Are Māori Thematic Consonants Predictable in Passives? A Reexamination of /-mia/, /-ria/, and /-hia/ Verbs

September 2023

Abstract

Māori, categorized as an Eastern Polynesian language, has garnered significant attention in contemporary linguistic studies due to its passive verb forms. Linguists have suggested up to seventeen distinct suffixes associated with the active/passive paradigm, prompting researchers of diverse backgrounds to explore the most appropriate means of describing and analyzing the status of the "thematic consonant" found in the passive form but absent in the active (Williams, 1988; Biggs, 1961; Hale, 1973; Ryan, 1989; Blevins, 1994). The probable unpredictability of the passive endings (Hale, 1973; Blevins, 1994) has been questioned by Moorfield (1988, p. 66), Harlow (2007, p. 117), and Parker Jones (2008). Following Parker Jones's approach (2008), this research endeavors to uncover potential characteristics or subregularities present in Māori verb roots that could predict their respective passive forms. I approach the problem from two complementary standpoints: a historical examination of thematic consonant shifts in Māori verb bases, if any, followed by a synchronic linguistic feature-based analysis of the data facilitated by computer-mediated techniques.

1 Introduction

Te reo Māori, the Māori language, is the traditional language of the Māori people of *Aotearoa*, New Zealand, and one of the official languages of the country. In the context of modern linguistic studies, Māori is, perhaps, best known for its passive verb forms (Hale, 1973, 1991; McCarthy, 1981; Sanders, 1990, 1991; Blevins, 1994). As presented in table 1, (where /ŋ/ = *ng*), as many as seventeen distinct suffixes have been associated with the active/passive paradigm, with a number of linguists from different backgrounds attempting to determine the most appropriate way to describe and analyze the status of the thematic consonant[1] present in the passive form but absent in the active, as in pairs like: *korēro* 'speak', *korērotia* 'was spoken' with thematic consonant *-t-*; *tango* 'remove', *tangohia* 'was removed' with thematic consonant *-h-*; or *inu* 'drink', *inumia* 'was drunk' with thematic consonant *-m-* (Williams, 1988; Biggs, 1961; Hale, 1973; Ryan, 1989; Blevins, 1994).

Although different sources refer to a varying number of passive endings in Māori as seen above, there is a general agreement about two main points: that the passive suffix /-tia/ has earned a default status in at least one major dialect (see Blevins, 1994, on dialect differences), and that there are no obvious phonological rules[2] determining the passive suffix that a stem will select, so much so that Hale (1973) suggested taking the thematic consonant to be lexically specified as part of the stem.

Prior to delving deeper into the intricacies of the present study, it is timely to address two important considerations: first, what is the "natural history" behind this pattern in Māori passives, as observed in table 1; and second, what the implications that this pattern bears for the child acquiring Māori as their native language?

The natural history of a synchronic sound pattern refers to the process of the imperfect language transmission through generations of speakers, which naturally results in such changes in

[1]Such stem-final consonants that are supported by suffixes are known as "thematic consonant", and I continue this practice in the literature without subscribing to any certain analysis.

[2]What is being referred to as a "phonological rule" is a purely phonological process as opposed to one that involves suppletive allomorphy.

Table 1: Māori passive endings according to five different sources.

	Williams (1988)	Biggs (1961)	Ryan (1989)	Hale (1973)	Blevins (1994)
1)	-a	-a	-a	-a	-a
2)	-ia	-ia	-ia		-ia
3)	-tia	-tia	-tia	-tia	-tia
4)	-hia	-hia	-hia	-hia	-hia
5)	-kia	-kia	-kia	-kia	-kia
6)	-mia	-mia	-mia	-mia	-mia
7)	-ngia	-ngia	-ŋia	-ngia	-ŋia
8)	-ria	-ria	-ria / ɾia	-ria	-ria
9)	-na	-na	-na		
10)	-ina	-ina	-ina		
11)	-nga	-nga			
12)	-whia	-whia			-whia
13)	-hina				
14)	-kina				
15)	-rina				
16)	-whina				
17)	-hanga				

the language, given its discontinuous aspect (Blevins, 2004; Hale, 1973). The alternating consonants of the surface patterns that become apparent only in the passive forms of Māori words in the active-passive paradigm is due to a sound change of word-final consonant deletion that took place in a language ancestral to Māori (and other Polynesian languages). Evidently, this pattern is a direct reflection of a historical sound change. Furthermore, this sound change is one of the primary innovations that characterize Proto-Oceanic, given its onset subsequent to the break-up of Proto-Oceanic, a reconstructed ancestral language which Proto-Polynesian belongs to as one of the lower-order subgroups. Consequently, the effects of this word-final consonant deletion rule is widely observable in Oceanic languages.[3] Despite the robust historical account of the sound pattern under discussion, however, it is not as easy to provide an explanation of how the structure is analyzed synchronically.

Within the paradigm of generative grammar, the knowledge possessed by a native speaker of

[3] Lynch et al. (2011) gives the count of Oceanic languages to be between 450 and 600 depending on how one counts them.

their language, along with the evolutionary shifts transpiring within the language, are construed as a grammar of rules. That is, the foundational structure of a language comprises rules, and linguistic changes that occur over time stem from the incorporation of new rules or the elimination (or simplification) of pre-existing ones within the linguistic system. We have already determined the exact rule that led to the change that Māori has inherited. Evidently, the existence of languages is contingent upon their utilization by individuals and their successive transmission to new generations of language acquirers. While this assertion finds consensus, it is paramount to underscore that language is not transferred with absolute fidelity to its original rules and attributes as previously indicated. The child acquires a language through the linguistic input they receive from proficient adult speakers; however, this input constitutes merely a partial subset of the comprehensive linguistic knowledge possessed by the adults. In other words, the child is exclusively exposed to a representative sample of the grammatical constructs present within the cognitive framework of the adult speakers. Consequently, the task facing the language-acquiring child involves the construction of an optimal grammar of rules, endowed with the capacity to generate the entirety of grammatically correct direct and indirect linguistic input received. Besides the fact that this is an exceedingly difficult task that young children miraculously accomplish within very short periods of time after being born, whether all sorts of linguistic tasks are at more or less the same level of difficulty for children is an empirical question to think about. From the perspective of a linguist, some languages have especially difficult properties to learn, and the Māori passives is one of them.

The child acquiring Māori encounters words exhibiting suppletive allomorphy; that is, passivizable Māori words show a variety of consonants in one form (i.e., in the passive voice) that are absent in their corresponding uninflected forms (i.e., in the active voice). From a phonological perspective, given the relevant historical facts, one might propose that the consonants that are now associated with the passive suffixes in Māori actually belong to the final position of the active words. This is one of the proposals Hale (1973) examines, as detailed in §1.2.1. While it is quite possible that this is the analysis an adult speaker of Māori for whom the word-final consonant deletion rule is active entertains, it is difficult to imagine that this is how the child synchronically analyzes the Māori passives in the course of language acquisition since they have no access to historical information.

An alternative to the abstract phonological analysis above is one that brings the surface canonical pattern[4] of Māori words and their analyses during language acquisition closer to each other: for the child learner, the once-word-final consonants are now reanalyzed as morpheme-initial for the passive suffix. While this is in line with the linguistic input the child is exposed to, it confronts the child with the problem of dealing with a serious case of suppletion that affects a high number of words with numerous alternants of the same suffix. We have already mentioned that one of these many passive suffixes, namely /-tia/, is treated as the default suffix. How, then, does the child learn which words to use the other suffixes with? Do they simply memorize such passive forms as individual idiosyncratic lexical items? Or are there associations that the child is able to establish for one or more of the non-default suffixes that enable them to memorize only the active forms whose passive forms they derive through the application of corresponding morpholexical rules (cf. Yang, 2005, p. 268)? While it is clear that the language-acquiring child triumphs in learning the Māori passives whatever the exact analysis they use is, it still presents the linguist with a puzzle that is yet to be solved.

It has long been assumed in the literature that the passive suffixes in Māori are unpredictable in the sense that given the active form of a word in Māori, it is not possible to predict the passive form of that word (Hale, 1973; Blevins, 1994, among others). More recently, however, the unpredictability of the passive endings has been questioned by Moorfield (1988, p. 66) and Harlow (2007, p. 117) who observe that /-ina/ only occurs after words ending with /a/ and /-mia/ only occurs after words ending with the round vowels /o/ and /u/. Inspired by the possibility of predicting passive endings, Parker Jones (2008) uses a neural network model to explore this issue concluding that the selection of the passive suffix may in fact be predictable based on what he calls the *phonotactic* properties of the active verb. In his conclusion, Parker Jones (2008, p. 47)

[4]See Hale (1973) for a discussion of this notion.

notes that: "The work described here is clearly preliminary with respect to the problem of predicting passives and gerunds... But the experimental results are suggestive, especially as they challenge the long held assumption that thematic consonants cannot be predicted."

The motivation for the present study stems from Parker Jones's findings and propositions (2008). It attempts to answer the following primary questions: first, what does a diachronic investigation into thematic consonants in Māori verb bases carrying a passive suffix reveal about the development and evolution of passive verb forms within the language, especially in cases of innovation; and second, what, if any, are the distinct characteristics or subregularities present in Māori verb roots that may contribute to the prediction of their respective passive forms? The central objective is to investigate the potential predictability of the passive forms of Māori verbs mainly based on the phonological properties of the verb roots.[5] To achieve this goal, two comprehensive approaches have been undertaken: The initial approach involves conducting a historical examination of the verb roots to discern the origins of thematic consonants observed in Māori passive verb forms (cf. Arms, 1973; Lichtenberk, 1978). This investigation aims to determine whether such consonants result from direct inheritance from ancestral languages, or if they represent innovations that either emerged locally within Māori or took place at a stage earlier than Māori itself and are reflected in the language in ways that may not be easy to analyze given the lack of clear Polynesian or Oceanic etymons. The importance of finding out the origins of these consonants lies in the implication that systematic innovations could be indicative of predictability behind the consonants. In the second approach, a computer program is developed to conduct a traditional linguistic feature-based analysis on the data based on 16 different categories (e.g., thematic consonants, vocalism, syllable count, etc.) in order to investigate whether there are any associations that arise for any class of words being inflected with the same passive suffix. What one can learn from this type of analysis is whether the formal surface properties of Māori base forms have any correlations with the suffixes used in the corresponding passive forms of the

[5] Arms (1973) claims a strong case of correlation between thematic consonants and meanings in Fijian, and a similar case is proposed by Lichtenberk (1978) for Manam. This paper does not consider this semantic correlation possibility beyond a surface-level examination in cases of innovation presented in §2; however, a more systematic and robust approach will be considered in future work.

words in the language. In other words, if one surface property (e.g., stem-final vowels such as /u/) exclusively co-occurs with a certain passive suffix (e.g., /-mia/), then one can consider that feature as being predictive of the corresponding suffix among other properties, if any.

It is important to emphasize that this study is specifically dedicated to investigating only three passive suffixes, namely /-mia/, /-ria/, and /-hia/, as indicated by the title, despite the presence of a total of eleven suffixes within the data set. For the sake of completeness, all eleven suffixes—the resources used to build the corpus in this study exhibit only eleven passive suffixes—are presented from most common to least: /-a/, /-tia/, /-hia/, /-na/, /-ria/, /ngia/, /-ia/, /-ina/, /-kia/, /-mia/, and /-nga/. The constrained scope of the research is mainly attributed to the temporal limitations within which the study was conducted. The criteria employed for the selection of the examined suffixes were based on a deliberate progression from a smaller class, exemplified by /-mia/, to larger classes, represented by /-ria/ and /-hia/ in this study. This approach was adopted to ensure manageable analysis and to encompass a diverse range of thematic consonants characterized by distinct articulatory features.

