A Puzzling Case: Māori Passives

Selin Alkan

1. Introduction

Te reo Māori (i.e., the Māori language), mainly spoken by the indigenous Māori people native to New Zealand and one of the official languages in the country, displays a curious morphological case that has intrigued many linguists so far: the many shapes of the passive suffix. Although different sources refer to a varying number of passive endings in Māori (a total of 17 suffixes in Williams (1957): *-a, -ia, -hia, -kia, -mia, -ngia, -ria, -tia, -whia, -na, -ina, -hina, -kina, -rina, -whina, -nga, -hanga*; 7 in Hale (1973): *-tia, -kia, -mia, -ŋia, -ria, -hia, -a*; 9 in Blevins (1994): *-tia, -kia, -mia, -ŋia, -ria, -hia, -fia, -ia, -a*; 12 in Biggs (2013): *-a, -ia, -hia, -ina, -kia, -mia, -na, -nga, -ngia, -ria, -tia, -whia*), there is an agreement about two main points: that the passive suffix *-tia* has a default status, and that there are no obvious rules what the passive form of a given stem will be.

Being a curious phenomenon, Māori passives are also being investigated in the current project (Gorman et al., 2022) which includes similar "problematic" cases from 9 other languages. This paper provides the reader with a synthesis of what the project (Gorman et al., 2022) is about and how work on the Māori data has been conducted, how passive endings in Māori are analyzed in two seminal studies by Kenneth Hale (1973) and Juliette Blevins (1994), and an approach to this construct from the Tolerance Principle (Yang, 2005) perspective. In Section 2, I provide further detail about the project including the sources used and the overall process of data preparation. I present a summary of how Hale (1973) and Blevins (1993) analyze the Māori passives in Section 3. Lastly, in Section 4, I examine the data used in the project through the Tolerance Principle perspective (Yang, 2005) and provide a brief analysis.

2. The Project (Gorman et al., 2022) and Māori Data Preparation

Morphological generation concerns the mapping between a lemma (the citation form of a word) and its lexeme(s), the set of all the inflected or alternating forms of a lemma, in a given paradigm based on certain necessary specifications such as inflectional rules and parts of speech. Although one may expect computational models to carry out morphological generation tasks without much trouble, Gorman et al. note that such models perform poorly in the following two situations where there is the problem of unpredictability: inflectional patterns which are based on properties inherent to lexemes that are not predictable by the models and abstract morphological patterns the forms of whose lemmas do not provide any structural clues for prediction. In their study, Gorman et al. propose to address these issues, in addition to increasing the quality of the data available for such tasks, by creating morphological generation tasks in 10 languages each of which is believed to have a difficult-to-handle lemma-inflected form mappings. Given its complexity, Māori verbs and their passive forms constitute one of the problem sets in the study.

The number of available sources that is useful for this type of research is limited for Māori; therefore, the Māori dataset is not as large as it is for some other languages in the study. However, we still had reliable sources to digitize and/or extract data from. It is important to note

that not all passives are formed by inflection via one of the suffixes given above in Māori; some words undergo stem changes such as partial or complete reduplication of some segments (e.g., *totohe* "to argue" > *toohea*) and/or loss of some others (e.g., *momotu* "to chop off" > *motuhia*), and those words are also included in our dataset.

Most of the verb-passive pairings that compose our dataset come from *English – Māori | Māori – English Dictionary* by Bruce Biggs (2009), and the rest comes from the dataset ʻŌiwi Parker Jones (2008) created based on *The Revised Dictionary of Modern Māori* by P. M. Ryan (1989) for their study titled *Phonotactic Probability and the Māori Passive: A Computational Approach*. I also consulted the online version of Te Aka Māori-English, English-Māori Dictionary and Index (Moorfield, 2022) in cases such as when Biggs (2009) and Ryan (1989) disagreed on the passive form of a verb, or when there were more than one passive endings given in those two sources for one single verb.

To extract the verb-passive pairs from Biggs (2009), I perused both sections of the dictionary and stored the pairs in a spreadsheet. Note that, in Māori, it is possible to passivize non-verb words (e.g., nouns or adverbs); however, to build a consistent dataset, we decided to include only the verbs (what Biggs (2009) calls 'universals') and their passive forms. In the spreadsheet, I entered the verbs that take a passive suffix in the second column separating the ending with a hyphen while leaving their corresponding first column cells empty. As for those verbs that go through stem changes when passivized, I stored the lemmas and the inflected versions in the first and second columns respectively. Two examples are given in Figure 1.

	where-a
whakarere	whakareerea

Figure 1. *where* meaning "to overcome" takes the *-a* suffix in the passive while *whakarere* "to abandon" has reduplication and the addition of a letter, *-a*, when passivized.[1]

The empty cells for the verbs that only take suffixes in the passive were later filled in via a short Python script that removes the suffixes. Some verbs in Biggs (2009) have optional segments given within parentheses in the dictionary; within-word optional segments were kept in our dataset (e.g., *tuu(w)hiti-tia* "expel" is included as *tuuwhiti-tia*) while those that come before or after verbs were omitted (e.g., *(aata) whakaaro-tia* "consider" is taken as *whakaaro-tia*). Verbs within compound structures are also included only by themselves (e.g., *(whaka)uu-ngia ki uta* "land" is included as *whakauu-ngia*). Lastly, in Biggs (2009), some words are suffixed with *-tia* despite being in question form (e.g., *aha-tia?* "what?", *peehea-tia?* "like how?"), which were excluded from the dataset.

