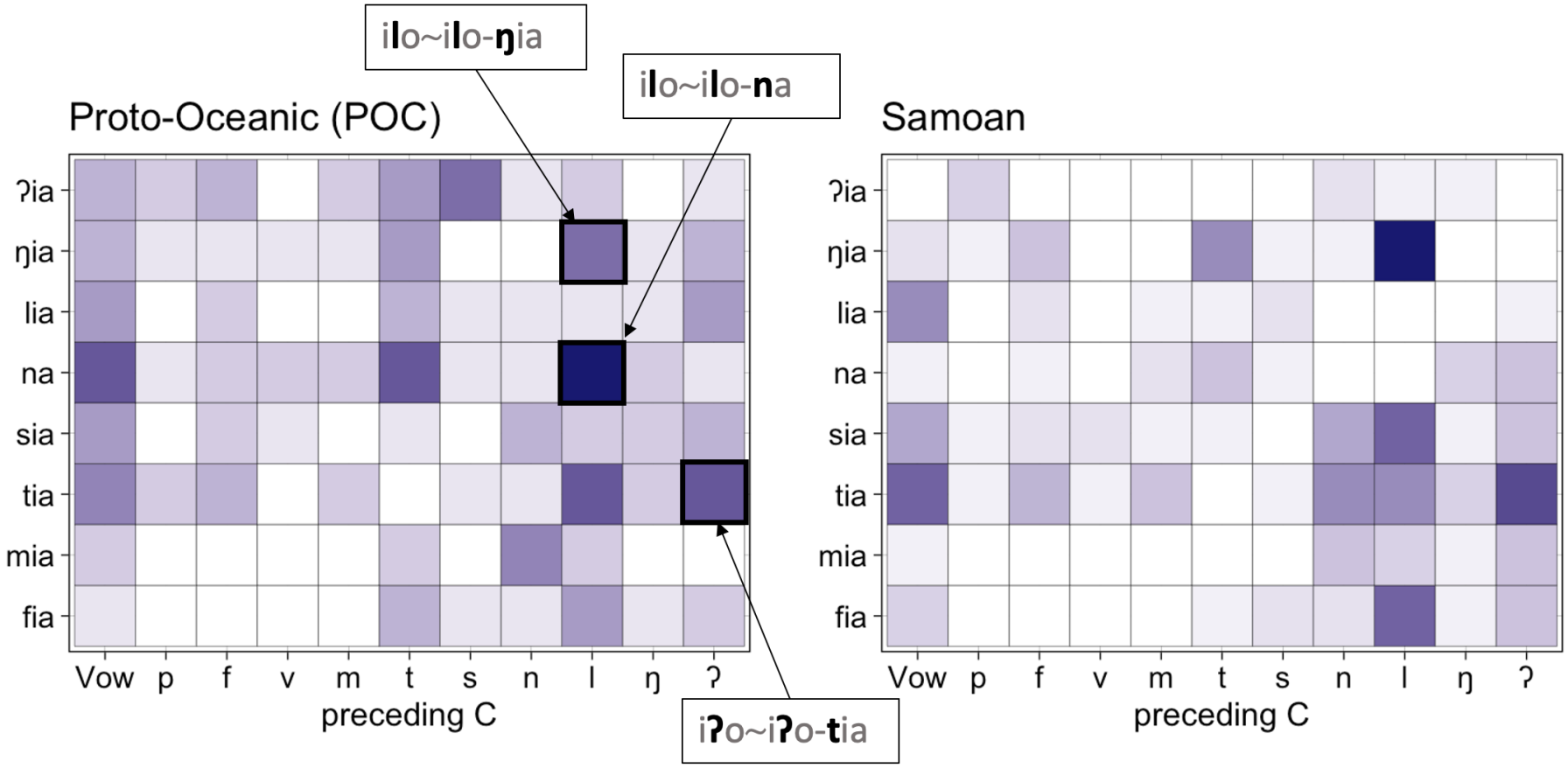
Reanalysis in Samoan

- **Data:** comparison of POC and Samoan
  - POC: 1023 protoforms, Austronesian Comparative Dictionary (Blust & Trussel 2010)
  - Samoan: 583 stem-suffix pairs (Milner 1966)

