

A reevaluation of the Māori-passive challenge to simplicity-based learning

Saar Yahalom, Tel-Aviv University Saaryahalom@mail.tau.ac.il

Summary: Simplicity-based learning in morpho-phonology (Chomsky & Halle 1968) recently returned in the form of Minimum Description Length (MDL; Rissanen 1978), showing significant success (Rasin et al. 2021). The Māori passive poses an old challenge to simplicity-based learning, where the *prima facie* simplest phonological analysis is shown by Hale (1973) not to be the one acquired by speakers. This was taken as evidence against simplicity-based learning in phonology, and various learning pressures were suggested to account for it. We take a closer look at the relevant possible grammars and attempt to show that, given newer empirical observations about Māori, the grammar learned by speakers is overall simpler – eliminating the problem.

Background: An early generative theory of language acquisition was that children acquire the *simplest* grammar that adequately generates the utterances they have observed (Chomsky 1965; Chomsky & Halle 1968). While this theory seemed promising in accounting for certain aspects of morpho-phonological acquisition, in the 1970-80s various challenges were raised – amongst them from optionality (Dell 1981; Baker 1979), and from the Māori passive. In later years the simplicity metric was abandoned. However, simplicity has recently returned in the form of MDL, which supports the learning of phonological rules and morphological segmentation from distributional evidence, and successfully accounts for challenges such as optionality and opacity (Rasin et al. 2021). MDL's recent success motivates a renewed look at the challenges previously presented to simplicity-based learning, in order to examine whether they pose a challenge to MDL as well and whether they can be informative with respect to the nature of representation (e.g., Rasin & Katzir 2020; Marco et al. 2025).

Māori – the basic problem: The puzzle concerns alternations such as the following (1):

- (1) Māori verbal paradigm (Hale 1973; Blevins 1994)

Verb	Passive	Gerund		Verb	Passive	Gerund	
mahue	mahuētia	mahuetaŋa	'leave'	inu	inumia	inumaŋa	'drink'
hopu	hopukia	hopukaŋa	'catch'	patu	patua	patuŋa	'strike, kill'

The *prima facie* simplest phonological analysis of the alternation, as would be expected in an introductory phonology class, goes along the following lines (shown as a rule-based grammar in (2)): a single UR /-ia/ for the passive suffix, and a (mildly abstract) UR for each root containing the thematic consonant, e.g. /hopuk/. Word-final consonant deletion accounts for the alternation in the stems, and vowel syncope for the [-a] realization of the suffix in the final example.

Hale (1973) called this the “*purely phonological alternative*”, and showed it is not the one acquired by Māori speakers. Instead, he argued (following Hohepa 1965) that the modern Māori grammar has a default /-tia/ suffix (used in loanwords, for passivizing nouns & adjectives, for derived causatives, etc.), and that other suffixes (e.g. /-kia/) are memorized for each verb. This can be implemented using verb-class diacritics which condition suffix-alternant selection (shown using Distributed Morphology style Vocabulary Insertion rules in (3)). Hale called this the “*conjugation alternative*”. Some examples of loanwords taking /tia/ are shown in (4)

- (2) *Phonological alternative*

Roots	/mahuet/, /hopuk/, /patu/
Suffixes	/-ia/ (default)
Rules	$C \rightarrow \emptyset / _\#$ $V \rightarrow \emptyset / V + _\$

- (3) *Conjugation alternative*

Roots	/mahue/, /hopu/[_{+K}], /patu/[_{+V}]
Suffixes	/-tia/ (default), /-kia/...
Rules	$[+PASS] \leftrightarrow /kia/ \quad / [+K] _\dots$ $[+PASS] \leftrightarrow /tia/ \quad (\text{default})$

- (4) **Loan-Verb** **Passive**

riiŋi	riiŋitia	'telephone'
parakitihi	parakitihitia	'practice'

Historically, the thematic consonants were indeed part of the root in Proto-Oceanic – surfacing both in isolation and before *V*-initial suffixes – but were lost in the unsuffixed form by the process of word-final consonant deletion in the Polynesian languages (Harlow 2007). Then, Hale argued, reanalysis occurred by following generations learning the paradigm, rebracketing [mahuet-ia] to [mahue-tia], and regularizing the /tia/ alternant. Thus, the puzzle is to explain the learning of the seemingly more complex *conjugation alternative*, based on the basic paradigm.

Various theoretical conclusions were drawn from this puzzle: It was indeed taken to be evidence against simplicity-based learning and the evaluation metric (Campbell 1974; Hyman 1975), and the Generative Phonology methodology at large (Lass 1984); or in favour of a functional pressure against ambiguity in language (Kaye 1975). In the generative literature, it was attributed to learning pressures against abstractness (Hale 1973; Halle 1979; Kenstowicz & Kisseberth 1977; Kiparsky 1971/82)

Additional Facts about Māori: Two descriptive elements about the Māori passive (unavailable to the theoretical discussion of the 70s-80s) play a role in our examination and the solution for the puzzle: (1) suffix-alternant frequency

statistics from Sanders (1991), which enable measuring the complexity of a representative part of the verbal lexicon; (2) the generalization about phonologically-conditioned suppletion by Blevins (1994). The latter is the observation that roots taking /tia/~/ia/ terminations are not generally contrastive – but are rather in complementary distribution, predictable by root-length. Short verbal roots (2 vowels) take /ia/, while others take /tia/. This is true for $\approx 95\%$ of relevant native verbs, and for all loanwords checked. Compare the loanwords in (4) with the following (5):

(5)	Loan-Verb	Passive	
	kiki	kikia	'kick'
	pine	pinea	'pin'

Solution - lexicon compression: In MDL learning, simplicity is measured (in particular) by the representational length of the lexicon. This motivates morphological segmentation in general because, for example, storing 3 roots and 1 affix in the lexicon is strictly shorter than representing 3 roots & 3 inflected forms. To explicitly measure the lexicon's length of various grammars, they are converted into string form, whose binary encoding length can be directly calculated (Rasin et al. 2021). We used this methodology to compare the lexicons of various observationally adequate grammars for the basic paradigm, and indeed found that a lexicon for Blevins' suppletion grammar is the simplest. The rule component is roughly equivalent between the grammars (see McCarthy 1981); grammar fit to the data, which MDL evaluates but the SPE-metric doesn't, is required to rule out a grammar left out for space reasons.

