Hiatus avoidance and the development of Māori passive allomorphy*

Jennifer Kuo

Cornell University

1. Introduction

In morphophonological learning, speakers are known to frequency-match, applying patterns in a way that matches the proportion at which they occur in a paradigm (Ernestus and Baayen 2003). However, the interaction of frequency-matching with other factors is not well-understood. To address this gap, the current paper examines how frequency-matching interacts with phonological markedness effects. Additionally, I also draw a distinction between *universal* markedness and language-specific markedness effects that are *active* in the language, in the form of stem phonotactics. This paper explores the idea that only *active* markedness effects can influence morphophonological learning.

The empirical focus of this paper is how morphophonological paradigms in Māori have been *reanalyzed* over time. Paradigms with conflicting patterns can be hard to learn, resulting in child errors (e.g. *mouse/mouses* vs. *mouse/mice* in English). Such errors can be adopted into speech communities, resulting in a type of language change I refer to as reanalysis. Examples of reanalysis in English include *help/halp→help/helped* (∼1300, OED) and *dive/dived→ dive/dove* (∼1800, OED). Patterns of reanalysis can be a window into morphophonological learning, if we assume that dispreferred patterns are prone to reanalysis.

Māori is an especially suitable test case because it has complicated patterns of allomorphy. The passive suffix has a large number of consonant-initial allomorphs whose distribution is mostly unpredictable. This makes it a pattern that is difficult to learn and therefore prone to reanalysis. I present data showing that while reanalysis in Māori is generally frequency-matching, it is also driven by the avoidance of vowel hiatus and heavy syllables.

2. Background

Māori is a Polynesian language spoken in mainland New Zealand. Māori passive allomorphy has received much theoretical attention, including: Hale 1968, 1973, Sanders 1990,

*Thank you to Bruce Hayes, Kie Zuraw, Claire Moore-Cantwell, members of the UCLA Phonology Seminar, and the audience at NELS 54 for their valuable input.

Blevins 1994, and de Lacy 2003. However, these studies focused on the synchronic representation of the passive allomorph, so patterns of reanalysis in Māori are still not well understood. Throughout this paper, basic descriptive facts about Māori phonology will be taken from Harlow 2007. Additionally, I draw on Blevins 1994 and de Lacy 2003 for generalizations about the synchronic distribution of passive allomorphs in Māori. Finally, all examples come from the Williams 7th edition dictionary (Williams 1971), which is primarily based off of speakers from Ngāpuhi, which falls under the Western dialect region.

2.1 Passive allomorphy in Māori

The Māori passive suffix has allomorphs /-a/, /-ia/, /-na/, /-ina/ and /-Cia/ (where C represents a variable consonant). The paradigms in (1) exemplify this allomorphy.[1,2]

(1)

stem	suffixed	gloss	stem	suffixed	gloss
fao	fao-**a**	'perforate'	ai	ai-**tia**	'to copulate'
paː	paː-**ia**	'stockade'	mataku	mataku-**ria**	'be feared'
aŋi	aŋi-**na**	'to blow'	rere	rere-**kia**	'carried by wind'
uta	uta-**ina**	'interior'	kuː	kuː-**ŋia**	'to coo'
inu	inu-**mia**	'to drink'	motu	motu-**hia**	'to separate'

The distribution of passive allomorphs is partly predictable. First, the relative distribution of /-na/ and /-ina/ is predictable, with /-ina/ surfacing after [a]-final stems and /-na/ elsewhere (e.g. [tipako-na] 'select' vs. [kata-ina] 'laugh at'). The vowel-initial allomorphs /-a/ and /-ia/ have a similar distribution, with /-ia/ surfacing after [a]-final stems and /-a/ elsewhere (e.g. [apu-a] 'cover' vs. [tapa-ia] 'recite'). This pattern is holds for ≈94% of stems that take /-a/ or /-ia/, and can be traced back to Proto-Oceanic (see Section 2.2).