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

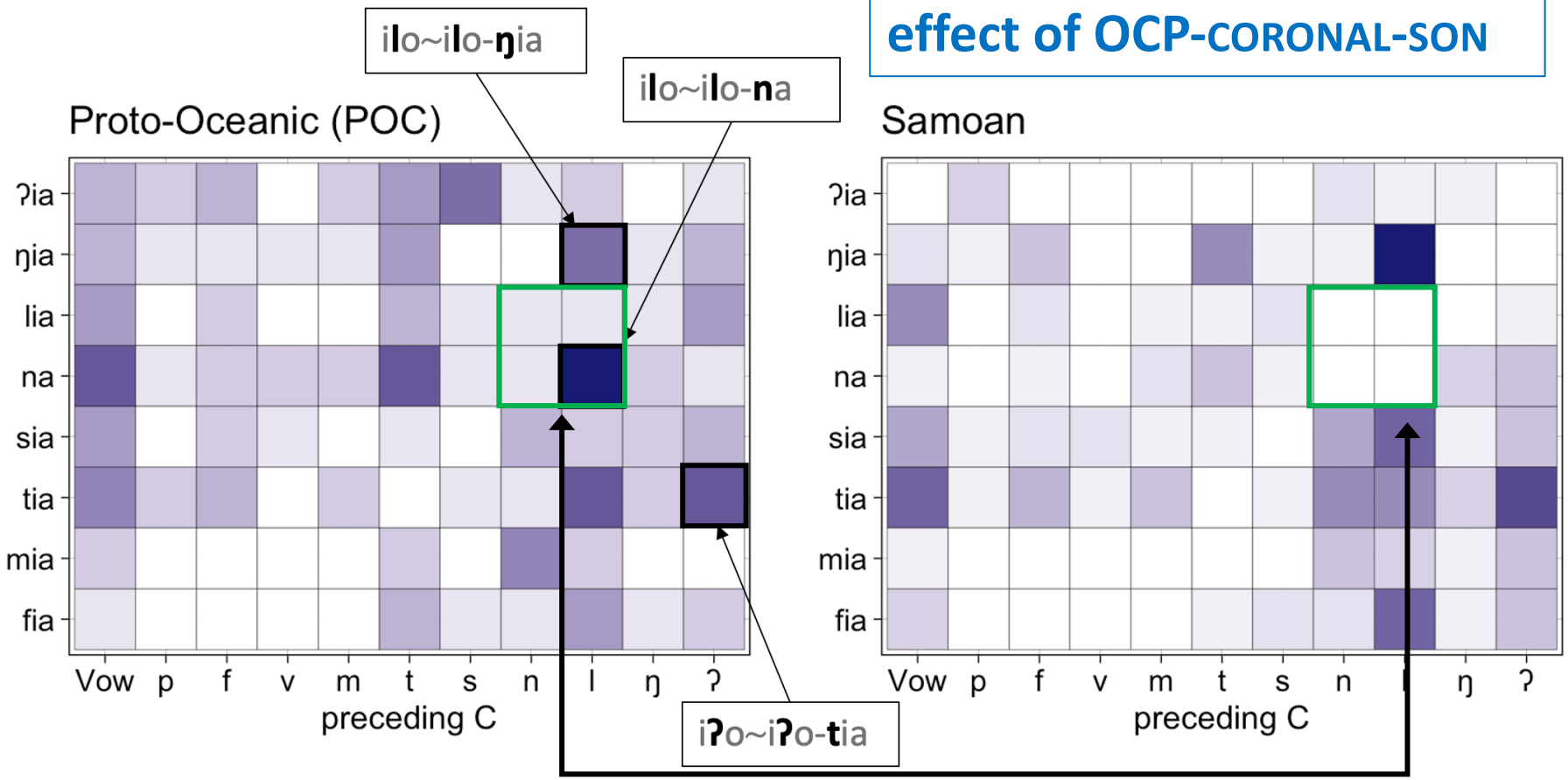


Cia allomorphs by identity of preceding C



Cia allomorphs by identity of preceding C

Proposal:
effect of OCP-CORONAL-SON



[ilo-na] is underrepresented in Samoan, given the POC distribution.
[significant in a Monte Carlo simulation]