As a final point, it is noteworthy to highlight several noteworthy contributions this research presents to the field. First, it represents an advanced and extended version of Parker Jones's pioneering experimental study (2008). Notably, the size of the data set utilized in this investigation is nearly twice the size of Parker Jones's, which provides further opportunities to discover associations or subregularities, if any, among the words that inflect with the same suffix. Additionally, this investigation holds significant value from a historical linguistics perspective. As far as I am aware, it stands as the first systematic and comparative inquiry into Māori thematic consonants, aiming to ascertain their precise origins. Specifically, the investigation seeks to discern whether these consonants are directly inherited within the language or whether they emerged as instances of innovation that have taken place at an earlier time before Māori or during the development of Māori itself. Lastly but importantly, a rigorous examination of sound correspondences between Proto Oceanic, Proto Polynesian, and Māori is undertaken to ensure that the reconstructions available for the Māori verbs within the data set align with the expected sound correspondences.

The paper's organization is structured as follows: §1 serves as the introduction, culminating in the formulation of research questions and the description of the utilized data. §1.2 provides a comprehensive overview of the seminal papers, namely, Hale (1973); Blevins (1994); Parker Jones (2008), underpinning the current study. Moving on to §2, the historical analysis of the /-mia/, /-ria/, and /-hia/ verbs is presented, accompanied by a preliminary overview of historical accounts. The synchronic analysis of the data is reported in §3. Finally, §4 and §5 conclude the paper.

1.1 Data

The data set I used in this study, where I have a total of 889 verbs with their passive suffixes, is based on *English – Māori | Māori – English Dictionary* by Biggs (1990), *The Revised Dictionary of Modern Māori* (1989) and *The Raupō Dictionary of Modern Māori* by Ryan (2012), and online *Te Aka Māori Dictionary* (Moorfield, nd; henceforth *Te Aka*). The primary sources I have consulted for the lexical reconstructions of Māori verb roots are *POLLEX-Online* by Greenhill and Clark (2011), *The Austronesian Comparative Dictionary (ACD)* by Blust and Trussel (2020) and all five volumes of *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society* by Ross et al. (1998, 2007, 2008, 2011, 2016).

In his work, Biggs (1990) presents the passive forms of passivizable verbs in two distinct manners: in cases where the passive is formed by a transparent affixation, the passive suffix is shown as appended to the verb root using a hyphen, exemplified by *awhi-tia* 'beset, besiege; embrace' (1990, p. 88). In other cases when the verb root undergoes alternation in the passive form, the passive version is denoted within parentheses, for instance, *titiro, (tirohia)* 'behold, examine, look, observe, stare' (1990, p. 138). While the data extracted from Ryan (1989) is the one Parker Jones used in his study (2008) and graciously shared with me, I used Ryan's later work (2012) for the purpose of comparison, particularly in instances where formal or semantic disparities between Biggs's data (1990) and Ryan's data (1989) arose. Lastly, *Te Aka Māori Dictionary* (Moorfield, nd) was utilized as a supplementary source to compare and corroborate the data sets of Biggs (1990) and Ryan (1989), with a few additional words contributed to the data set.

The curious reader will quickly notice that in these sources mentioned, it is not uncommon to see more than one passive suffix per verb as in: *ahu-a, -ria* 'tend, foster'; *ahu(ahu)-ngia* 'earth up, heap up' (Biggs, 1990, p. 85); *kuhu(-a-ngia)* 'enter, go into, hide, put on clothes' (Ryan, 2012, p. 112); *hopu (-a,-kia,-kina,-ngia)* 'to seize, catch, snatch, detect, take in the act, capture, arrest' (Moorfield, nd). For the sake of manageability in the diachronic approach and computational ease in the synchronic and computational approaches, it is necessary that each verb be associated with only one suffix in the data set; therefore, it was necessary to make some assumptions while building the corpus. It is a straightforward matter in cases of homophony as in the Biggs example above (1990, p. 85); the homophonous verbs that take different passive suffixes as distinct entries with their corresponding passive suffixes were included as is in the data set. In cases of multiple passive suffixes per verb, the four sources were compared with each other and took one of these steps: if Biggs (1990) provides only one suffix while the others provide multiple for the same verb or in cases of consistency, the verb with the suffix suggested by Biggs (1990) was included.[6] If Biggs also provides multiple suffixes and in cases when the passive form is not given in either Biggs (1990) or Ryan (1989), the one in majority was picked. For example, for the verb *ahu* with the meaning 'tend, foster', Biggs (1990, p. 85) suggests /-a/ and /-ria/; there are no passives available for this verb in Ryan (1989, 2012); *Te Aka* (Moorfield, nd) suggests /-a,-ngia,-ria,-tia/. This verb was included as *ahu-ria* in the corpus based on the majority criterion.

Lastly, it is relevant to mention that the passive forms enclosed in parentheses in Biggs's work (1990) have been excluded from the data set unless they are also listed separately with a clear passive suffix indicated by a hyphen. This decision is attributed to the presence of morpheme-internal alternation within the verb roots. Unlike the other verbs, these verbs manifest distinct phonotactic patterns, rendering them difficult to analyze in a purely phonological fashion in the feature-based analyses. The main reason why I consider them as such under the current approach is due to the fact that there is reduplication involved, a morphological process, which makes the word stems difficult to segment, and consequently, what is going on in such verbs are not ob-

[6]When available, gerund forms of the verbs are also useful in such cases of inconsistency since, if retained, the thematic consonant in the passive suffixes and gerund suffixes are in agreement in Māori (Hale, 1973; Blevins, 1994).

viously relevant to the thematic consonants in question since there is more involved than pure phonology. In support of these difficulties with segmentation observed, it is worth mentioning that Biggs (1990) also does not demarcate a passive suffix in such parenthesized forms, as exemplified by *kakati (katia)* denoting ‘bite, sting; acrid, sour, bitter’ (1990, p. 98) and *kukume (kuumea)* ‘drag, haul, pull, stretch, tow’ (1990, p. 106).

1.2 Background

In this section, I present a concise overview on two seminal investigations delving into the essence of the thematic consonants found in Māori passives. The works in question are those of Hale (1973) and Blevins (1994). Finally, a summary of Parker Jones’s study (2008), which serves as the foundational inspiration for the current research, is provided to introduce his ideas around the predictability of Māori passives and to establish the framework on which this study is built.

1.2.1 An Overview of Hale (1973)

Hale (henceforth, H), in his paper titled *Deep-Surface Canonical Disparities in Relation to Analysis and Change: An Australian Example* (1973), investigates language change via rule simplification and rule generalization looking at some languages from Australia and Polynesia, one of which is Māori. H reports that at some point in the history of Polynesian languages, which might be referred to as “pre-Polynesian”, the grammar acquired a phonological rule which deleted word-final consonants as shown in (1):

(1) $C \rightarrow \varnothing$ / _#

After the addition of this rule, all words that ended in a consonant lost their word-final consonants giving rise to changes such as **inum* > *inu* ‘to drink’ in Māori. Given that the rule only applied to word-final consonants, a stem-final consonant is retained before suffixes. To illustrate, we witness the presence of the original final /m/ of Proto Oceanic **inum* in its passive form in modern-day Māori, /inumia/, and the gerundive form, /inumaŋa/.

The effect of the consonant-deletion rule in (1) is that in modern Māori, all morphemes end in vowels at word boundaries. However, the pre-Polynesian word-final consonants are continued in the suffixed forms of the words. H provides the following representative Māori verb roots, their corresponding passive forms, and glosses presented in (2).

(2)

Verb	Passive	Gloss
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'

H draws attention to the fact that the passive suffix in the last two examples is /-a/ in table **??** while the ending has a consonant associated with it in the other examples. H makes the claim that there are two classes of alternants for the passive suffix in Māori; one with a consonant except for /p/ and /w/ and another without a consonant. H emphasizes the unpredictable nature of the passive endings given the lack of consonants in the bare form of the verbs.

H suggests that the natural reaction of a linguist given such a problem would be to propose a purely phonological explanation and state that the verbs which have consonants in their passive forms underlyingly end in their corresponding consonants as is shown in (3):

(3)

Underlying Forms
/awhit/
/hopuk/
/arum/
/tohuŋ/
/maur/
/weroh/

Under this analysis, the passive ending has only two alternants whose distribution can be described in purely phonological terms by the following rules:

(4) Pass $\rightarrow$ *-ia* / C+_

(5) Pass $\rightarrow$ *-a* / V+_

These rules state that the passive ending (i.e., Pass) is phonologically /-ia/ when following a consonant and /-a/ when following a vowel. The consonant deletion rule given in (1) applies in cases of bare verbs ending in a consonant. The following derivations are provided by H based on the examples and the rules in (4) and (5):

(6) *awhit* + Pass $\rightarrow$ *awhit+ia*

awhit# $\rightarrow$ *awhi#*

hopuk + Pass $\rightarrow$ *hopuk+ia*

hopuk# $\rightarrow$ *hopu#*

patu + Pass $\rightarrow$ *patu+a*

H indicates that this analysis is further supported by the gerundive ending, which also supports the appearance of the postulated stem-final consonants. Under this analysis, the consonant deletion rule in (1) is still a part of the synchronic grammar of Māori, and by making this assumption, H emphasizes the possibility of describing the Māori passive verb forms in a purely phonological way with a small set of completely natural rules, and he names this analysis the "phonological alternative."

Another analysis H proposes as being the null hypothesis is what he refers to as the "conjugation alternative", which suggests the following: the endings are identified with the material to the right of the subsequence which matches the stem in its uninflected form. Therefore, it follows that the passive ending is /-tia/ in /awhitia/, /-kia/ in /hopukia/, and so on. It also follows from this proposal that all verb stems are consistently vowel-final since the ancestral consonant is now associated with the suffix instead of the verb itself. This consonant-to-suffix assignment results in a variety of passive ending alternants which differ from each other in their morpheme-initial consonants. In order to associate the correct type of suffix with the correct verb stem (i.e., /awhi + tia/, not */awhi+ a/, etc.), H draws attention to the necessity of attributing to each stem a lexical diacritic feature,[7] which serves as a conditioning environment for the rules that specify the phonological shapes of the passive suffixes. This lexical solution denies the existence of the consonant deletion rule in (1) in modern Māori since the rule is redundant in the grammar given the apparent absence of root-final consonants. This analysis also suggests that the passives must have undergone a reanalysis if the consonant deletion rule existed at some point in the language.

H considers the phonological alternative to be the more elegant one and expects a linguist to favor it over the lexical analysis, but he takes a critical approach to it and questions the basis on which the phonological alternative would be preferred. He claims that if the basis is simplicity, the two alternatives are not comparable in terms of that since ”one analysis posits abstract phonological markers (final consonants) at the expense of the generalization that all Māori morphemes end in vowels; the other posits abstract diacritic features (conjugations), at the expense of the simple and purely phonological rules which spell suffixal alternants” (p. 416). He suggests that we should examine the preference of the phonological alternative over the other.

1.2.2 An Overview of Blevins (1994)

Blevins (1994; henceforth, JB) presents a reanalysis of the thematic consonants in Māori passives and gerunds as shown in (7)[8] from a perspective that essentially merges Hale's (1973) morphological solution (i.e., the conjugation alternative) with a phonological one.

[7]This refers to unpredictable suppletive allomorphy in which case speakers have to memorize which allomorph goes with which verb.