Additionally, the allomorphs /-(i)a/ and /-tia/ are the most frequent, and are described as being the most productive. Their distribution is partially conditioned by the prosodic shape of the base of suffixation (where 'base' includes the root and any prefixal material); roughly speaking, /-tia/ is used on longer bases, as seen in examples like [riri-**a**] 'quarrel' vs. [kairiri-**tia**] 'oppose' (Blevins 1994, de Lacy 2003).

2.2 Diachronic development of passive allomorphy

Māori passive allomorphy arose as a result of a historical process of final consonant deletion, which affected all Polynesian languages. If we trace the suffix back to Proto-Oceanic (POc), which precedes Proto-Polynesian (PPn), there used to be two main allomorphs *-ia and *-a, with *-a appearing after vowels other than *a (*i, *e, *o, *u), and *-ia appearing elsewhere (Evans 2001); this is the pattern shown in row (2a) below. The allomorph *-ina was also observed occasionally, although its distribution is unclear. The subsequent loss of

[1] Similar allomorphy is observed in the nominative suffix, which has allomorphs /-ŋa, -Caŋa/. I focus on the passive because it is more productive and well-attested.

[2] Three other allomorphs, /-hina, -fia, -nia/, are rare and each observed for just 1-2 words.

final consonants in PPn, shown in (2b), resulted in ∅∼C alternations. The result is that in Māori, the passive suffix now has many consonant-initial allomorphs of the shape /Cia/.

(2) *Development of Māori passive suffix*

	/ia/	/a/	/Cia/	
a.	*bebe/*bebe-a	*tupa/*tupa-ia	*bikit/*bikit-ia	POc
b.	–	–	biki/bikitia	PPn, Final C deletion
c.	pepe/pepe-a	tufa/tufa-ia	piki/piki-tia	Māori

In words that were *n-final, the observed suffix allomorphs are /-ina/ and /-na/, rather than /-nia/. /-ina/ arose by metathesis of historical *-nia when the verb base ended in *a. /-na/ likely also arose by metathesis of *-nia to *ina, followed by deletion of the *i vowel.

In summary, if no reanalysis has occurred, we should observe the following correspondences: historically vowel-final stems should take /-a/ and /-ia/, with /-ia/ surfacing after [a] and /-a/ surfacing after other vowels (e.g. *fulu/fulu-a*, <POc *pulu). Historically consonant-final stems should take /-Cia/, /-na/, or /-ina/, and the C that is observed should reflect the historic stem-final consonant (e.g. *inu/inu-**m**ia*, <POc *inu**m**). Some POc consonants (*q, *R) were deleted in all positions in Māori; forms that historically ended in these consonants should therefore also behave like vowel-final stems.

3. Reanalysis in Māori

In this section, I compare the distribution of final consonants in POc to the distribution of passive allomorphs in modern Māori. Where the POc and Māori distributions mismatch, this suggests that reanalysis has taken place. For example, [kopi-a] (<POc *kobi**t**) 'doubled', which should take /-tia/ but instead takes /-a/, has undergone reanalysis of t→∅.

The POc data is a set of 1,023 protoforms; forms were sourced from the ACD (Blust et al. 2023) and must have at least 6 cognates in the Oceanic language family. The Māori data is a set of 1167 stem-passive pairs (de Lacy, p.c.), which are primarily sourced from the Williams dictionary, supplemented with data from de Lacy's fieldwork.

3.1 Historical distribution of final consonants (POc)

Table (3) shows the distribution of final segments in POc, organized by the expected Māori allomorph. Vowel-final stems, corresponding to /-(i)a/, were by far the most frequent. /-tia/ and /-hia/ are the next most frequent allomorphs, each taking up 7% of the POc corpus. Even if /-a/ and /-ia/ are treated as separate allomorphs, each are individually more frequent than any of the other allomorphs. Recall that /-ia/ occurs after [a]-final stems, and /-a/ occurs elsewhere. Based on the distribution of final vowels in POc, this means that /-a/ would be the most frequent allomorph (n=347, p=0.34), followed by /-ia/ (n=325, p=0.32).

If reanalysis is generally frequency-matching, it should mainly occur towards /-a/ and /-ia/. As discussed in Kuo 2023, there is also no strong distributional evidence in POc for reanalysis based on other phonological factors like the identity of the stem-final vowel.