Modeling reanalysis in Samoan

- **Goal:** Explicit comparison of frequency-matching vs. markedness-biased models
- **Elements of the model:**
  - →**MaxEnt HG** to capture gradient alternations (Goldwater and Johnson 2003)
  - →**Stem phonotactic grammar** used to derive markedness effects (Hayes and Wilson 2008)
  - →**Bias** implemented as a Gaussian prior (Wilson 2006; White 2013)
  - →**Iterative**: Predictions of one iteration are input to next iteration (de Boer 2000; Kirby 2001; Brighton 2002; Kirby, Griffiths, & Smith 2014, etc.)

Modeling reanalysis in Samoan

- **Goal:** Explicit comparison of frequency-matching vs. markedness-biased models
- **Elements of the model:**
  - →**MaxEnt HG** to capture gradient alternations (Goldwater and Johnson 2003)
  - →**Stem phonotactic grammar** used to derive markedness effects (Hayes and Wilson 2008)
  - →**Bias** implemented as a Gaussian prior (Wilson 2006; White 2013)
  - →**Iterative**: Predictions of one iteration are input to next iteration (de Boer 2000; Kirby 2001; Brighton 2002; Kirby, Griffiths, & Smith 2014, etc.)

Phonotactic grammar

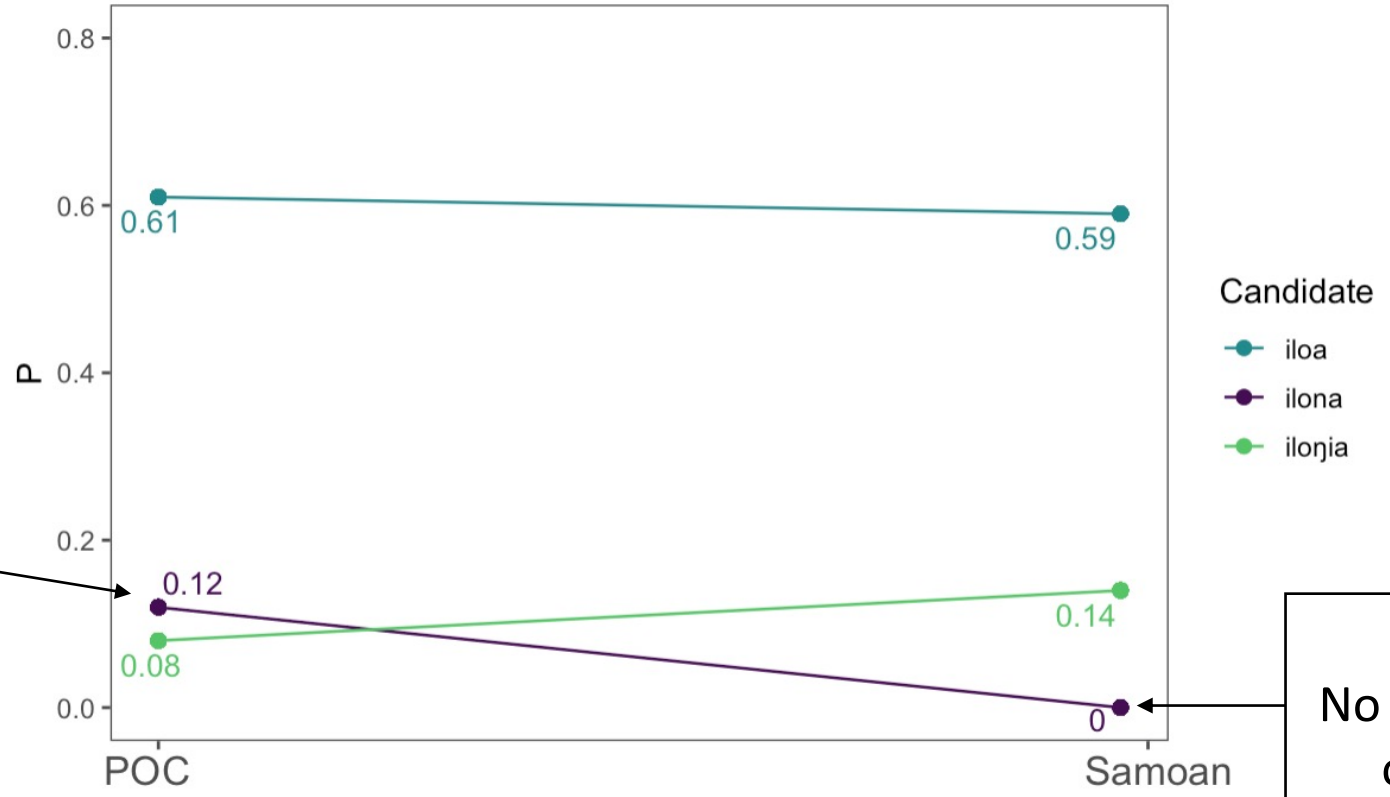
- Two phonotactic grammars:
  - **OCP-PLACE:** constraint set is all possible combinations of OCP-PLACE constraints (OCP-LABIAL, OCP-CORONAL, OCP-DORSAL) with the subsidiary features [sonorant], [voice], and [continuant].
  - **Bigram**: constraint set is all possible consonant bigram combinations
 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