[8]JB uses f to represent [ɸ].

(7)

Verb	Passive	Gerund	Stem Gloss
mahue	mahuetia	mahuetaŋa	‘leave’
hopu	hopukia	hopukaŋa	‘catch’
inu	inumia	inumaŋa	‘drink’
tohu	tohuŋia	tohuŋa	‘point out’
mau	mauria	mauraŋa	‘carry’
waru	waruhia	waruhaŋa	‘scrape’
fao	faofia	faofaŋa	‘put in’
noho	nohoia	nohoaŋa	‘sit’
ehu	ehua	ehuaŋa	‘bail out’
tahu	tahuna	tahuŋa	‘burn’

JB characterizes the thematic consonants appearing in the passive and gerundive forms of such verbs as in (7) as *a consonant of unpredictable quality* and introduces her analysis characterizing it as a synthesis of a purely lexical and a purely phonological solution. JB states that, according to her approach, the passive and nominal suffixes of Māori, exhibit two distinct lexical forms: one beginning with a vowel and the other with a consonant. Based on an examination of a wide range of constructions from the passive forms of English borrowings into Māori (e.g., *miraka; miraka-tia* ‘milk’ in contrast to *wepu; wepu-a* ‘flog, whip’) to the passives of nominal, adverbial, and adjectival bases, which take /-tia/ in the passive (e.g., *poro; poro-tia* ‘butt, end’ (noun), *noa, noa-tia* ‘without conditions’ (adverb), and *kau; kau-tia* ‘alone, only’ (adverb)), JB concludes as her primary claim that both /-ia/ and /-tia/ have the default status in the passive.

JB’s secondary argument relies on autosegmental phonology, reinforced by the concept of underspecification. JB presents the underlying phonological forms of the two passive allomorphs, shown in (8) and her morphological analysis presented above with phonological representations and rules that are able to account for the passive forms of both vowel- and consonant-final stems.

(8) Phonological form of the two passive allomorphs (Blevins, 1994, p. 44)

a.
```
 i a
 | |
-VV
```

b.
```
  i a
  | |
-CVV
```

As seen in (8b), the thematic consonant *-t-* of the default suffix /-tia/ is not an underlying element of the consonant-initial suffix; instead, there is an empty slot to be filled with a consonant in accordance with the phonotactic structure of each verb root to be passivized. Examples of the consonants to fill the empty slot include /-k/ of *hopukia*, /-m/ of *inumia*, and /-r/ of *mauria* (Blevins, 1994, 45). Within the framework of autosegmental analysis, JB recognizes such thematic consonants as final floating melody elements as parts of the verb stems and proposes an additional allomorphy rule (1994, 45) to ensure correct suffix assignment. JB further explains that the absence of root-final consonants in surface forms of active verb roots is due to their inability to syllabify in accordance with the canonical structure of Māori syllables, denoted as (C)V.

JB's sophisticated autosegmental analysis is also able to account for the surface-level variation observed in the consonant-initial default suffix in three dialects of Māori. In her conclusion, JB highlights the effectiveness and applicability of her proposed morphological and phonological solutions compared to the previously posited approaches.

1.2.3 An Overview of Parker Jones (2008)

In the introductory section, Parker Jones discusses the two principal approaches that have been presented and discussed in the existing literature regarding the Māori passive and gerund suffixes, namely, the morphological and phonological perspectives. While central to his emphasis is his conclusion that both these approaches ultimately assume the lexicalization of thematic consonants in the suffixes in question, the main point Parker Jones challenges is the assumption that the thematic consonant quality is unpredictable in the inflected forms of Māori verb roots, which is the inherent to these two approaches.

In order to test this long-held assumption of unpredictability, Parker Jones builds an artificial neural network model that uses what he calls the phonotactic probabilities of the active verb forms to predict their passive and gerundive counterparts.

The functional equivalent of the network Parker Jones designs for the study is one that takes the representation of the verb root as its input and maps it to a passivized form among the set of outputs of various passive formations. The network has a 3-layer feed-forward architecture with 199 input units, 100 hidden units, and 10 output units. One fully activated bias is used per every unit in the hidden and output layers. In the network, Parker Jones uses a sigmoid activation function, and learning was facilitated through back-propagation and a learning rate of 0.1.

Parker Jones's objective is to construct a model capable of predicting both passive and gerundive forms of Māori verbs. To achieve this, he incorporates relevant information into the input layer to ensure the network's awareness during the learning process. In the input layer, the verb roots are represented through three distinct coding schemes: one segment-based and two syllable-based. In his analysis, Parker Jones considers 10 passive suffix alternants, namely /-hia/, /-kia/, /-mia/, /- ŋia/, /-ɾia/, /-tia/, /-ia/, /-ina/, /-na/, and /-a/. Consequently, the output layer consists of ten output units.[9] Notably, he accommodates situations where multiple passive suffixes can be associated with a single verb root, a departure from the approach adopted in this study. In order to equip the network with appropriate computational power, preventing under- and over-fitting, Parker Jones takes an empirical approach to determine the required number of hidden units based on his data. After achieving satisfactory results, he chooses to employ 100 hidden units in his model. This decision is grounded in his endeavor to optimize the performance and accuracy of the neural network.[10]

The active-passive data Parker Jones uses come from *The Revised Dictionary of Modern Māori* (Ryan, 1989), which is 476 in number minus 12 English loans. Since Ryan's dictionary (1989) included only two gerund patterns, Parker Jones extracted data from Māori Broadcast Corpus, which provided 139 gerund-like types.

His experiments show that the predictive power of both of the syllable-based schemes (the percentages of accuracy are 91.74 and 93.91) is minimally higher than that of the segment-based one (the accuracy percentage is 90.43). Parker Jones concludes that the model does well in predicting the passive forms of Māori verb roots on average challenging the unpredictability assumption.

[9]By way of treating these 10 suffixes as separate outputs, Parker Jones is essentially endorsing the suppletive allomorphy (Hale's conjugation alternative) analysis.

[10]The interested reader is kindly referred to Parker Jones (2008) for further details of the network.

Lastly, as an attempt to illuminate aspects of the neural network, which is often subject to criticism for its opaque nature, Parker Jones investigates a few possible features present in the verb roots that lead to the predictability of the passive forms, and he makes suggestions regarding types of features, e.g., final syllables as opposed to penultimate or antepenultimate syllables, that might enhance the accuracy of predictions. Through this examination, Parker Jones seeks to enhance our understanding of the neural network's underlying processes and identify factors that influence its predictive capabilities.

Based on the successful predictions from segmental and syllabic representations of the active Māori verbs, Parker Jones argues that the lexicalization is not the only solution to the thematic consonant puzzle in Māori, and that the widely-discussed 'phonological' and 'morphological' approaches may not be necessary to frame the Māori passives when one considers them as active-passive verb pairings.

Finally, Parker Jones concludes his study by pointing to the success of his model in terms of predicting the correct passive suffixes and highlighting the experimental outcomes that indicate potential predictive patterns of phonotactic probabilities of active forms, facilitating the mapping of a verb root to its passive counterpart. While acknowledging the preliminary status of the research, he emphasizes the efficacy and versatility of computational approaches in executing extensive statistical analyses over substantial data sets. Furthermore, he underscores the significance of methodological studies similar to this one (2008) in contributing to investigations of allomorphy within Polynesian languages.

2 A Diachronic Analysis of /-mia, -ria, -hia/

As previously indicated, Māori is derived from Proto Polynesian, which, in its ancestral lineage, traces its origins back to Proto Oceanic. A widely recognized characteristic of Polynesian languages, such as Māori, is the manifestation of a thematic consonant within their inflected verbal structures. Conversely, this particular consonantal element remains absent in the non-inflected forms of identical verbs on the surface. These thematic consonants are interpreted as reflexes of the stem-final proto-sounds of Proto Oceanic prior to the sound change that started to delete the consonantal sounds in stem-final positions following the break-up of Proto Oceanic. Nevertheless, one cannot be completely sure that every identified thematic consonant present in modern Polynesian languages originates directly from Proto Oceanic. This uncertainty arises mainly due to language change, which takes place in varied complex ways, and the phenomenon of language contact. Consequently, the potential for certain thematic consonants to represent innovative replacements for proto-sounds cannot be dismissed.

The purpose of this section is to investigate the origins of the thematic consonants *-m-*, *-r-*, and *-h-* that surface in the passive forms of the Māori verbs associated with the /-mia, -ria, -hia/ passive endings in the data set. To better appreciate the importance of such a historical investigation for the current research, it is noteworthy to reiterate the overarching goal of the study: it is to determine whether it is possible to predict which passive suffix—the current study is limited to the analysis of /-mia, -ria, -hia/—is used with what active Māori verb. If a thematic consonant showing itself in the passive form of any verb being inflected with one of the three suffixes under consideration is found to be directly inherited from Proto Oceanic, then the long-held view of unpredictability remains unchallenged. If, however, any instance of these thematic consonants in passivized Māori verbs is determined to be an innovation in the language, then it is reason enough to not be so eager to accept these suffixes as unpredictable. In the case of an innovation, the first question that one should ask is what is the reason behind the innovation? One of the two main possible answers to such a question is that the innovation took place in one or more Oceanic or Polynesian languages at earlier stages than the emergence of modern Māori and that Māori adopted the said innovation for some reason (e.g., language contact). The other possibility is that the innovation took place locally in Māori. While the former does not tell us much about the Māori language itself or its users, the latter case is intriguing and more informative; one needs to ask what might have caused Māori speakers to replace a proto-consonant with another one instead of directly reflecting it from Proto Oceanic, or what might have made the language users insert a consonant into a position where there was none etymologically (i.e., vowel-final Proto Oceanic verb stems). Local innovations could be suggestive of predictability of passive suffixes in Māori under certain circumstances where there is evidence of some kind of systematicity since it would appear to be the case that Māori speakers might have detected or created certain patterns associated with the innovated thematic consonant. This possibility is especially stronger when the same thematic consonant appears as an innovation in multiple distinct verbs. A hypothetical local innovation where all the once-root-final *n was replaced by /m/ would be an example.

The investigation of the origins of the thematic consonants is conducted by examining the earlier versions of the Māori verb roots in question through linguistic reconstruction. Linguistic reconstruction provides information about an earlier stage of a language, for which no direct evidence exists. This process involves employing the comparative method, wherein modern languages belonging to different first order subgroups of a common ancestor are compared to trace the form and semantics of linguistic expressions (Biggs, 1978, pp. 691, 701; Ross et al., 1998, p. 7).

The principal reference for comparing Māori verb roots lies in POc for three main reasons: first, POc serves as the closest ancestor of Māori, with substantial attestation of word-final consonants (Ross et al., 1998, p. 17). Second, the sound changes that have taken place since the break-up of POc till the development of modern-day Māori have been studied extensively. Lastly, a wealth of proto-forms has been reconstructed for POc based on present-day Oceanic languages, rendering POc reconstructions a reliable resource for discerning the origins of once-root-final consonants in Māori.