(3) *Distribution of final segments in POc*

POc	Allomorph	n	P
vow (or *q, *R)	(i)a	672	0.66
*m	mia	20	0.02
*t,*j, *d	tia	67	0.07
*n, *ñ	na/ina	60	0.06
*r,*l,*dr	ria	36	0.04
*k	kia	52	0.05
*ŋ, *mw	ŋia	40	0.04
*s, *p	hia	66	0.07

3.2 Comparison of POc and Māori

Fig. (4) compares the proportion of allomorphs in POc and Māori; /-na/ and /-ina/ are pooled with the /-Cia/ allomorphs. Recall that in Māori, the distribution of /-(i)a/ and /-tia/ is partly conditioned by the prosodic shape of the base (Section 2.1). As such, the figure also has a column 'moras <=4', which shows the subset of stems that are expected to prefer /-(i)a/ (i.e. with shapes LL, LLL, and LLLL).

Overall, the proportion of stems taking /-Cia/ allomorphs has decreased between POc and Māori, which is unsurprising given that /-Cia/ allomorphs were the least frequent in POc. However, there has been a decrease in /-(i)a/ and increase in /-tia/ that is not predicted by historical distributions. For the subset of stems expected to prefer /-(i)a/, the increase in /-tia/ is weaker, but still present. This suggests that reanalysis away from /-(i)a/ and towards /-tia/ has happened across stems of different prosodic shapes.

(4) *Distribution of passive allomorphs in POc vs. Māori, by prosodic shape of stem*

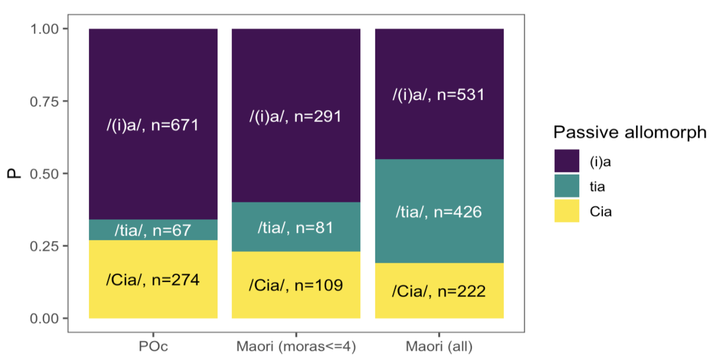
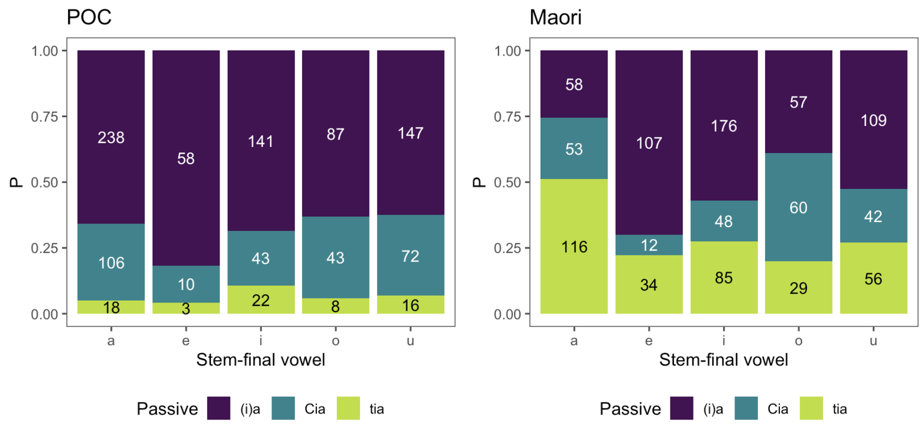


Fig. (5) shows the distribution of passive allomorphs by identity of the stem-final vowel. There is an effect of the stem-final vowel, such that the shift away from /-(i)a/ has primarily happened in [a]-final stems. This suggests that reanalysis has been primarily away from /-ia/ rather than /-a/, since /-ia/ is the vowel-initial allomorph that appears after [a].