Digitizing Biggs (2009) yielded 854 word pairs in total. Ryan (1989) dataset has 481 distinct words given in the passive; however, since most of them are the same as what was already given

[1] Since, unlike for the other affixes in the dictionary, Biggs (2009) did not use a hyphen to separate the final *-a* from the passive *whakareerea*, I am not sure if it is the same as the passive ending *-a* which some verbs that do not go through stem change take. However, all verbs undergoing stem change in the passive end in *-a* through the addition of different combinations of letters that all look like one of the passive endings listed in *Introduction*.

in Biggs (2009), the final version of our dataset has 915 verb-passive pairs after removing the duplicates with the following command line script in (1):

(1) `sort -u BRsplit.tsv -o BRsplit.tsv`

There were two main challenges to take care of before successfully merging the two datasets. First, while Biggs (2009) used double vowels to indicate long vowels, Ryan (1989) used a macron. Since using macrons seems to be the more common practice (Māori Language Commission, 2012, Te Whare Wānanga o Waitaha | University of Canterbury (UC), 2020). I manually converted all the double vowels to single vowels with a macron in the Biggs (2009) dataset except for 13 pairs since they constitute exceptional cases according to Te Taura Whiri i te Reo Māori Guidelines for Māori Orthography (2012) and University of Canterbury Te Reo Māori Style Guide (2020). Second, unlike Biggs (2009), Ryan (1989) provides more than one passive ending for some verbs, and so does Te Aka (2022). To ensure optimal computational performance, it was necessary to select only one passive ending per verb. To this end, I compared the entries in all three sources and picked the suffixes in agreement prioritizing Biggs (2009) and Te Aka (2022) over Ryan (1989) in cases of contrast. Lastly, aiming to increase the word count, I emailed some contacts given on the Te Aka website (2022); however, my attempts did not yield any helpful results.

3. Māori Passives According to Hale (1973) and Blevins (1994)

Kenneth Hale (1973) analyzes Māori as part of a larger study, *Deep-Surface Canonical Disparities in Relation to Analysis and Change: An Australian Example*, where he is intrigued by the considerable typological diversity of the Australian languages despite their generic relation to one another, and how language change takes place. Hale (1973) gives the following examples from the language:

(2)

verb	*passive*	
awhi	awhitia	‘to embrace’
hopu	hopukia	‘to catch’
aru	arumia	‘to follow’
tohu	tohuŋia	‘to point out’
mau	mauria	‘to carry’
wero	werohia	‘to stab’
patu	patua	‘to strike, kill’
kite	kitea	‘to see, find’

Hale (1973) claims that at some point in the history, Polynesian languages acquired a phonological rule in their grammars which affected words ending in consonants; the rule’s effect was to delete word-final consonants:

(3) Final Consonant Deletion Rule
$C \rightarrow \varnothing$ / __ #

The rule in (3) implies that the consonants that are seemingly associated with the passive suffixes in the examples in (2) were once a part of the verb stems themselves. To account for the passive endings, Hale (1973) proposes two alternative explanations. The first one, *phonological alternative*, argues that the underlying representations of the verbs have the final consonants themselves (e.g., *awhit*), the passive ending has only two alternants (i.e., *-ia* following a consonant, e.g., *awhit + ia* and *-a* following a vowel, e.g., *patu + a*), and the final consonant deletion rule applies when a consonant-final word ends in word boundary (e.g., *awhit #* → awhi #).

Hale's second proposal, *conjugation alternative*, suggests that the once-stem-final consonants are now assigned to the beginnings of the passive terminations, which leads to a proliferation of suffixes (e.g., *-tia*, *-kia*, *-ngia-*, and so on) instead of only two alternants (i.e., *-ia* and *-a* as suggested by the phonological alternative), all stems end in vowels, and each stem has its own diacritic feature to ensure correct passive ending attachment. Lastly, Hale (1973) also suggests that *-tia* is now the regular passive ending of all the alternants.

In her paper titled *A Phonological and Morphological Reanalysis of the Maori Passive*, Juliette Blevins (1994) gives a brief summary of the phonological and lexical solutions by Hale (1973) and a phonological solution by Sanders (1990, 1991) and argues for a phonological and morphological analysis of the Māori suffixes that she describes as a merger of the lexical and phonological proposals by Hale (1973, 1991) and Sanders (1990, 1991) respectively. According to Blevins' proposal (1994), the Māori passive suffix comes in two lexical forms: one vowel-initial form and one consonant-initial form.