The following three subsections are designed to provide the reader with essential contextual information regarding the historical background and phonological aspects of Māori. Additionally, these sections aim to provide supplementary details concerning the methodological approaches employed in this paper.[11]

[11] **Key to abbreviations:** PAN: Proto Austronesian; PMP: Proto Malayo-Polynesian; PWMP: Proto Western Malayo-Polynesian; POc: Proto Oceanic; PPn: Proto Polynesian; PPh: Proto Phillippines; NCV: North/Central Vanuatu; PN: Polynesian; MAO: Māori, ACD: The Austronesian Comparative Dictionary (Blust and Trussel, 2020); LPO: Lexicon of Proto Oceanic (Ross et al., 1998, 2007, 2008, 2011, 2016); V: Verb; VT: Transitive verb; N: Noun

2.1 Te Reo Māori and Subgrouping

With approximately thirty other languages, Te Reo Māori, the Māori language, belongs to the clearly demarcated subgroup known as Polynesian, situated under the larger Oceanic branch of the Austronesian language family (Biggs, 1971; Besnier, 1992; Pawley, 2001). The Polynesian languages are indigenous to the geographical region referred to as the Polynesian Triangle, encompassing Hawaii, Easter Island, and New Zealand as its apexes, with parts of Melanesia and Micronesia on the northwest, within the broader expanse of Oceania.[12]

For a comprehensive understanding of this section, it is essential to provide further information about the collateral and lineal relationships among Polynesian languages and the hypothesized protolanguages. This preliminary contextual information will aid in precisely situating Māori within its linguistic context. The necessity of protolanguages is due to the historico-comparative approach taken in this research.

The subgrouping theory I assume in this study aligns with those adopted by Biggs (1978) and Ross et al. (1998). The primary higher-order subgroups of Oceanic, to which Māori belongs, are listed here in chronological sequence. Note that each of them has their corresponding interstage proto versions: Eastern Oceanic, Remote Oceanic, Central Pacific, Tokelau-Fijian, and Polynesian. Within Polynesian, as categorized by Biggs (1978, p. 697), Māori finds its place in the Tahitic lowest order subgroup of the Central Eastern Polynesian subgroup of the Eastern Polynesian subgroup of Nuclear Polynesian, which ultimately, together with its sister, Tongic, forms Polynesian.

2.2 Māori Phonology, Phonotactics, and Orthography

This subsection provides a concise overview of Māori phonology and phonotactics coupled with the orthographical practices adopted in this paper.

Like other Polynesian languages, Māori has a simple phonology with a small size of consonant

[12] See Biggs (1971, pp. 488–89), Ross et al. (1998, p. 8), and Ross et al. (2016, pp. 638–39) for the maps showing these geographic and linguistic areas.

inventory and an average sized vowel and simple phonotactics (Biggs, 1971, 1978; Bauer, 1993; Harlow, 2007). There are 10 consonant phonemes in Māori: /p, t, k, ɸ, h, m, n, ŋ, r, w/. Similar to other Polynesian languages, the number of distinct vowel sounds of Māori is 5: /i, e, o, u, a/ (cf. Bauer, 1993, p. 524). Each vowel is either short or long;[13] the latter is reflected by a macron in the standard orthography of the language as in Māori.[14] Māori phonemes, organized in accordance with conventional articulatory terms, are presented in table 2.[15]

Table 2: The phonemes of Māori.

	labial	**dental/alveolar**	**velar**	**glottal**
stops	p	t	k	
fricatives	ɸ			h
nasals	m	n	ŋ	
liquid		r		
glide	w			
	front	**central**	**back**	
high	i		u	
mid	e		o	
low		a		

Additionally, Māori has a variety of diphthongs, and since most of the data comes from Biggs's dictionary (1990), I adopt those that he provides as diphthongs, given here orthographically (cf. Bauer, 1993, p. 534 and Harlow 2007, p. 69): *ae, āe, ai, āi, ao, āo, au, āu, ou, ōu, ei, ie, eo, eu, ea, ia, oa, ua, oi, oe, iu, io* (Biggs, 1990, p. 7).

Concerning the phonotactics, Māori demonstrates very simple patterns like other Polynesian languages. In Māori, all syllables are open, there can be at most one consonant in syllable onsets, and consequentially, consonant clusters are not attested (Bauer, 1993, p. 542, cf. Harlow 2007, pp. 71–5).

[13] See Harlow (2007, pp. 66–7) for a discussion of different approaches to vowel inventory size. I treat short and long vowels as distinct phonemes in my synchronic approach detailed in §3.

[14] For phonetic details, see Biggs (1961), Biggs (1990), Hohepa (1967), Bauer (1993), and Harlow (1996, 2007).

[15] This is adapted from Harlow (2007, 63). In standard orthography, /ŋ/ is *ng*. Additionally, there is extreme variation in the realization of what I adopt as the /ɸ/ sound, which is the primary reflex of PPn **f* (Biggs, 1971, 1978) and is spelled as *wh* in orthography (cf. Bauer (1993, p. 522); Biggs (1978, p. 707)). Following Biggs (1971) and (1978), I assume that this phoneme is /ɸ/ (i.e., *wh*) in word initial positions before unrounded vowels (e.g., PPn **faka-* > *whaka-* CAUS); it is realized as *h* word-medially and initially before round vowels (e.g., PPn **folo* > *horo* 'swallow'), and *w* before a vowel followed by **s* or **f* (e.g., PPn **fafine* > *wahine* 'woman').

Lastly, I have adopted the standard orthography for the representation of Māori words (Māori Language Commission, 2012) while providing the reconstructed forms in the orthography that their respective resources use.

2.3 Sound Correspondences

Despite the existence of sound reconstructions for POc based on Oceanic languages (Ross et al., 1998, 2007, 2008, 2011, 2016; Lynch et al., 2011, among others) and for PPn based on Polynesian languages (Biggs, 1971, 1978), it was necessary for me to have the sound correspondences between POc, PPn, and Māori presented in a systematic way. The references I have consulted for the POc reconstructions of the relevant Māori verbs do not always include the intermediate PPn reconstructions. Given that regular sound changes have occurred between POc, PPn, and Māori, it was necessary for me to establish how each sound in each position in the POc reconstructions is reflected in PPn (cf. Harlow, 2007).

To this end, I made use of the standard comparative method to examine the cognate reflexes of POc lexical reconstructions across several modern Polynesian languages, mostly belonging to subgroups distinct from that of Māori. Within the Polynesian language family, two primary lower-order subgroups exist, namely, *Tongic* and *Nuclear Polynesian*. Tongan and Niuean are classified as *Tongic*, whereas all other Polynesian languages fall under *Nuclear Polynesian*.

Given the conservative nature of Tongan (Pawley, 2001, p. 203) and its relative genetic distance from Māori, the cognates found in Tongan were regarded as the most dependable sources. Additionally, cognates from Samoan, Rarotongan, and Hawaiian were systematically examined. Other Polynesian languages were consulted in the absence of cognates in one or more of these five languages including Māori.

Once an adequate set of cognates is collected, demonstrating how POc reconstructions are reflected in modern Polynesian languages, one can use the sound correspondences Biggs (1971, 1978) provides between PPn and Polynesian languages to determine the PPn reflexes of POc sounds. The POc, PPn, and Māori sound correspondences are provided in table 3:[16]

Table 3: Sound correspondences between POc, PPn, and Māori.

POc	*p, *p^w	*b, *b^w	*t, *d	*r, *dr	*l	*s	*s	*s	*j, *c	*k, *g	*m	*m^w, ŋ	*n, *ñ	*w	*y	*q	*R
PPn	*f	*p	*t	*l, *r	*l	*s	*h	*t		*k	*m	*ŋ	*n	w	∅	*ʔ	∅
MAO	ɸ	p	t	r		h	∅	t		k	m	ŋ	n	w	∅	∅	∅

2.3.1 A Māori Passive Suffix: /-mia/

The verbs that take the /-mia/ suffix in the passive construction constitute the third smallest class in the data set; there are only 6 distinct verb roots and 11 total verbs in this class. Of these 6 verbs, 3, shown in (9), have firm POc reconstructions that end in /m/;[17] therefore, I assume that the thematic consonant is inherited in these verbs:[18]

(9)

Māori Base	Passive	Gloss	POc	Gloss
horo	horomia	'gulp'	*polom	'swallow'
inu	inumia	'drink'	*inum	'drink'
tanu	tanumia	'bury, plant'	*tanum	'bury, plant'

2 of the remaining 6 verb roots have unknown etymologies, which prevents me from being able to determine the origin of the thematic consonant; that is, whether the final *m* was inherited from an earlier form of each verb, or whether it is an innovation in or before Māori in each case.

(10)

Māori Base	Passive	Gloss
aru	arumia	'chase'
nao	naomia	'handle'

The last verb root in this class is the one that I propose to be clear case of innovation in the language: *ngaro; ngaro-mia* 'out of sight' reflects the POc *$^*m^w aloq$* 'submerged rock'.

[16] The correspondence between POc **s, *j* > PPn **t* > MAO *t* has been observed in some words such as: PAN **ŋusuq* > POc **ŋusu(q)* > PPn **ŋutu* > MAO *nutu* 'lip; beak' (ACD); POc **Rusan* > PPn **uta* 'cargo, freight (of canoe)' (Blust, 1976, p. 347); POc **jila* > PPn **tila* > MAO *tira* 'mast of a canoe; yardarm' (ACD). See Blust (1976) for a discussion of this curious sound change where he describes PPn **s*, **h*, and **t* as reflexes of the PAN palatal stops, an inexplicable phenomenon under the theory of PAN Blust assumes. Also, see: POc **la(s,c)am* > MAO *rata* 'tame, quiet; familiar, friendly' (ACD).

[17] See Ross et al. (2016, pp. 257–58) for a discussion of final /-m/ vs. /-n/ in POc **polom.*

[18] The meanings given for both Māori verbs and reconstructed forms are condensed or abbreviated in consideration of readability and spatial constraints.

Based on the discussion of complex lexical structure of the terms for forgetting, Ross et al. (2016, pp. 557–59) demonstrate the semantic connection between the POc *m^waloq* and *ngaro* in Māori with an intermediate stage where *m^waloq* is reflected as *ŋalo* 'out of sight, disappeared, forgotten, lost' in PPn. Assuming that the stem-final /q/ Ross et al. (2007, p. 113) and Ross et al. (2016, p. 557) reconstructs for *m^waloq*, which got lost after the break-up of POc, is accurate, and given the widely accepted proposal that there was no /q/ phoneme in PPn and that the glottal stop, *ʔ*, reconstructed for PPn corresponds to zero in Māori (Biggs, 1965, 1971, 1978), the thematic consonant used in the passive form of the Māori reflex, *ngaro*, must be an innovation in the language. Now, the question to ask is why Māori speakers chose /-mia/ as the passive ending for *ngaro* and its causative form, *whakangaro*, over any of the other passive endings including the ones with and without the morpheme-initial thematic consonants.

One way to approach the question raised above is to analyze the cognates of the Māori *ngaro* in any other modern Oceanic languages to see what thematic consonant they might have. Importantly, finding a cognate with the same thematic consonant, i.e., *-m-*, in any other Oceanic language would suggest that the innovation must pre-date Māori. Ross et al. (2007, p. 113) provides cognates from 15 Western and Eastern Oceanic languages, and cognates from 5 other Polynesian languages come from Ross et al. (2016, pp. 558–59). However, none of those 20 cognates show any thematic consonant; all but one end in /o/ while the one outlier has a final /e/. Given the lack of evidence for the /m/ generalization to be older than Māori itself, I assume that /m/ is a within-language innovation, and to examine this further, I now focus on another approach.