In fact, if we look more closely at forms expected to take /-a/ or /-ia/, there has been much more reanalysis away from /-ia/. Table (6) shows form-by-form comparisons of POc protoforms with their Māori reflexes, for just the subset of forms that were expected to take /-a/ or /-ia/. When the expected allomorph is /-ia/, around 63% of forms (n=19/30) have been reanalyzed, compared to 29% for stems expected to take /-a/ (n=22/75).

(5) *Distribution of passive allomorphs in POc vs. Māori by stem-final vowel*



(6) *Mismatches between POc and Māori*

POc	match	n	p
/a/	yes	53	0.71
	no	22	0.29
/ia/	yes	11	0.37
	no	19	0.63

Overall, results of this section suggest that /-ia/ has been reanalyzed more than would be expected given its high frequency in POc. There has also been an overall increase in /-tia/-suffixed forms, particularly for [a]-final stems. I propose that reanalysis of /-ia/→/-Cia/ occurred to avoid outputs that are marked in terms of syllable structure. Because /-tia/ was the most frequent /-Cia/ allomorph, it was most frequently the result of reanalysis. As reanalysis made /-tia/ more frequent, it was extended to more general environments. The result is the modern Māori system, where /-a/ and /-tia/ are the productive allomorphs.

4. Reanalysis as the avoidance of marked syllable structures

In this section, I propose that reanalysis away from /-ia/ can be explained as the combined effect of syllable structure constraints that penalize vowel hiatus and heavy syllables (i.e. ones with long vowels or diphthongs). These constraints are also already active in Māori stem phonotactics, consistent with the active markedness proposal put forth in Section 1.

The marked status of both vowel hiatus and heavy nuclei is well-substantiated. Hiatus avoidance is widespread (e.g. Casali 1997), and typically argued to be aimed at removing onsetless syllables (Blevins 1994). Heavy nuclei, and in particular diphthongs, face conflicting demands between maximizing perceptual distinctiveness and minimizing articulatory effort (Minkova and Stockwell 2003). Below, I use ONSET to penalize vowel hiatus, and *LONGNUCLEUS to penalize heavy nuclei.

(7) *Hiatus in passive forms of [a]-final stems*

		STEM	SUFFIXED	GLOSS	MARKEDNESS
a.	/STEM-ia/	[ho.ka]	[ho.kai.a]	‘run out’	ONS, *LONGNUC
b.	/STEM-tia/	[wa.ha]	[wa.ha.ti.a]	‘raise up’	ONS
c.	/STEM-ina/	[ka.ta]	[ka.tai.na]	‘laugh at’	*LONGNUC

The examples in (7) demonstrate how syllable structure constraints can motivate reanalysis towards /-tia/ (or more generally any /-Cia/ allomorph). First, when an [a]-final stem takes /-ia/ as the passive allomorph, the resulting suffixed form violates both ONSET and *LONGNUC; an example of this is given in (7a). In contrast, [a]-final stems which take /-tia/ violate ONSET but not *LONGNUC, as in (7b); all other /-Cia/ allomorphs would have the same violation profile as /-tia/. In (7c), the stem which takes /-ina/ avoids violations of ONSET, but incurs a violation of *LONGNUC.[3]

Importantly, all suffixed forms violate either ONSET or *LONGNUC, but stems which take /-ia/ violate both and are therefore the most marked. Note that following de Lacy (2003) and Harlow (2007), I adopted a syllabification where forms like /hokaia/ surface as [ho.kai.a], rather than as *[hoka.i.a]. However, even in the latter case, [ho.ka.i.a] violates ONSET twice, and is therefore more marked than other suffixed form options.

Additionally, although I argue that reanalysis is sensitive to ONSET and *LONGNUC, allomorphs which violate these constraints have been maintained in Māori. For example, all of the /-Cia/ allomorphs violate ONSET. This could be because learners are allowed to select allomorphs in a way that reduces markedness, but the allomorphs themselves cannot be changed, unless as part of regular sound change.

