Hiatus avoidance and the development of Māori passive allomorphy

Paradigms with conflicting data patterns can be difficult to learn, resulting in child errors (e.g. *mouse/mouses* instead of *mouse/mice* in English). Such errors can be adopted into speech communities, resulting in a type of change over time I refer to as *reanalysis*. Existing models of morphophonology predict reanalysis to be frequency-matching, occurring in a way that matches probabilistic distributions of the input data (e.g. Hare and Elman, 1995; Albright, 2002). I propose that instead, reanalysis is sensitive to both frequency and the reduction of phonological markedness.

Specifically, I present evidence that in Māori (Austronesian, New Zealand), reanalysis of passive allomorphy is not just sensitive to frequency, but also motivated by avoidance of long syllable nuclei and vowel hiatus. Results are confirmed using iterated learning models; I compare two models: one that is frequency-matching, and one which has a markedness learning bias. I find that the latter model performs better in predicting the Māori patterns of reanalysis.

The Māori passive suffix has allomorphs /-a/, /-ia/, /-na/, /-ina/ and /-Cia/ (where C represents a variable consonant). This allomorphy is exemplified in (1) (Hohepa, 1967). Passive suffix allomorphy developed as follows: in Proto-Oceanic (POc), the suffix had two allomorphs *-a and *-ia, whose distributions were phonologically predictable, with *-ia surfacing after stems ending in *a, and *-a surfacing elsewhere. The /-Cia/ allomorphs developed when the Polynesian languages underwent a process of final consonant loss. As a result, consonants were deleted at the end of unsuffixed stems, but maintained in suffixed forms (e.g. **inum/*inum-ia* 'to drink' → *inu/inu-mia*).

(1) *Ergative suffix allomorphy in Māori*

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

Based on this historical development, the Māori allomorph which surfaces should be traceable back to the historic stem-final segment in POc (Pawley, 2001). Historically vowel-final stems are expected to take /-a/ or /-ia/ (e.g. [fao]/[fao-a] < POc *paRo), while historically consonant-final stems are expected to take /-Cia/, where C corresponds to the historic final consonant (e.g. [inu]/[inu**m**ia] < POc *inu**m**). However, the observed alternant often does *not* match the historic POc one; for example, Māori [puru] (<POc *bulu**t**) should have the suffixed form [puru-**t**ia], but instead [puru-a] is observed. These mismatches suggest that language learners have carried out extensive reanalyses.

To investigate the expected direction of reanalyses, I collected 1023 POc forms from the Austronesian Comparative Dictionary (Blust, Trussel, and Smith, 2023). Based on the distribution of final segments in this POc corpus, /-a/ (n=347/1023, 34%) and /-ia/ (n=347/1023, p=0.34) should be the most frequent allomorphs. In a frequency-matching model, we would therefore expect reanalysis to be towards /-a/ and /-ia/ (and away from /-na/, /-ina/, and /-Cia/).

In Māori, /-a/ is in fact very frequent, but /-ia/ is unexpectedly rare; reanalysis has generally been away from /-ia/ and towards another 'default' allomorph. This default (i.e. frequent and productive) allomorph is generally /-tia/. However, at least one dialect has default /-hia/ (Blevins, 1994), and one dictionary lists /-ina/ as the most frequent allomorph (Ngata, 1971).

I argue that reanalysis away from /-ia/ and towards other allomorphs (/-tia/, /-hia/, /-ina/) is motivated by avoidance of long syllable nuclei (long vowels and diphthongs) and vowel hiatus. Consider (2), which shows different syllabified suffixed forms, given an [a]-final stem. Vowel-vowel sequences with falling sonority (e.g. /ai, au/) are syllabified as diphthongs (e.g. /aitu/, [ai.tu]). Other vowel-vowel sequences belong to separate syllables (e.g. /ruaki/, [ru.a.ki]).

When an [a]-final stem is suffixed with /-ia/, the resulting output violates both *Hiatus and *LongNuc, as seen in (2a). In contrast, forms which take another allomorph are always less marked. First, when the suffix is /-Cia/, the resulting output violates *Hiatus but not *LongNuc, as in (2b,c). When the suffix is /-ina/, the output violates *LongNuc but not *Hiatus, as in (2d). Going back to (2a), note that even if /ai/ was syllabified as two syllables (i.e. [pa.ka.i.a]), the resulting output would have multiple violations of *Hiatus and still be more marked than competing outputs.

(2) *Markedness of different suffixed form outputs*

	CIA	**Example suff.**	***Hiatus**	***LongNuc**
a.	ia	[pa.kai.a]	*	*
b.	tia	[pa.ka.ti.a]	*	
c.	hia	[pa.ka.hi.a]	*	
d.	ina	[pa.kai.na]		*

These results are confirmed using a model of reanalysis implemented in MaxEnt Harmonic Grammar (Goldwater and Johnson, 2003). To simulate the cumulative effects of reanalyses over time, the model is iterated. In other words, at each "generation", a learner induces a grammar based on input data, then uses this grammar to generate data that is passed down to the next generation. Two models are compared: 1) a frequency-matching baseline model, and 2) a markedness-biased model where *Hiatus and *LongNuc are biased to have high weight using the method laid out by Wilson (2006). I find that the markedness-biased model performs significantly better than the purely distributional baseline model in predicting the directions of reanalysis in Māori.

Notably, the iterated learning paradigm also predicts variation across model runs. The markedness biased model, in particular, predicts reanalysis to be variably towards one of /-tia/, /-hia/, or /-ina/, which are the historically most frequent allomorphs after /-a/ and /-ia/. These results actually match the dialectal variation found in Māori (as /-tia/, /-hia/, and /-ina/ are the different 'default' allomorphs), showing the utility of iterated learning for modeling dialect divergence.

References. [1] A. C. Albright. "The identification of bases in morphological paradigms". PhD thesis. UCLA, 2002. [2] J. Blevins. "A phonological and morphological reanalysis of the Maori passive". In: *Te Reo* 37 (1994), pp. 29–53. [3] R. Blust, S. Trussel, and A. D. Smith. *CLDF dataset derived from Blust's "Austronesian Comparative Dictionary" (v1.2) [Data set]*. Zenodo. 2023. [4] S. Goldwater and M. Johnson. "Learning OT constraint rankings using a maximum entropy model". In: *Proceedings of the Stockholm workshop on variation within Optimality Theory*. 2003, pp. 111–120. [5] M. Hare and J. L. Elman. "Learning and morphological change". In: *Cognition* 56.1 (1995), pp. 61–98. [6] P. W. Hohepa. *A Profile Generative Grammar of Maori*. Waverly Press, 1967. [7] H. M. Ngata. *English-Maori Dictionary*. Wellington, N.Z.: Learning Media, 1971. [8] A. Pawley. "Proto Polynesian *-CIA". In: *Issues in Austronesian Morphology*. Pacific Linguistics, 2001. [9] C. Wilson. "Learning phonology with substantive bias: An experimental and computational study of velar palatalization". In: *Cognitive science* 30.5 (2006), pp. 945–982.