Harlow (2007, pp. 116–17) discusses two processes which he describes as having the tendency to bring regularity to the passive formations in Māori. He focuses on the passive suffixes which have earned the "default" status in Māori and other closely related Polynesian languages. However, it is mainly the second process that is of relevance here. Citing Moorfield (1988, p. 66), Harlow (2007) proposes that certain subregularities have developed in Māori, one of which is about /-mia/; he reports that /-mia/ is used only after /-o/ and /-u/ (e.g., *whakangaro-mia* 'be made to disappear', *inu-mia* 'be drunk'). The data at hand is in alignment with Harlow's observation about /-mia/ since all the 11 /-mia/ verbs in the data set end in either /o/ or /u/. In support of this approach, considering the possibility that the final V or CV sequence of the verb root may indeed be the determining factor coupled with the possibility that a phenomenon similar to Arms's (1973) observation in Fijian transitive—where he proposes semantic associations with the thematic consonants—might have also played a role in Māori, I now turn to phonological and semantic neighborhoods of *ngaro-mia* to examine whether the Māori *m* innovation in place of POc **q* might be a product of lexical analogy. In order to find possible neighbors, the primary resources used in this study were consulted. First, a phonological approach were undertaken to search for Māori words or reconstructed forms showing the thematic consonant in question, and then, those which are deemed to be semantically close enough were examined further.[19]

4 words that I could find, which seem to fall within the phonological and semantic neighborhoods criterion, are given below:

(11) Phonological and Semantic Neighbors of Māori *ngaro*

POc	**ralom*	'below; under'	ACD
POc	**rodrom*	'dark; be night'	(ACD; Ross et al. (2007, p. 308))
POc	**marom*	'be dark'	(Ross et al., 2007, 304)
Māori	*toromi*	'drowned'	(Biggs, 1990, p. 129)

POc **ralom* 'below; under; deep' is reflected as *raro* 'the bottom; down; beneath' in Māori (ACD) with the expected **l* > *r* change. Despite three Polynesian reflexes of the first, there is not any Māori data available for either POc **rodrom* or **marom*. Lastly, Biggs (1990, p. 139) gives *toremi* with the meaning of 'disappear, disappeared, drowned' and *toromi* 'drowned, be'[20] (1990, p. 139). *Toromi* resembles one POc reconstruction that Lynch (2001) and Ross et al. (2016, pp. 257–65) talk about at length: **tolom* 'swallow'. What is interesting about this reconstruction is that, first, **tolom* is one of the two main hypothetical POc forms for 'swallow' (the other one is **polom*), and second, even though Ross et al. (2016, p. 258) do not agree with this, Lynch (2001)

[19] It is important to point out the informal nature of this approach. A systematic search for phonological neighbors is employed in §3.1.13.

[20] It is possible that the definition was intended to be "be drowned" instead.

suggests a clear connection between the earlier forms, (i.e., from PMP and PWMP in this case), meaning 'sink' and forms meaning 'swallow'. A similar suggestion for reconstructed forms from higher level proto languages with the same meanings was also made by Mills (1975, p. 854). I am unable to find a POc form for Māori *toromi*; however, if there is any such connection as Lynch (2001) suggests, then it may be possible that *toromi* and POc **tolom* share a connection. It is noteworthy to point out that in Māori, *horo* 'gulp', which is a /-mia/ verb, reflects POc **polom* 'swallow' instead of **tolom*. Regardless, given the identical root-final CV sequences and possible semantic similarities, the assumed local /-m/ innovation for *ngaro* might have been due to such words in the language and its ancestors. However, more stable data and a more involved analysis are due to support the claim that the innovation of /m/ in the passive form of *ngaro* is based on such a pattern that the Māori speakers detected rather than pure chance.

2.3.2 A Māori Passive Suffix: /-ria/

In the subsequent analysis of verb classes within the data set, I direct my attention towards those verbs exhibiting the passive marker /-ria/. As observed previously with the /-mia/ verbs, there are several words derived from the same verb roots in this category, which include reduplicated and causative forms. Consequently, while there are 37 verbs overall in this class, the actual count of distinct verb roots is 26. Among these 26 roots, 18 demonstrate reliable POc reconstructions, each of which ends in one of eight distinct POc sounds, as illustrated in (12):

(12)

Māori Base	Passive	Gloss	POc	Gloss
kākahu[21]	kahuria	‘clothe’	*kaput	‘wrap’
kau	kauria	‘swim’	*kaRu	‘swim’
mātau	mātauria	‘adept’	*ma-taqu	‘know’
takapau	takapauria	‘spread out’	*tabakau	‘mat’
takoto	takotoria	‘lie down’	*keteŋ	‘straighten’
tau	tauria	‘count’	*sau	‘tattoo’
tīmata	tīmataria	‘begin’	*mata	‘source’
tohu	tohuria	‘sign’	*tusuq	‘point at’
tupu	tupuria	‘grow’	*tubuq	‘grow’
tū	tūria	‘stand’	*tuqur	‘stand’
whakatipu[22]	whakatipuria	‘bring up’	*tibu	‘ancestor’
whāwhā	whāwhāria	‘feel’	*paRa	‘hand, arm’
whakaako	whakaakoria	‘educate’	*akop	‘learn’
whakaatu	whakaaturia	‘point out’	*atu[23]	‘away’
whakakai	whakakairia	‘pendant’	*kawil	‘hook’
whakamataku	whakamatakuria	‘fright’	*matakut	‘be afraid’
whakatapu	whakatapuria	‘sanctify’	*tabu	‘taboo’
whatu	whaturia	‘weave’	*patuR	‘weave’

As for the remaining 8 verb stems, even though there are no intermediate POc reconstructions available, applying the comparative method yielded some comparable Proto Western Malayo-Polynesian (PWMP) forms for two, as shown in (13):

(13)

Māori Base	Passive	Gloss	PWMP	Gloss
apiti	apitiria	‘add to’	*tampir; tampil	‘attach’
mākutu	mākuturia	‘bewitch’	*kutuk	‘oath; curse’

Lastly, I am unable to provide earlier forms for the remaining 6 verb roots since there are no reconstructed forms available at the time of writing this paper, and I was unable to reach any

[22]Even though *kākahu* appears to be a reduplicated form of *kahu* ‘cloak, garment’, Ross et al. (1998, pp. 154–55) provides *kākahu* ‘clothing’ as the Māori reflex of POc **kapu(t), *kaput-i-* ‘wrap, cover; cover food prior to cooking’ and of PPn **kafu* (N) ‘clothing or covering for the body’; (V) ‘cover the body.’

[23]Biggs (1990, 138) notes that *tipu* is the Eastern dialect form of *tupu*; however, considering it reflects POc *tibu* ‘ancestor, grandparent’ and the meaning nuance between *tipu* and *tupu*, *whakatipu* is treated separately than *tupu.*

[24]This reconstruction is given as PMP in ACD.

helpful data using the comparative method, either. The following are the words of unknown etymologies:

(14)

Māori Base	Passive	Gloss
ahu	ahuria	‘tend, foster’
apu	apuria	‘heap’
mau	mauria	‘carry’
whakamātaotao	whakamātaotaoria	‘to cool’
whakatūpato	whakatūpatoria	‘warn’
whītiki	whītikiria	‘belt’

Now, I turn to the verb roots with POc reconstructions, given in (12), to dissect them further. 2 of the 18 POc reconstructed forms provided in (12) end in **r* and **l*; these words are repeated below in (15) for ease of reading:

(15)

Māori Base	Passive	Gloss	POc	Gloss
tū	tūria	‘stand’	*tuqur	‘stand’
whakakai	whakakairia	‘pendant’	*kawil	‘hook’

As shown in §2.3 where I lay out the sound correspondences between POc, PPn, and Māori, the proto sounds **r* and **l* merged as /r/ in Māori. The thematic *-r-* that appears in the passive form of these verbs in (15) is, then, the expected, inherited consonant.

9 of the remaining 16 distinct POc reconstructions end in a vocalic sound; that is, there is no final consonant in their proto-forms as repeated in (16):

(16)

Māori Base	Passive	Gloss	POc	Gloss
kau	kauria	‘swim’	*kaRu	‘swim’
mātau	mātauria	‘adept’	*ma-taqu	‘know’
takapau	takapauria	‘spread out’	*tabakau	‘mat’
tau	tauria	‘count’	*sau	‘tattoo’
tīmata	tīmataria	‘begin’	*mata	‘source’
whakatipu	whakatipuria	‘bring up’	*tibu	‘ancestor’
whāwhā	whāwhāria	‘feel’	*paRa	‘hand, arm’
whakaatu	whakaaturia	‘point out’	*atu	‘away’
whakatapu	whakatapuria	‘sanctify’	*tabu	‘taboo’

The stem-final sounds of the POc reconstructions in (16) are reducible to three vocalic phonemes: /u/ for four out of eight proto-forms, /a/ for two, and /au/ for the last two. The pattern that we observe in (16) is that there is perfect correspondence—except for the length difference in *whāwhā*—between the stem-final vowels in proto-forms in POc and their reflexes in Māori and that the majority of the stem-final sounds is [+back, +round].

Following one of the proposals made in §2.3.1, namely that it is possible for the final V (or CV) segments to have some influence on the thematic consonants in Māori, it is noteworthy to point out that final /u/ or /au/ constitute the majority; however, this is a tentative suggestion considering the low number of data at hand and the small vowel system.

Finally, the remaining 7 verb stems out of the 18 Māori forms whose history can be traced to POc present us with the most interesting case in the /-ria/ class; the respective POc reconstructions of these 7 verb stems have one of the following sounds as final consonant: **t, *ŋ, *p, *q,* and **R* as presented in (17):

(17)

Māori Base	Passive	Gloss	POc	Gloss
kākahu	kahuria	'clothe'	*kaput	'wrap'
takoto	takotoria	'lie down'	*keteŋ	'straighten'
tohu	tohuria	'sign'	*tusuq	'point at'
tupu	tupuria	'grow'	*tubuq	'grow'
whakaako	whakaakoria	'educate'	*akop	'learn'
whakamataku	whakamatakuria	'fright'	*matakut	'be afraid'
whatu	whaturia	'weave'	*patuR	'weave'

A review of the relevant sound correspondences between Proto Oceanic, Proto Polynesian, and Māori is in order:

- There is regular correspondence between the POc **t*, PPn **t*, and Māori /t/.
- PPn and Māori inherited POc **ŋ* as is.
- POc **p* displays a less regular patterning where it corresponds to **f* in PPn, which is reflected as /ɸ/, /w/, or /h/ in Māori (Biggs, 1978).

- POc **q* corresponds to the glottal stop **ʔ* in PPn, which Māori reflects as zero.
- POc **R* is reflected as zero in PPn and Māori.

It is clear that none of these POc final consonants showing up in the proto-forms in (17) exhibit correspondence with the Māori /r/ sound. Nonetheless, the Māori reflexes of these POc reconstructed forms in (17) inflect with the /-ria/ suffix in their passive forms. It follows from this observation that the thematic consonant *-r-* in these words must be an innovation. Once again, the question is what might be the reason why Māori speakers decided to use /-ria/ to passivize these verbs over any other passive ending. Is there a regular pattern that led Māori speakers to this path?

The first approach to take, as I did with the /-mia/ verbs, is to look for cognates in other modern Oceanic languages to see whether any of the innovations are local, i.e., happened in Māori, or whether any took place at a time that is pre-Māori.

POc **kaput* 'wrap, cover; cover food prior to cooking'

In light of the sound correspondences established in §2.3, the expected stem-final consonant for **-t* is †*-t*[24] in Māori. Ross et al. (1998, pp. 154–55) provides 18 reflexes of **kaput*, including Māori *kākahu* 'clothe'; however, even though there are remnants of the final consonant in some reflexes as in *gavus* 'wrap, cover, cook on embers in wrapper' in Raga, a North/Central Vanuatu language, and *kaputi* 'overspread, cover the whole surface' in Niuean, another Polynesian language, none of them show /-r/. Thus, I assume that the *-r* innovation in question here is a local one in Māori.