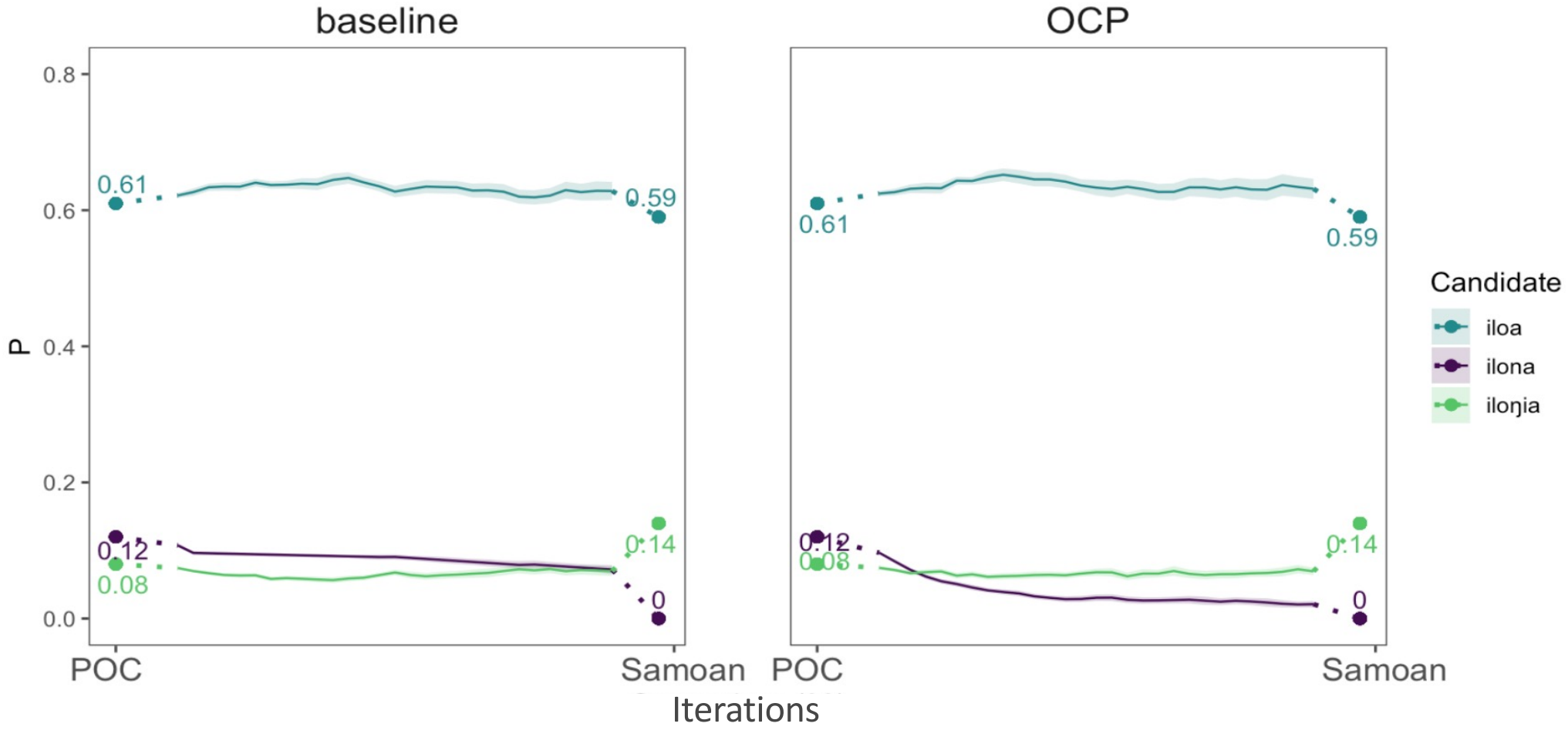
Data pattern for /ilo/ type stems



POC:
Relatively high rate of [ilo-na]

Samoan:
No occurrences of [ilo-na]

Model results



Conclusion

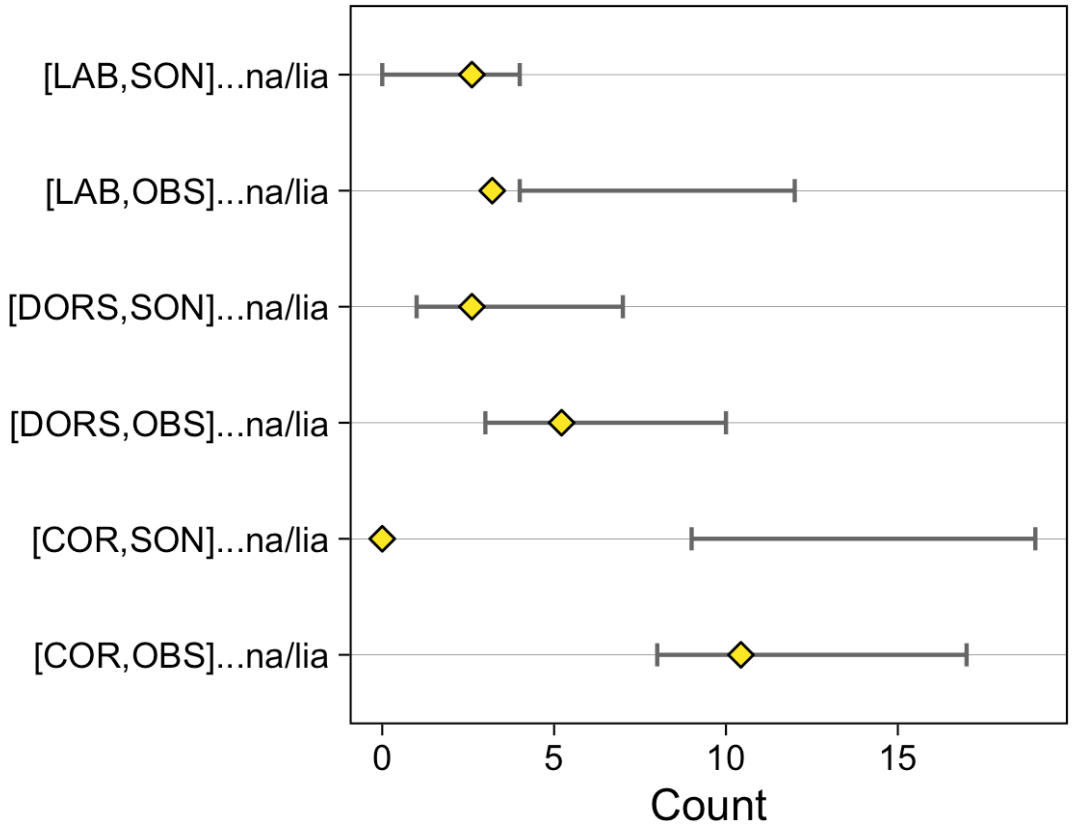
- In Samoan, reanalysis shows effects of both frequency-matching and markedness avoidance.
- Not all phonotactic generalizations are picked up by learners.
- Language change can be a **“natural laboratory”** for studying these effects (Kiparsky 1965; 1968; 1978, et seq)

Thank you!