4.1 Syllable structure constraints in Māori phonotactics

In this section, I show that the two constraints argued to motivate reanalysis (ONSET and *LONGNUC) are both present in Māori stem phonotactics. This finding is consistent with the active markedness proposal, where only constraints that are already active in the language’s phonological grammar can influence morphophonological reanalysis.

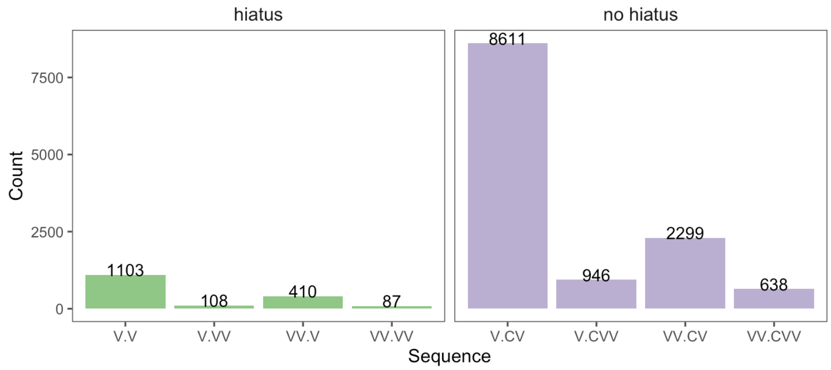
Fig. (8) shows the distribution of syllable-syllable pairs in Māori. This figure is based on 7430 headwords taken from the Williams dictionary, pre-processed to remove prefixes, reduplication, and compounding. Forms are grouped by whether syllables were light (V) or heavy (VV), and by whether or not there was vowel hiatus at the syllable boundary. Both diphthongs and long vowels are represented as VV.

The effects of both *LONGNUC and ONSET are observed in this figure. First, there is a clear dispreference for hiatus; syllable-syllable sequences with hiatus are overall much lower in frequency. Additionally, sequences which include long (VV) nuclei are also dis-

[3] In principle, [a]-final stems could take the allomorph /-a/, resulting in forms like [pa.na.a]. However, this option violates a restriction against sequences of identical vowels that is generally upheld in Māori.

preferred. In fact, monophthongs account for 89% of syllable types in the Williams corpus (n=6047), compared to long vowels (4%, n=248) and diphthongs (7%, n=454).

(8) *Counts of syllable-syllable combinations in the Williams dictionary*



Although not shown here, these patterns were also confirmed using a MaxEnt phonotactic model, following the procedure laid out by Wilson and Obdeyn (2009). The same patterns were also found in Proto-Polynesian, which represents an older stage of Māori. Details of the modeling are discussed in Kuo 2023:ch. 5.2.2.

5. Modeling reanalysis in Māori

In this section, I use a model of reanalysis to test the proposal that reanalysis of Māori passives is sensitive to both frequency and markedness. Following the active markedness proposal, markedness effects are directly derived from phonotactics. I provide a general overview of the model architecture, while detailed descriptions are given in Kuo 2023.

5.1 Model implementation

To formally implement the interaction between frequency-matching and phonotactics, I use Maximum Entropy Harmonic Grammar (MaxEnt; Goldwater and Johnson 2003, Wilson 2006), which is a probabilistic model of phonological learning that uses weighted constraints. A preference for outputs that obey the phonotactics can then be implemented as a *prior* (see White 2017 for a similar implementation).

MaxEnt is well-suited to the type of learning behavior being modeled, because its general mechanism of weighting constraints according to the principle of maximum entropy results in frequency-matching. In this study, the model is trained on stem-passive paradigms, and will match the frequencies of this data. The subsequent addition of a prior allows for us to model frequency-matching that is constrained by markedness.

In addition to modeling the interaction of frequencies and markedness, the model should also capture the cumulative of reanalyses over time. To do this, I adopt an iterated learning paradigm, where one iteration of the model becomes the input to the next iteration (e.g. Ito and Feldman 2022). In the first iteration, the model is trained on the historical (POc) distribution of passive allomorphs. After a set number of iterations, the model is then evaluated on its fit to the modern Māori distribution of passive allomorphs. In the current study, each model was run for 20 iterations, and all model results were averaged over 30 trials.