PMP/POc **keteŋ* 'straighten out, of a limb, etc.'

Considering the sound correspondences, the expected stem-final consonant for **-ŋ* is †*-ŋ* in Māori. For **keteŋ*, ACD provides cognates from five Oceanic languages, none of which has a final consonant. Thus, the assumption I make for *-r* in this word is that this is also a local innovation in Māori.

[24]Following the convention used by Ross et al. (2016), I use the dagger symbol, *†*, to mark the expected forms.

POc **tusuq* 'point at'

In light of the sound correspondences, **-q* corresponds to zero in Māori. For **tusuq*, ACD shares 11 reflexes from 8 languages, one of which is Tongan, a Polynesian language. (Ross et al., 2016, p. 178) lists 18 cognates from 13 Oceanic languages, 3 of which are Polynesian. In both sources, only Tongan shows the reflex of the final consonant **q*: *tusuʔ-i* 'point' (VT), and all the rest end in vowels, except for one. Interestingly, Nakanai, a Meso-Melanesian language, shows an /r/ in its reflex: *tur-i* 'point to, point out' (Ross et al., 2016, p. 178). In order to figure out how POc **q* is reflected in Nakanai, I investigated the Nakanai reflexes of numerous POc reconstructions provided in Ross et al. (1998, 2007, 2008, 2016) and conclude that POc **q* corresponds to /h/ in most cases as in **saqi(t)* 'sew' > *sahi* 'sew' and **qapu* 'ashes, dust' > *havu* 'lime (made from burnt shells)' (1998, p. 147) and sometimes corresponds to zero as in **loqi* 'make thread by rolling fibres on the thigh' > *loi* 'rub between the hands' (1998, p. 288) and **taqe-* 'faeces' > *ta-tae* 'excrement' (2016, p. 202). Ross et al. also mentions in footnote 16 that POc **q* is reflected as *h* in Nakanai (2007, p. 211). Furthermore, Ross et al. suggests that Nakanai /r/ reflects POc **s* and **c* (2008, 400), and two examples where we witness this pattern are **qaco* 'sun, daytime' > *haro* 'sun; day' (2007, p. 179) and **qasam* 'fern used for tying and binding, Lygodium circinnatum' > *hara* 'Lygodium circinnatum' (2008, p. 233). Lastly, there are no other Nakanai reflexes of transitive POc proto verbs whose final consonant is retained in these four volumes of Lexicon of Proto Oceanic (Ross et al., 1998, 2007, 2008, 2016) that I consulted. Clearly, then, the final *r* in *tur-i* 'point to, point out' reflects **...s* in **tusuq* instead of the final **q*. Thus, I conclude that the *-r* innovation in this word is also local.

POc **tubuq* 'grow, swell'

ACD (Blust and Trussel, 2020) lists data from 20 languages, in addition to Māori, and Ross et al. (1998, pp. 134–35) lists data from 13 languages with reflexes of **tubuq*; however, there is no stem-final consonant in any. Therefore, I conclude the *-r* innovation in question we also observe in **tubuq* is local like **tusuq*. Given the similarity between these two proto verbs, the innovation in either might have affected the other provided one predates the other.

POc **akop* 'learn'

The sound correspondences established indicate that *-*p* is expected to be reflected as †-*wh* in Māori. Besides Māori, reflexes from 8 languages are provided in Ross et al. (2016, p. 565), and none of them appears with a stem-final consonant. Thus, I assume the -*r* innovation examined here is yet another Māori-based innovation.

POc **matakut* 'be afraid'

This proto verb is widely reflected in all major Oceanic subgroups. ACD lists reflexes from 49 Oceanic languages, besides Māori, and Ross et al. (2016, p. 584-85) provides reflexes from 30 languages. Many of them retained the root final **t* in the form of expected reflexes. Kove, a North New Guinea language, reflects **matakut* in its transitive form, **matakut-i-* 'to fear (s.t.)' as *mataur-i* 'afraid'; however, it is noted that *r* reflects **R* (Ross et al., 2016, p. 584). Ross et al. also reports that compared to the intransitive form, the transitive form has fewer reflexes and highlight the possibility of parallel innovations having taken place across various Oceanic languages (2016, 584). In the light of these two pieces of information, it seems to be the case that -*r* in Kove is an innovation. Unlike Kove, Māori reflects the intransitive form of the verb; however, despite the genetic distance between the two languages, it is possible that the innovation in Māori might have a Kove influence. Alternatively, as in the case of previous verbs discussed above, this is an independent, local innovation in Māori.

POc **patuR* 'to plait, weave'

The sound correspondences point to an expected zero for *R. ACD provides information from 6 Oceanic languages and Ross et al. (1998, pp. 81–2) lists 14 languages for **patuR*. Almost most of the languages do not reflect the final consonant, 3 Southeast Solomonic languages in Ross et al. reflects *-*R* > *l* in Tolo and Lau; *-*R* > *r* in Arosi. In fact, these languages seem to reflect **R* in all positions in a word as illustrated by many other reflexes (Ross et al., 1998). As mentioned in §2.3 and repeated above, POc **-R* corresponds to zero in Māori, and **r* and **l* merged as /r/ in the language. Considering that both Southeast Solomonic languages and Māori belong to the Eastern Oceanic branch, it is likely that the innovation in Māori might have been influenced by the fact that these languages reflected the transitive form, **patuR-i-* 'tie, plait, weave (mats, baskets +)', of the verb retaining the final **R* either as *-l* or *-r*. Clearly, though, these languages do not present an innovative case, which leads me to assume that the *-r* replacement in Māori is local.

Now that we have established all the replacements as hypothetical local innovations in Māori, the next step is to examine whether they might be cases of lexical analogies within phonological and semantic neighborhoods.

POc **kaput* 'wrap, cover; cover food prior to cooking' > *kākahu* 'cloak; clothe, dress'

Ross et al. suggests that there are two POc reconstructions available for wrapping food for cooking: **kapu(t)*, **kaput-i* and **kopu* noting that both of which refer specifically to clothing or blankets in Polynesian languages (1998, p. 154). Despite the relatively larger semantic space, none of the words bearing lexical resemblance show final **ur* or *ur* to account for the current *-r* innovation.

PMP/POc **keteŋ* 'straighten out, of a limb, etc.' > *takoto* 'lie down'

Ross et al. (2016, 378) point out that although **qenop* is reconstructed for 'lie, rest horizontally', it was replaced by other terms in many subgroups. In the case of Māori *takoto*, it is a reflex of PCP **koto* 'lie down' and PPn **ta-koto*[25] 'lie down'. Since there is no final consonant in these PCP and PPn etymons, it is clear that the choice of /-r/ is innovative as the word itself. As for lexically and phonologically similar words that might have led to the /r/ thematic consonant, one possible candidate is the Fijian *davo*, *davaro* 'to lie down, lie on' (Gatty, 2009, p. 62), which carries a final /r/; however, my search in four other Polynesian languages – Samoan, Tahitian,

[25]It is also suggested that PPn **ta-* seems to reflect POc **ta-*, a spontaneity marker (Ross et al., 2016, p. 378).

Tongan, and Mangarevan—in addition to Māori itself, did not yield any helpful words with a final /r/ within the phonologically and semantically proximal space (Tregear, 1891; Pratt and Newell, 1891). Among the words I have looked up are 'horizontal', 'nap', 'recline', 'relax', 'rest', and 'stretch (out)'.

POc ****tusuq* 'to point at' > ***tohu*** 'sign; signal'**

Another /-ria/ verb, *whakaatu* 'point out, show, draw attention to', seems to bear some phonological and lexical resemblance to *tohu*; however, despite phonological neighbors within the same class such as *ahu* 'tend, foster' and *kākahu* 'clothe, dress', no other verb seems to be both phonologically and semantically close within POc and Māori.

POc ****tubuq* 'grow, swell' > ***tupu*** 'grow, sprout'**

The closest match within the same class for *tupu* is *whakatipu* 'grow, bring up, educate, nurture' from POc **tibu* 'ancestor, grandparent'; however, the lack of a final consonant does not help with explaining *-r* innovation. One word of interest is *huri* 'seed; young shoot, sprout' < POc **suli* 'plant shoot, sprout, cutting'. A metaphoric extension of such terms to kinship terms, especially grandchildren, was noted by Fox (1971). Therefore, it may be possible that *huri* might have given a reason to select the thematic consonant *-r-* for *tupu*. Lastly, as previously alluded, it remains plausible that the innovation observed in **tusuq* could have influenced the occurrence of the one in **tubuq*, or vice versa, depending on which one transpired earlier in the historical development of Māori, given their phonological similarity.

POc ****akop* 'learn' > ***whakaako*** 'educate'**

A comparison suggested by Ross et al. (2016, 565) provides a lexically relevant reconstruction with the meaning of 'to teach': POc **paka-usuri*. Despite the phonological distance, a lexical analogy might have been drawn and the *-r-* in *-usuri* might have been adopted for *whakaako-ria*.

POc **matakut* 'be afraid' > *whakamataku* 'fright, frighten, frightened, scare, terrify'

As mentioned above, **matakut* is a widely reflected form across Oceanic languages, and only Kove has the same *-r* innovation. Considering that I am unable to find any other POc reconstructions and Māori forms to lead to the innovation based on lexical analogy in *whakamataku-ria*, this one may indeed be a Kove effect.

POc **patuR* 'to plait, weave' > *whatu* 'weave'

Despite the lack of reflexes from any Central Pacific languages and the unexpected vowel correspondence, POc **piji(r), *pijir-i-* 'braid, twist together' proposed by Ross et al. (1998, pp. 288–89) shows close lexical resemblance to **patuR*.

Thanks to the higher number of verb roots in the /-ria/ class, it is relatively easier to look for patterns as opposed to the previously examined /-mia/ class to see whether there is any kind of phonological regularity that might lie at the core of the hypothetical /-ria/ pattern. One preliminary observation about the verb roots that conjugate with /-ria/ in the passive is that majority of them (i.e., 16 out of 26 distinct verb roots) end in /u/, and one ends in /ū/ as is seen in the following examples:

(18)

Māori Base	Passive	Gloss	POc	Gloss
apu	apuria	'heap'		
mātau	mātauria	'adept'	*ma-taqu	'know'
tū	tūria	'stand'	*tuqur	'stand'

The similarity in terms of vowel sequences also stands out; a final pattern of /*...a(C)u(C)(a)(u)*/[26] seem to dominate slightly more than half of the verb stems in the data set (i.e., 14 out of 26 distinct verb stems) as illustrated with a couple of examples in (19):

[26]The symbol C is used to refer to any consonantal segment of Māori, and the parenthesis notation indicates optionality.

(19)

Māori Base	Passive	Gloss	POc	Gloss
kākahu	kahuria	‘clothe’	*kaput	‘wrap’
mākutu	mākuturia	‘bewitch’	*kutuk	‘oath; curse’
takapau	takapauria	‘spread out’	*tabakau	‘mat’

It is important to point out neither of these final *u/ū* and /...*a(C)u(C)(a)(u)*/[27] patterns holds for every single verb stem in this class as illustrated by *tīmata; tīmataria* ‘begin’ < POc **mata* ‘source.’

To conclude, there is no one single phonological pattern that shows up in all or a majority of the verbs in the /-ria/ class that may be considered as the underlying pattern why these verbs all take /-ria/ and why innovations might have taken place. Further synchronic analysis on the forms of the verbs is provided in §3.