The table in (6) presents cost calculations for selected grammars. For each grammar (column) and root type ([mahuetia / hopukia / patua]), it contains the representation for alternant selection specified in the lexicon (a thematic consonant in UR, or verb-class diacritic), under the assumption that the default need not be specified. The final row lists the binary encoding length for these specification aggregated by their relative frequency. The simplest lexicon is the one with the lowest length measure. Note that besides Hale's conjugation alternative with default /tia/ (a), we must also consider a conjugation-based grammar with default /ia/ (b), a more common alternant by Sanders' count.

	Gram.	Type Freq. (Sanders)	Phonological deflt. /ia/	Conjugation (a) deflt. /tia/	Conjugation (b) deflt. /ia/	Suppletion (Blevins) deflt. /tia/~/ia/
(6)	-tia roots	31%	/t/	-	[+T]	5% [+T]
	-kia, -mia, -hia, ...	25%	/k/	[+K]	[+K]	[+K]
	-(i)a roots	44%	-	[+V]	-	8% [+V]
	Aggregate (bits / 1000 verbs)		2187.9	2292.1	1860.3	1040.3

The main insight is that the length-based suppletion enables compression of the lexicon. The speaker need not store a separate /t/ for hundreds of roots such as [mahuetia, arohatia] 'leave-PASS, like-PASS', storing them instead as /mahue, aroha/; at the same time, true vowel-final roots such as [patua, taaia] 'strike/kill-PASS, beat/tattoo-PASS' need not be marked as to avoid them wrongly taking /tia/ [*patutia].

Conclusions: A closer look at the relevant grammars showed that the simplest grammars is actually the one chosen by learners, under standard representational assumptions. This accounts for Hale's learning puzzle. Thus – contrary to previous claims – Māori is not a problem for simplicity-based learning in morpho-phonology.

Baker, Carl L. 1979. "Syntactic theory and the projection problem." *Linguistic inquiry* 10 (4). **Blevins, Juliette.** 1994. "A Phonological and Morphological Reanalysis of the Maori Passive." *Te Reo* 37. **Campbell, Lyle.** 1974. "Theoretical implications of Kekchi phonology." *International journal of American linguistics* 40 (4). **Chomsky, Noam.** 1965. *Aspects of the Theory of Syntax*. MIT press. **Chomsky, Noam, & Morris Halle.** 1968. *The sound pattern of English*. New York: Harper / Row. **Dell, François.** 1981. "On the learnability of optional phonological rules." *Linguistic Inquiry* 12 (1). **Hale, Kenneth.** 1973. "Deep-Surface Canonical Disparities in Relation to Analysis and Change: an Australian Example." In *Diachronic, areal, and typological linguistics*. Current Trends in Linguistics v. 11. **Halle, Morris.** 1979. "Formal vs. Functional Considerations in Phonology." In *Studies in diachronic, synchronic, and typological linguistics: Festschrift for Oswald Szemerényi on the occasion of his 65th birthday*. **Harlow, Ray.** 2007. *Māori: A linguistic introduction*. Cambridge University Press. **Hohepa, Patrick Wahanga.** 1965. "A Profile Generative Grammar of Maori." PhD diss., Indiana University. **Hyman, Larry M.** 1975. *Phonology: theory and analysis*. ISBN: 9780030121418. **Kaye, Jonathan.** 1975. "A Functional Explanation for Rule Ordering in Phonology." In *Papers from the Parasession on Functionalism*. Chicago Linguistic Society. **Kenstowicz, Michael, & Charles Kisseberth.** 1977. "The Problem of the Abstractness of Underlying Representations." In *Topics in phonological theory*. Academic Press. **Kiparsky, Paul.** 1971/82. "Historical Linguistics." In *Explanation in Phonology*. <https://doi.org/10.1515/9783111666242>. **Lass, Roger.** 1984. *Phonology: An introduction to basic concepts*. Cambridge Textbooks in Linguistics. **Marco, Eyal, Roni Katzir & Ezer Rasin.** 2025. "The complementary-distribution challenge for phonological learning: The case of [h] and [ŋ] in English." Available at lingbuzz/009103. **McCarthy, John J.** 1981. "The Role of the Evaluation Metric in the Acquisition of Phonology." Chap. 8 in *The Logical Problem of Language Acquisition*. MIT Press. **Rasin, Ezer, Iddo Berger, Nur Lan, Itamar Shefi & Roni Katzir.** 2021. "Approaching explanatory adequacy in phonology using Minimum Description Length." *Journal of Language Modelling* 9 (1). **Rasin, Ezer, & Roni Katzir.** 2020. "A conditional learnability argument for constraints on underlying representations." *Journal of Linguistics* 56 (4). **Rissanen, Jorma.** 1978. "Modeling by shortest data description." *Automatica* 14 (5). **Sanders, Gerald.** 1991. "Levelling and Reanalysis in the History of Polynesian Passive Formations." *The Journal of the Polynesian Society* 100 (1).