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

Appendix: Monte Carlo simulations

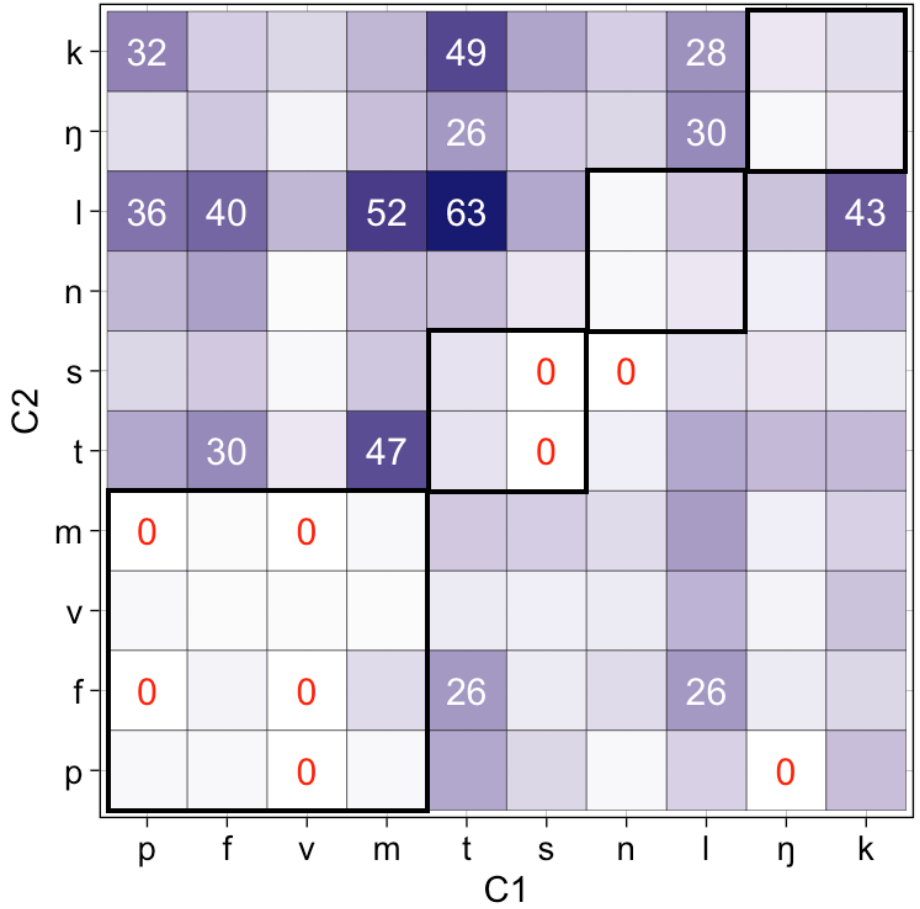
- **Method**: randomly resample POC corpus 10,000 times to arrive at chance-level distribution of each suffix by the preceding consonant.
- **Intervals**: chance-level distribution, given the historical POC data.
- ◆ = observed count in Samoan