2.3.3 A Māori Passive Suffix: /-hia/

The final category of verbs examined in this research comprises verb roots that undergo conjugation with /-hia/ in the passive construction. Nonetheless, prior to delving into the specifics of the diachronic analysis, an investigation into the phonological aspect of Māori /h/ is warranted.

The investigation into the phoneme correspondences between POc, PPn, and Māori, as laid out in §2.3, has yielded two notable observations pertaining to the phoneme /h/ in Māori, which were similarly reported by Biggs (1971, 1978) and Blust (1976) among others: Māori /h/ has two distinct origins.

Table 4: POc and PPn sound correspondences for Māori /h/.

POc	*p	*p^w	*s	
PPn	*f		*h	*s
MAO	ɸ		∅	h

Since, as detailed in §2.3, there usually are not many PPn reconstructions even though there is a great number of POc reconstructions and their reflexes in Polynesian languages, I used the

[27] This pattern is to be interpreted beginning from the final position of the verb stem with an optional stem-final /u/, preceded by an optional /a/, preceded by an optional consonant, preceded by /u/, preceded by an optional consonant, which is in turn preceded by /a/.

POc reconstructions and their reflexes in a few Polynesian languages that are (i) close to Māori and (ii) are relatively conservative in terms of sound changes they have gone through in order to reconstruct the expected sounds in PPn to confirm this finding. Some examples are provided below.

Table 5: Expected PPn reflexes of POc **p* in all positions based on cognates from Polynesian languages.

POc Sound	**POc Etymon**	**MAO**	**TON**	**SAM**	**RAR**	**HAW**	**PPn**
**p-*	*patu 'stone'	whatu	fatu	atu	atu	haku	**f-*
**-p-*	*api 'fire'	ahi	afi	afi	aʔi	ahi	**-f-*
**-p*	*qatop 'thatch'	ato; atoh-ia	ʔato	ato	ato	ako	**-f*

Table 6: Expected PPn reflexes of POc **s* in all positions based on cognates from Polynesian languages.

POc Sound	**POc Etymon**	**MAO**	**TON**	**SAM**	**RAR**	**HAW**	**PPn**
**-s*	*saman 'float'	ama	hama	ama	ama	ama	**h-*
	*siko 'tail'	hiku	hiku	siʔu	iku	hiʔu	**s-*
**-s-*	pose 'canoe paddle'	hoe	fohe	foe	ʔoe	hoe	**-h-*
	*lasor 'testicles'	raho	laho	laso	raʔo	laho	**-s-*
**-s*	*taŋis-ia 'cry'	tangi-hia	*teŋi-hia	tāŋi-si-a	–	kani-hia	**-s*
	*lakas 'to stride'	†raka-hia	laka	laʔa; laka-sia	–	–	

Even though there are not many POc reconstructions with *pw, and the ones that are accessible do not have a large number of reflexes available, I did not include them in the table; however, the reflexes provided for them still support the hypothesized POc **p*^{w}, **p* > PPn **f* reconstruction.[28] As a reminder, it is important to highlight that even though PPn **f* corresponds to /ɸ/, orthographically *wh*, in Māori, there are some exceptional environments: PPn **f* is reflected as /w/ in initial positions preceding a vowel that is followed by **s* or **f* and /h/ medially and before round vowels (Biggs, 1971, 1978), which explains the discrepancy between Māori *whatu* 'stone' and *ahi* 'fire' in table 5. Given these correspondences, it can be reasonably inferred that any POc

[28]Cf. POc **[p*^{w}*ano]p*^{w}*ano* 'Guettarda speciosa' > PPn **(f,p)ano* 'Guettarda speciosa' (Ross et al., 2008, p. 165). See Ross et al. (1998, p. 16) for the reconstruction of POc **p*^{w}.

reconstruction that has a final **p* or **s* indicate inheritance behind the thematic consonant in /-hia/ for this collection of verbs.

Now, we can turn our attention back to the data set at hand: Analogous to the previously analyzed /-mia/ and /-ria/ classes, there exist verbs that are related to each other by virtue of being derived from the same roots. Consequently, there is a total of 66 verbs in total, with 54 representing distinct verb roots, that inflect with /-hia/ in the passive construction. Among these 54 verb roots, 26 appear to possess reliable POc reconstructions presented in (20):[29]

[29] There is irregularity between the initial vowels of POc **kotos* > MAO *kato* and between the vowels of POc **tabe* and Māori *tauapo*.

(20)

Māori Base	Passive	Gloss	POc	Gloss
ao	aohia	‘scoop up’	*agup	‘pick up’
ato	atohia	‘to thatch’	*qatop	‘thatch’
hora	horahia	‘spread out’	*polas	‘spread out’
hou	houhia	‘bind’	*paqus	‘bind, lash’
huri	hurihia	‘convert’	*p^{w}uri-	‘roll’
kato	katohia	‘pick flowers’	*kotos	‘break off’
kōwhiti	kōwhitihia	‘twitch’	*pidik	‘to fillip’
mākiri	mākirihia	‘debone’	*kili	‘turn over’
maro	marohia	‘girdle’	*malo	‘mulberry’
motu	motuhia	‘break’	*motus	‘be broken’
ringi	ringihia	‘pour out’	*liŋis	‘spill’
rongo	rongohia	‘hear, feel’	*roŋoR-	‘hear s.t.’
ruku	rukuhia	‘dive’	*ruku	‘dive’
rūrū	rūrūhia	‘handshake’	*ruru	‘shake’
tangi	tangihia	‘cry, sound’	*taŋis	‘cry, lament’
tango	tangohia	‘take, grasp’	*taŋop	‘take hold of’
tapa	tapahia	‘cut, recite’	*taba(s)	‘cut’
tauapo	tauapohia	‘carry’	*tabe	‘carry’
tīkaro	tīkarohia	‘tear out’	*karis	‘scrape’
tiro	tirohia	‘look, check’	*tirop	‘look intently’
unu	unuhia	‘draw out’	*unus	‘withdraw’
whānako	whānakohia	‘to cheat’	*panako	‘steal’
whakamana	whakamanahia	‘empower’	*mana	‘power’
whakaoti	whakaotihia	‘finish’	*qoti	‘finished’
whakapono	whakaponohia	‘belief, faith’	*bonor	‘true, correct’
whakapoto	whakapotohia	‘shorten’	*botoŋ	‘short’

Regarding the remaining 28 verb stems, while there are no reliable intermediate POc reconstructions available, the application of the comparative method resulted in the identification of comparable higher-level forms for five of them, as evidenced in (21):[30]

[30] According to Ross et al. (2008, 208), the Proto-Oceanic verb root **polas, *polas-i-* ‘spread (s.t.) out’ is reflected in Tongan and Samoan, two Polynesian languages, as *fola* ‘spread’. Given that the /f/ in Tongan and Samoan corresponds to PPn **f*, which typically appears as either /ɸ/, *wh*, or /h/ in Māori (Biggs, 1971, 1978), the forms *hora* ‘spread out’ and *whera* ‘spread’ contend as potential reflexes of the Proto-Oceanic root **polas*. Considering the unexpected

(21)

Māori Base	Passive	Gloss	Reconstructions	Gloss
kapo	kapohia	‘seize, snatch’	PPh **kapét*	‘hold onto’
kimi	kimihia	‘look for’	PAN **kiRim*	‘seek, search’
oro	orohia	‘grind’	PWMP **hurus*	‘slip’
waru	waruhia	‘peel, scrape’	PAN **karus*	‘to scratch’
whera	wherahia	‘spread’	POc **polas*	‘to spread’

Finally, it is worth noting that the earlier forms for the remaining 23 verb roots, except for one,[31] could not be traced due to the unavailability of reconstructed forms at the time of writing this paper. Additionally, my attempts to acquire relevant data through the comparative method were unfruitful. As a result, the following words stand as examples of unknown etymologies:

---

irregularity in the initial vowel of *whera* < POc **polas*, I consider *hora* as the reliable reflex and *whera* as a doubtful one.

[31]The exceptional verb root is *rīngi* ‘ring’, which is a clear loanword from English. This word is not listed in (22).

(22)

Māori Base	Passive	Gloss
amo	amohia	‘bear, carry’
apo	apohia	‘gather up’
ara	arahia	‘arouse’
hura	hurahia	‘discover’
kaiponu	kaiponuhia	‘stingy’
karo	karohia	‘evade, dodge’
karanga	karangahia	‘call out, hail’
kotē	kotēhia	‘squeeze’
manako	manakohia	‘hope for’
mekemeke	mekemekehia	‘box’
neke	nekehia	‘move, shift’
nuku	nukuhia	‘move, shift’
patō	patōhia	‘crack’
rapi	rapihia	‘scratch’
ruke	rukehia	‘throw away’
tīwara	tīwarahia	‘cleave’
wero	werohia	‘pierce’
whakaemi	whakaemihia	‘to assemble’
whakairo	whakairohia	‘carve, figure’
whakangāwari	whakangāwarihia	‘soften, melt’
whakarato	whakaratohia	‘serve (food)’
whakataratara	whakataratarahia	‘provoke’

Now, I direct my attention towards the verb roots with agreed-upon POc reconstructions, as presented in (20), in order to subject them to further analysis. Among the 26 POc reconstructed forms listed in (20), four of them have a final **p* and nine have **s* as their stem-final consonant. In light of the prior discussion on POc **p* and **s* as the sources of Māori /h/, I deduce that the thematic consonant *-h-* in these 13 Māori verbs is an expected result of inheritance. For the sake of clarity and ease of reference, these specific words are reiterated below.

(23)

Māori Base	Passive	Gloss	POc	Gloss
ao	aohia	‘scoop up’	*agup	‘pick up’
ato	atohia	‘to thatch’	*qatop	‘thatch’
hora	horahia	‘spread out’	*polas	‘spread out’
hou	houhia	‘bind’	*paqus	‘bind, lash’
kato	katohia	‘pick flowers’	*kotos	‘break off’
motu	motuhia	‘break’	*motus	‘be broken’
ringi	ringihia	‘pour out’	*liŋis	‘spill’
tangi	tangihia	‘cry, sound’	*taŋis	‘cry, lament’
tango	tangohia	‘take, grasp’	*taŋop	‘take hold of’
tapa	tapahia	‘cut, recite’	*taba(s)	‘cut’
tīkaro	tīkarohia	‘tear out’	*karis	‘scrape’
tiro	tirohia	‘look, check’	*tirop	‘look intently’
unu	unuhia	‘draw out’	*unus	‘withdraw’

Among the remaining 13 POc reconstructions, nine of them has vowels as their stem-final segments. Consequently, the status of the thematic consonant in the Māori reflexes of these reconstructions remains inconclusive, and no definitive conclusions can be drawn at this stage.[32] These verbs are reiterated below.

(24)

Māori Base	Passive	Gloss	POc	Gloss
huri	hurihia	‘convert’	*p^{w}uri-	‘roll’
mākiri	mākirihia	‘debone’	*kili	‘turn over’
maro	marohia	‘girdle’	*malo	‘mulberry’
ruku	rukuhia	‘dive’	*ruku	‘dive’
rūrū	rūrūhia	‘handshake’	*ruru	‘shake’
whānako	whānakohia	‘to cheat’	*panako	‘steal’
whakamana	whakamanahia	‘empower’	*mana	‘power’
whakaoti	whakaotihia	‘finish’	*qoti	‘finished’

Finally, the examination of the remaining 4 verb stems from the set of 26 distinct Māori verb roots introduces the most intriguing scenario within the /-hia/ class. Notably, as shown in (25),

[32] It is possible that the thematic consonant used in the passive forms of these verb roots is also a case of innovation. If that is indeed the case, then the need to find an explanation behind the innovations in Māori is even stronger. The phonotactics-based analysis is undertaken in the following section, which should shed light on the current question, provided there is a pattern. The semantic-based account is to be considered in future work.

the respective POc reconstructions of these 4 verb roots exhibit one of the following four final consonants: **k, *R, *r,* and **ŋ*.