Model inputs are 500 stem-suffix pairs based on POc, which were simplified and pooled by the identity of the stem-final vowel (/a e i o u/). Conditioning effects of prosodic shape are not considered, so an input like /ito/ represents all stems where the final vowel is /o/. Candidates are all the possible suffixed passive forms. For example, given an input /ito/, possible candidate outputs include [itofia], [itomia], [itotia], [itoa], etc.

The constraint set is composed of violable morpheme exponence constraints of the form 'PASS=/tia/', which demand a particular exponent for a particular morpheme (Kager 1996). To implement phonotactic markedness bias, I include a constraint PHONOTACTICSCORE, which is assigned violations based on the phonotactic well-formedness of suffixed form candidates. The prior is set so that this constraint is given a high preferred weight.

Phonotactic well-formedness scores are assigned using the UCLA Phonotactic Learner (UCLAPL; Hayes and Wilson 2008), which is itself a MaxEnt grammar that learns phonotactic constraint weights. The input to the phonotactic model is a corpus of 1645 Proto-Polynesian protoforms taken from POLLEX (Greenhill and Clark 2011). The corpus was modified to reflect the regular sound changes that have happened between Proto-Polynesian and Māori, and is meant to represent the phonotactics of an earlier stage of Māori.

Two markedness-biased models of reanalysis are compared, differing in the phonotactic grammar used. In the NATURAL CLASS model, the UCLAPL learned 50 constraints and was given no prespecified constraints. In the HIATUS model, the UCLAPL was given five prespecified constraints on syllable structure, and was allowed to learn 45 more constraints. In both models, PHONOTACTICSCORE is biased to have a higher weight than other constraints. These models are compared against respective BASELINE models, which are identical except that all constraints have the same preferred weight.

5.2 Modelling results

Table (9) compares the log-likelihood of the models, fit to modern Māori.[4] Overall, both markedness-biased models outperform the BASELINE model, and the HIATUS model performed the best. This suggests that the improvement in model fit is mainly driven by the hypothesized markedness constraints, *LONGNUC and ONSET.

(9) *Model results: log likelihood*

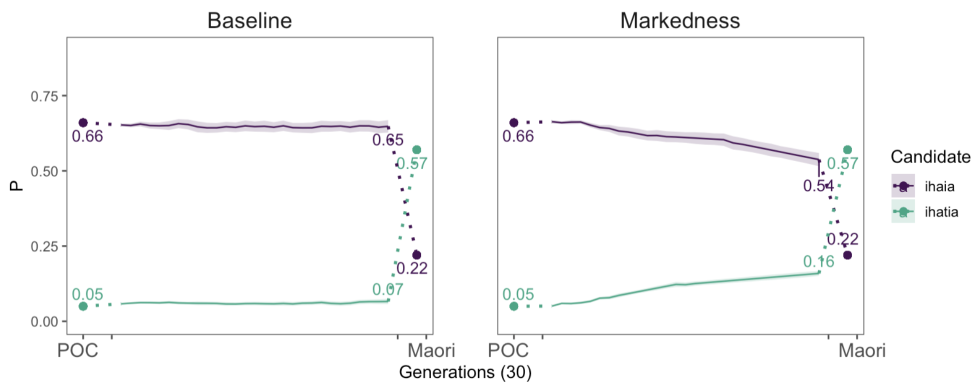
	L	ΔL
BASELINE	-1883.51	–
NATURAL CLASS	-1815.21	68.30
HIATUS	-1702.12	181.39

The HIATUS model differs from the BASELINE primarily in its prediction for [a]-final stems. This is illustrated in Fig. (10), which compares predictions of the BASELINE and

[4] Both BASELINE models behaved similarly, so only the baseline model for the HIATUS model is shown.

HIATUS models for [a]-final stems. In this figure, [iha-ia] represents all [a]-final stems followed by /-ia/. For ease of comparison, only predictions for /-ia/ and /-tia/ are shown.