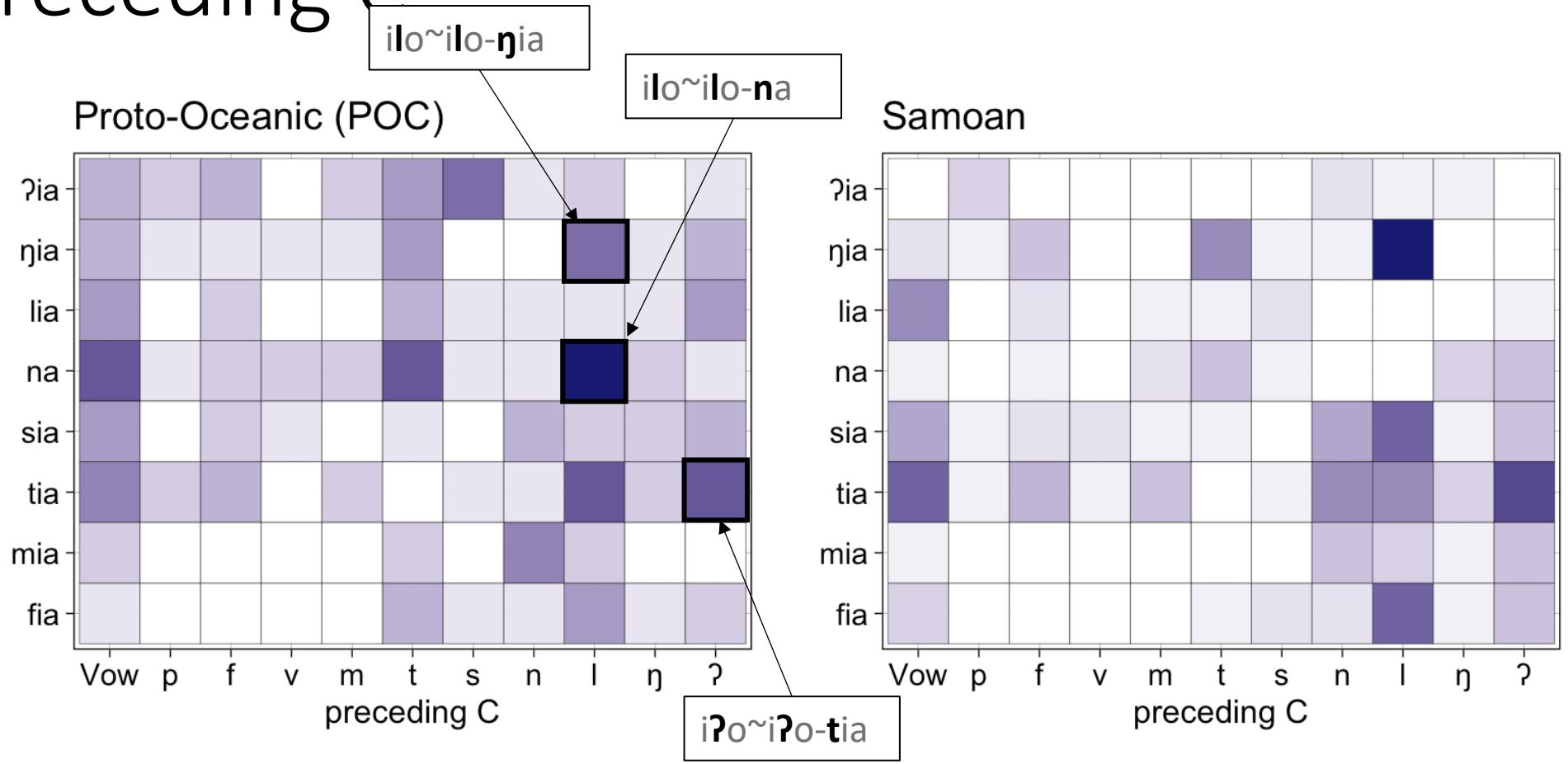
Appendix: OCP-place in Proto-Polynesian

- **Data:** Protoforms from the Polynesian Lexicon Project (POLLEX; Greenhill & Clark 2011).
- Active OCP-place constraints based on a Maxent phonotactic model (Wilson & Obdeyn 2009)

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



Appendix: Cia allomorphs by identity of preceding C



Appenfix: Development of CIA allomorphs

Development of Samoan ergative suffix

Suff.	/ia/	/ina/	/a/	/Cia/	/ina/	/na/	
POc	*nofo-i-a	*tabu-i-na	*tari-i-a	*puat-i-a	*pulan-i-a	*talun-i-a	CHANGE
	-	-	tali-a	-	-	-	*i-deletion[A]
	-	-		-	pula-ina		Metathesis[B]
	-	-		-		talun-a	*nia -> na
Sam.	nofo-ia 'sit'	tapu-ina 'be banned'	tali-a 'wait'	fua-tia 'bear fruit'	pula-ina 'be bright'	talu-na 'forest'	

A. *i-deletion: *i is deleted after *i,*e[1]

B. Metathesis of *nia to *ina after *a