

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

Jennifer Kuo<sup>1</sup>

<sup>1</sup>*University of California, Los Angeles*  
Jenniferkuo2018@ucla.edu

Paradigms with conflicting data patterns can be difficult to learn, resulting in child errors (e.g. *go/goed* instead of *go/went* in English). Such errors can in turn be adopted into speech communities, resulting in a type of change over time I refer to as *reanalysis*. Existing models of morphophonology, such as Albright's [1, 2] Minimal Generalization Learner, predict reanalysis to be frequency-matching, occurring in a way that matches probabilistic distributions within a paradigm. I propose that in fact, reanalysis responds to two factors: both frequency matching and the reduction of markedness.

In this study, I use iterated learning models to investigate this issue in a set of Samoan alternations. Two models are compared: one that is frequency-matching, and one which has a markedness learning bias. I find that the latter model performs better. I further propose that the markedness effects allowed to influence reanalysis are restricted to those already active in the language (e.g. in root phonotactics), and show that Samoan is consistent with this proposal.

In some Samoan suffixes, a consonant of unpredictable quality surfaces, as exemplified in (1) for the ergative suffix [3]. This pattern arose due to a historic process of final consonant loss. As a result, all consonants were deleted at the end of unsuffixed stems, but maintained in suffixed forms (e.g. *\*inum/\*inum-ia* 'to drink' → *inu/inu-mia*).

In general, the allomorph that surfaces should be traceable back to the historic stem-final consonant in Proto-Oceanic (POc) [4]. For example, *[inu]/[inu-mia]* 'to drink' comes from POc *\*inum*, and *[pulu]/[pulu-tia]* 'to plug up' comes from *\*bulut*. However, in modern Samoan, the observed alternant often does *not* match the historical POc one; for example, *[ŋuŋu]* (<POc *\*ŋuŋul*) 'arthritis' should have the suffixed form *[ŋuŋu-lia]*, but instead *[ŋuŋu-a]* is observed. These mismatches suggest that language learners have carried out reanalysis in multiple instances. To investigate the direction of reanalyses, I collected 358 POc forms with known Samoan reflexes, taken from the Austronesian Comparative Dictionary [5]. POc stems were compared against 584 Samoan stem/ergative pairs collected from Milner's Samoan Dictionary [6].

I find that reanalysis is sensitive to transvocalic consonant OCP effects. In particular, suffixed forms are more likely to be reanalyzed if they violate an OCP constraint against coronal sonorants (*\*[+COR,+son]...[+COR,+son]*), which assigns violations to stems such as *[lanu]* 'color'. In fact, in modern Samoan, there are almost no suffixed forms of the type *[puli-na]* or *[puni-lia]* (*n=2/584*). Using a Monte Carlo simulation [7], visualized in Figure 1, forms that violate coronal sonorant OCP are shown to be underrepresented in modern Samoan, given the distribution of final consonants in POc.

Moreover, I find that OCP[+COR,+son] is also active in Samoan root phonotactics. Specifically, a probabilistic constraint-based phonotactic model [UCLA Phonotactic Learner; 8] was trained on 1600 Samoan roots from Milner's Samoan Dictionary [6]. The resulting model assigns significant weight to the constraint OCP[+COR,+son]. This finding is compatible with the proposal that markedness effects in reanalysis are restricted to those already active in the language.

These results are confirmed using a model of reanalysis implemented in Maximum Entropy Harmonic Grammar [Maxent; 9]. The model is iterated to simulate the cumulative effects of reanalyses over time. In other words, at each "generation", the learner induces a grammar based on input data and then uses this grammar to generate data that is passed down to the next generation. Two models are compared: 1) a baseline model that is purely frequency-matching, and 2) a markedness-biased model in which the constraint OCP[+COR,+son] is biased to have high weight using the method laid out by Wilson [10]. I find that the markedness-biased model performs significantly better than the purely distributional baseline model. In sum, the Samoan data supports

the view that reanalysis (and more generally morphophonology) is guided both by the statistical patterns that learners encounter and by principles of markedness.

(1) *Ergative suffix allomorphy in Samoan*

ERG	STEM	SUFFIXED	GLOSS
a	rere	rere-a	‘to take’
ina	iloa	iloa-ina	‘to see, perceive’
tia	pulu	pulu-tia	‘to plug up’
sia	laka	laka-sia	‘to step over’
ŋia	tutu	tutu-ŋia	‘to light a fire’
fia	utu	utu-fia	‘to draw water’
mia	inu	inu-mia	‘to drink’
lia	tatau	tatau-lia	‘to hang up’
na <sup>1</sup>	ʔai	ʔai-na	‘to eat’
ʔia	momo	momo-ʔia	‘to break in pieces’

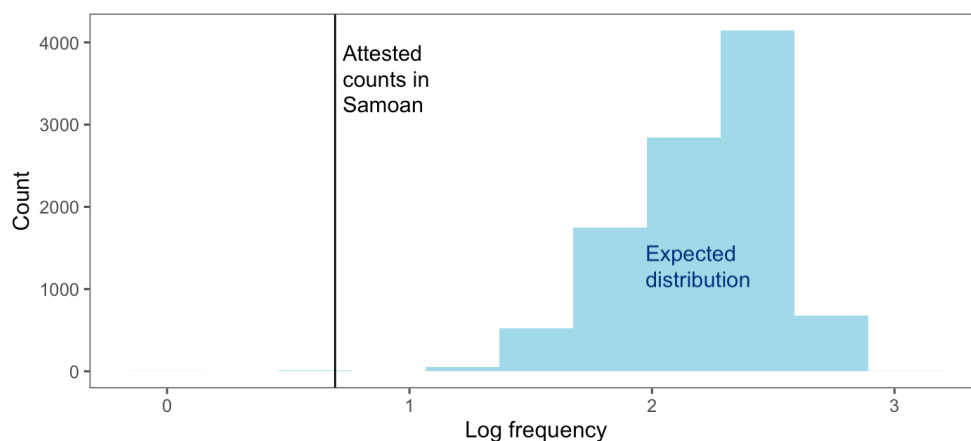


Figure 1: Attested [puli-na]/[puni-lia] words vs. expected distribution from POc

References

- [1] Albright, A. (2002). *The identification of bases in morphological paradigms* [PhD thesis, UCLA].
- [2] Albright, A. & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90(2), 119–161.
- [3] Mosel, U. & Hovdhaugen, E. (1992). *Samoan reference grammar*. Scandinavian Univ. Press.
- [4] Pawley, A. (2001). Proto Polynesian \*-CIA”. In J. Bradshaw & K. L. Rehg (eds.), *Issues in Austronesian Morphology: A festschrift for Byron W. Bender*. Pacific Linguistics.
- [5] Blust, R. and Trussel, S. (2020) *Austronesian Comparative Dictionary*. <https://www.trussel2.com/ACD>
- [6] Milner, G. B. (1966). *Samoan Dictionary; Samoan-English, English-Samoan*. ERIC, 1966
- [7] Mooney, C. Z. (1997). *Monte Carlo Simulation*. Thousand Oaks, CA: Sage Publications.
- [8] Hayes, B. & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3), 379–440.
- [9] Goldwater, S. & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In J. Spenser, A. Eriksson, & Ö. Dahl (eds.), *Proceedings of the Stockholm workshop on variation within Optimality Theory* (pp. 111-120). Stockholm University, Sweden.
- [10] Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive science*, (30)5, 945–982.

<sup>1</sup>Note that when the ergative suffix starts with /n/ the allomorph is /na/ rather than /nia/