Based on well-motivated phonological and morphological grounds, Blevins (1994) suggests the following:

(4) Both /-ia/ and /-tia/ are default passive forms in Māori,
(5) The underlying phonological representations (UR) of /-ia/ and /-tia/ are as shown below:
   a. UR of /-ia/:
      ia
      | |
      -VV
   b. UR of /-tia/:
      ia
      | |
      -CVV
(6) The C-slot of (5.b) will either be filled by the floating final consonant of a stem, or by phonological default rules.

4. Māori Passives through the Tolerance Principle Lens (Yang, 2005)

In *On Productivity*, Charles Yang (2005) describes language learning as a remarkable process since exceptions do not prevent the child from figuring out systematic regularities and using them productively in their language. Towards the goal of building a learning model that is able to learn both productive processes and unproductive exceptions, Yang (2005) develops a decision

procedure named the *Tolerance Principle* that can determine whether a linguistic process is productive or not.

Simply put, if there is a generalization observed in a set of N words, M of which are exceptions, the Tolerance Principle gives us a threshold value, M_c, at which a rule for the generalization becomes productive if $M \leq M_c$ based on the following theorem:

(7) $M_c \approx N / \ln N$

Yang (2005) argues that the default rule is, by definition, necessarily productive. Making the assumption that *-tia* is the default passive suffix in Māori, we can look into whether the Tolerance Principle recognizes passivization by the attachment of *-tia* as a productive procedure in Māori.

The passive forms of each one of the 915 verbs in our dataset are constructed either through the addition of one of the 13 passive suffixes given in Biggs (2009), Ryan (1989) or both, or via stem change. Another short Python script that I wrote gave us the counts of the words per passive suffix and the *irregulars* (those verbs that go through stem change) presented in Table 1:

Table 1. Counts of verbs per passive suffix and stem change

Māori Passive Suffix	*Count of Verbs*
-a	345
-tia	308
-hia	56
-na	38
-ngia	35
-ria	32
Irregulars (stem change)	31
-ia	26
-ina	18
-kia	15
-mia	8
-kina	1
-tanga	1
-nga	1
TOTAL	*915*

Considering *-tia* is the default suffix, $N = 915$, the rule following items, $N - M$, is 308, the number of exceptions, M, turns out to be 607. Plugging our N into the theorem given in (7), we get the following:

(8) $M_c \approx 134$

Since $607 \not\leq 134$, our *-tia* rule is not recognized to be productive by the Tolerance Principle (Yang 2005).

This situation with Māori passives seems similar to the German plurals problem Yang (2005) mentions; Yang (2005) reports that, according to the CELEX database, only 7% of nouns (by type) make up the default *-s* class of the German noun plural system, and that this implies that the remaining 93% of the nouns would have to be stored in memory. However, Yang (2005) also reports that the German plural system is in fact made up of a cascade of rules some of which are productive, and the "add -s" rule applies when none of the more specific rules applies for a noun. Lastly, Yang (2005) also notes that most of the words in the "add -s" category in German plurals are foreign words, which reminds me of the agreement between Hale (1973) and Blevins (1994) on the point that borrowings from English take the *-tia* ending in the passive in Māori.

It is conceivable to think that the results of the Tolerance Principle on Māori passives will be different when applied to a larger corpus. However, the two similarities noted for the German *-s* and the Māori *-tia* are also suggestive of a parallelism between these two suffixes considered default in their respective processes. Exploring this similarity in future research could yield useful insights about the Māori puzzle.

Bibliography

Biggs, B. (2009). *English – Māori | Māori – English Dictionary*. Auckland: Auckland University Press.

Biggs, B. (2013). *Let's Learn Maori*. Auckland: Auckland University Press.

Blevins, J. (1994). A Phonological and Morphological Reanalysis of the Maori Passive. *Te Reo*, 29-53.

Gorman et al. (2022). In preparation.

Hale, K. (1973). Deep Surface Canonical Disparities in Relation to Analysis and Change: An Australian Example. *Current Trends in Linguistics*, 401-458.

Jones, ‘Ō. P. (2008). Phonotactic Probability and the Māori Passive: A Computational Approach. *Proceedings of the Tenth Meeting of ACL Special Interest Group on Computational Morphology and Phonology* (pp. 39-48). Columbus, Ohio: Association for Computational Linguistics.

Māori Language Commission. (2012). *Te Taura Whiri i Te Reo Māori Guidelines for Māori Language Orthography*. Te Taura Whiri i te Reo Māori, Māori Language Commission.

Moorfield, J. C. (2022). *Te Aka Māori Dictionary*. Retrieved from Te Aka Māori-English, English-Māori Dictionary and Index: https://maoridictionary.co.nz/

Ryan, P. M. (1989). *The Revised Dictionary of Modern Māori*. Auckland: Heinemann Education.

Te Whare Wānanga o Waitaha | University of Canterbury (UC). (2020). *UC Te Reo Māori Style Guide*. Retrieved from https://www.canterbury.ac.nz/style-guide/uc-te-reo-maori-style-guide/

Williams, H. W. (1957). *A Dictionary of the Maori Language*. Retrieved from New Zealand Electronic Text Collection: https://nzetc.victoria.ac.nz/tm/scholarly/tei-WillDict.html

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook 5*, 265-302.