(10) *Predicted reanalysis in [a]-final stems*



Between POc and Māori, there is a drop in the proportion of stems which take /-ia/ and conversely, an increase in stems which take /-tia/. Because the POc inputs show a strong preference for /-ia/, the frequency-matching model is not able to predict this shift towards /-tia/. On the other hand, the HIATUS model is able to predict change in the right direction.

6. Conclusion

Reanalysis of Māori passives is found to generally be in direction of /ia/→/-tia/, which is unexpected in a purely frequency-matching model. I propose that this change is driven by avoidance of marked syllable structures, and present a MaxEnt model of reanalysis which supports this proposal. I additionally made the distinction between 'universal' markedness and 'active' markedness. The Māori findings are consistent with an approach where only active markedness can influence reanalysis.

References

Blevins, Juliette. 1994. A phonological and morphological reanalysis of the Maori passive. *Te Reo* 37:29–53.

Blust, Robert, Stephen Trussel, and Alexander D Smith. 2023. CLDF dataset derived from Blust's "Austronesian Comparative Dictionary" (v1.2) [data set]. Zenodo. URL `https://doi.org/10.5281/zenodo.7741197`.

Casali, Roderic F. 1997. Vowel elision in hiatus contexts: Which vowel goes? *Language* 73:493–533.

Ernestus, Mirjam Theresia Constantia, and R Harald Baayen. 2003. Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79:5–38.

Evans, Bethwyn. 2001. A study of valency-changing devices in Proto-Oceanic. Doctoral dissertation, Australian National University, Canberra.

Goldwater, Sharon, and Mark Johnson. 2003. Learning OT constraint rankings using a maximum entropy model. In *Proceedings of the Stockholm workshop on variation within Optimality Theory*, 111–120. Stockholm: Stockholm University Department of Linguistics.

Greenhill, Simon J, and Ross Clark. 2011. POLLEX-online: The Polynesian lexicon project online. *Oceanic Linguistics* 50:551–559.

Hale, Kenneth. 1968. Review of Hohepa (1967)—'a profile generative grammar of Maori'. *Journal of the Polynesian Society* 77:83–99.

Hale, Kenneth. 1973. Deep-surface canonical disparities in relation to analysis and change: an Australian example. In *Diachronic, areal, and typological linguistics*, ed. by Henry Hoenigswald. Berlin: De Gruyter.

Harlow, Ray. 2007. *Maori: a linguistic introduction*. Cambridge, UK: Cambridge University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39:379–440.

Ito, Chiyuki, and Naomi H. Feldman. 2022. Iterated learning models of language change: a case study of Sino-Korean accent. *Cognitive Science* 46:e13115.

Kager, René. 1996. On affix allomorphy and syllable counting. In *Interfaces in phonology*, ed. by Kleinhenz Ursula, 155–171. Berlin: Akademie Verlag.

Kuo, Jennifer. 2023. Phonological markedness effects in reanalysis. Doctoral dissertation, University of California, Los Angeles.

de Lacy, Paul. 2003. Maximal words and the Maori passive. In *Optimality theory in phonology: a reader*, ed. by John McCarthy, 495–512. Malden: Blackwell.

Minkova, Donka, and Robert Stockwell. 2003. English vowel shifts and 'optimal' diphthongs: Is there a logical link? In *Optimality theory and language change*, ed. by Eric Holt, 169–190. Dordrecht: Springer.

Sanders, Gerald. 1990. On the analysis and implications of Maori verb alternations. *Lingua* 80:149–196.

White, James. 2017. Accounting for the learnability of saltation in phonological theory: a maximum entropy model with a P-map bias. *Language* 93:1–36.

Williams, Herbert W. 1971. *A dictionary of the Maori language, 7th edition*. Wellington, N.Z.: Government Printer.

Wilson, Colin. 2006. Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive science* 30:945–982.

Wilson, Colin, and Marieke Obdeyn. 2009. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms, Johns Hopkins University, Baltimore, MD.

Jennifer Kuo
jk2767@cornell.edu