(25)

Māori Base	Passive	Gloss	POc	Gloss
kōwhiti	kōwhitihia	'twitch'	*pidik	'to fillip'
rongo	rongohia	'hear, feel'	*roŋoR-	'hear s.t.'
whakapono	whakaponohia	'belief, faith'	*bonor	'true, correct'
whakapoto	whakapotohia	'shorten'	*botoŋ	'short'

Similar to the approach I took in the /-ria/ class, it is appropriate to review the relevant sound correspondences between POc, PPn, and Māori.

- There is regular correspondence between the POc **k*, PPn **k*, and Māori /k/.
- POc **R* is reflected as zero in PPn and Māori.
- POc **r* is split into **r* and **l* in PPn, and PPn **r* and **l* merged as *r* in Māori.
- POc **ŋ* is directly inherited in PPn and Māori.

Table 7: POc and PPn Sound Correspondences for Māori /k, r, ŋ/.

POc	*k	*R	*r		*ŋ
PPn	*k	∅	*r	*l	*ŋ
MAO	k	∅	r		ŋ

The absence of correspondence between these POc final consonants and the Māori /h/ sound is evident. Nevertheless, it is significant that the Māori reflexes of these POc reconstructed forms in (25) exhibit inflection with the /-hia/ suffix in their passive forms instead of the respective reflexes of the POc stem-final consonants. From this observation, it can be deduced that the thematic consonant *-h-* in these words represents an innovative development. Once again, the question arises: what rationale led Māori speakers to choose the /-hia/ suffix for passivizing these verbs, as opposed to the expected /-kia/, /-ia/ or /-tia/, /-ria/ and /-ngia/? Is there a discernible regular pattern that influenced Māori speakers' selection of this path?

The initial step to undertake, akin to the examination conducted with the /-mia/ and /-ria/ verbs, involves investigating cognates in other contemporary Oceanic languages. This analysis aims to discern whether the observed innovations are confined to Māori, signifying local developments, or if they transpired prior to the emergence of Māori.

POc _*pidik, *p(^{w})idik_ 'to fillip, flick with the finger'

Based on the sound correspondences elucidated in §2.3 and repeated above, the anticipated stem-final consonant for _*-k_ is _†-k_ in Māori. Ross et al. (2016, p. 338) present 13 reflexes of _*p(^{w})idik_ from seven Oceanic language, none of which are Polynesian. ACD also provides seven reflexes from four Oceanic languages, all of which are non-Polynesian. While remnants of the final consonant are evident in some reflexes, such as _pidik_ 'to sting, of insects, nettles, etc., tap with the fingers' in Tolai, a Meso-Melanesian language and _vidik-a_ 'flick s.t., fillip' in Bauan, also known as the standard Fijian language, none of them demonstrate their respective reflexes that is equivalent to Māori /h/. Hence, it is reasonable to surmise that the _-h_ innovation in this word is a local development unique to Māori.

POc _*roŋoR-_ 'hear s.t., listen to s.t.'

Considering the established sound correspondences, the anticipated stem-final consonant for *-ʀis †∅ in Māori. Thanks to its wide semantic range, _*roŋoR-_ has a substantial number of reflexes in Oceanic languages, which allows for a detailed look into how the final _*R_ is reflected. Out of 39 reflexes of 29 languages provided in Ross et al. (2016, pp. 501-02), three show a consonant that is equivalent in position to the final _*-R_ in _*roŋoR, roŋov-i_ 'hear s.t., listen to s.t.' from Bugotu, a Southeast Solomonic language; _roŋo-ð-a_ 'hear s.t.' (its respective immediate reconstruction lower than POc is PCP _*roŋo_ 'hear, be heard') from Bauan Fijian; _oŋo-ʔi_ 'hear s.t.; perceive, feel – pain, pleasure, taste, smell, etc.' from Tongan, a Polynesian language.

A closer inspection of the Bugotu reflexes of POc forms available in Ross et al. (2016, pp. 54-

5) reveals that POc **p* is reflected as /v/ in Bugotu[33] as illustrated by **papine* 'woman, female; sister of man' > *vavine* 'opposite-sex sibling' and **qarop, qarop-i-* 'feel pity, empathy, be sorry for' > *(r)arov-i* 'to pity' (Ross et al., 2016, p. 587). Based on this evidence, I assume that Bugotu, a higher-order language than Māori, has replaced the final **R* in POc **roŋoR-* with /v/, its reflex of POc **p*.

In contrast to the direct sound correspondence of POc **p* > Bugotu *v*, the POc sounds Bauan Fijian reflects with /ð/ are somewhat more varied. With **s* being in majority, **-c-*, and **j-* are among the proto sounds from POc that Bauan Fijian reflects as /ð/ in the data provided in Ross et al. (1998, 2016),[34] e.g., **pasu-* 'facial bony ridge, especially cheek bone' > *vaðu-* 'eyebrows' (2016:119-20); **dumus-i-* (VT) 'suck on, suck up (liquid)' > *domið-* 'sip, suck , as a child at the breast' (2016:248-9); **kiri(s), *kiris-i-* 'tickle' > *kiri, kirið-a* 'tickle under the armpits' (2016, pp. 471–72).

In their discussion of derivational morphology of verbs, Ross et al. points out a transitive verb resegmentation that took place in Fijian where the root-final POc consonants are now considered as allomorphs of the transitive suffix instead of a part of the verb root, and this sometimes led to either the replacement of the etymological consonant by another or the insertion of a consonant where none is expected (1998, p. 24). They illustrate the latter with *roŋo-ð-a* 'hear s.t.' in Bauan Fijian reporting that POc **R* is either reflected as zero or sometimes *r* but never *ð*. In light of this information, it is, then, clear that Fijian innovated the consonant *ð* in *roŋo-ð-a* 'hear s.t.'. As to the question of why /ð/, it is pertinent to highlight that Ross et al. (2016, p. 24), in their examination of the transitivizing morphology **-i* and **-akin[i]*, report that the final consonant preserved in verbs with the transitive suffix ended up being interpreted as part of the suffix in Oceanic languages leading to its replacement by another consonant in languages of the Southeast Solomonic and Fijian groups. The developments witnessed in Bugotu and Bauan Fijian, discussed above, then, may find their underlying explanation in this particular phenomenon.

Lastly, Tongan seems to have replaced the final **R* in **roŋoR-* since /ʔ/ in *oŋo-ʔi* is the reflection

[33]See Ross et al. (2016, p. 248) for a discussion of POc **sosop, *sop-i* 'put lips to, kiss, suck, absorb (moisture)' > Bugotu *sop-i* 'suck'.

[34]Examples include **icuŋ* 'nose' > *uðu-* 'nose' (2016, p. 123) and **jika* 'be soiled, weakened' > *ðika(i)* 'be destroyed, be weakened.'

of **q* (Biggs, 1971, 1978)[35] as is illustrated by examples such as **tau paqoRu* 'young person of marriageable age' > *tāupoʔou* 'virgin, maiden, an esp. attractive young woman' (2016, pp. 65–6) and **Ruqa-* 'neck' > *uʔa* 'neck' (2016, pp. 137–38).

In light of these three innovations and the possibility of Māori having borrowed words from other languages like Bugotu, I propose that this innovation in Māori may be a result of the Bugotu influence considering both of them have a final consonant reflecting POc **p* instead of **R*. Thus, I assume this innovation to be older than Māori.

POc **bonor* 'true, correct'

There is only a curiously small set of reflexes of this proto form available in ACD and Ross et al. (2016, p. 554), all of which are from three Eastern Polynesian languages in total: *pono* in Māori, Hawaiian, and Marquesan. Given the lack of reflexes with final consonants from other languages than Māori, I assume *-h* in question here was a local innovation in Māori.

POc **botoŋ* 'short'

Likewise, a limited number of reflexes are accessible for **botoŋ*, with none of them demonstrating a final consonant (Blust and Trussel, 2020; Ross et al., 2007). Hence, I assume that the innovation of *-h* replacing POc *-ŋ* is also exclusive to the Māori language.

The subsequent course of action, akin to my previous approach, involves scrutinizing the plausibility of these three, presumably Māori-internal innovations being attributed to lexical analogies influenced by phonological and semantic neighbors.

POc **pidik, *p$^{(w)}$idik* 'to fillip, flick with the finger' > *kōwhiti* 'twitch'

The proposed reflex of POc **pidik* in Māori has a larger semantic coverage than the proto form; therefore, it is likely that the innovation might have been brought about by words that are not within the immediate lexical neighborhood. Unfortunately, I am unable to propose any modern

[35] See §2.3 for PPn – POc – Māori sound correspondences.

or proto-forms that share semantic similarity while having the final /h/. For example, despite the apparent phonological similarity in terms of vowel sequence and lexical closeness, the Māori word *tākiri* 'jerk' takes the /-tia/ suffix instead of /-hia/ making it an unlikely candidate to trigger the innovation via lexical analogy.

POc **bonor* 'true, correct' > *whakapono* 'believe, admit as true'

Likewise, no modern or proto-forms within the semantic vicinity of **bonor* or *whakapono* appear to share the final /h/ that could potentially serve as a trigger for the innovation.

POc **botoŋ* 'short' > *whakapoto* 'abbreviate, abridge, shorten'

Finally, once more, my investigation did not yield any lexically similar forms that also exhibit the final /h/, providing no basis for this innovation to be rooted in lexical analogies. Nonetheless, it is prudent to contemplate the potentiality that, akin to the *-r* innovation observed in **tusuq* and **tubuq*, presented in the /-ria/ section, a corresponding hypothetical association could exist between **bonor* and **botoŋ*. In other words, the *-h* innovation in one of these forms might have had an impact on the other, considering their formal resemblance.

The final task, once again, entails searching for a formal pattern that prevails among the verbs in the /-hia/ class, which could also serve as a crucial factor in comprehending the innovations observed in this context. One that draws attention is that many of the verbs have a final /o/ or /ō/ sound; 21 out of 55 distinct verbs roots to be exact.

(26)

Māori Base	Passive	Gloss	POc	Gloss
ato	atohia	'to thatch'	*qatop	'thatch'
patō	patōhia	'crack'		
whakapono	whakaponohia	'belief, faith'	*bonor	'true, correct'

Additionally, 8 of the remaining 34 distinct verb roots end in /u/ or /ū/.

(27)

Māori Base	Passive	Gloss	POc	Gloss
hou	houhia	'bind'	*paqus	'bind, lash'
rūrū	rūrūhia	'handshake'	*ruru	'shake'

There are also 16 distinct verb roots that have the shape /...*(C)a(C)o*/ and 11 other distinct verbs roots have a similar shape except for a round vowel instead of /a/.

(28)

Māori Base	Passive	Gloss	Reconstructions	Gloss
ao	aohia	'scoop up'	POc **agup*	'pick up'
waru	waruhia	'peel, scrape'	PAN **karus*	'to scratch'

Similar to the /-ria/ class, the pattern that we see here is also verbs with a final round vowel, some of which are preceded by /a/ and some others by a round vowel. However, this is a competing pattern with the hypothetical one I tentatively suggested for /-ria/. This especially poses a problem regarding the proposed innovations: given the assumption that the correspondences between POc **k, *t, *r, *R* and Māori *k, t, r, ∅* are regular, the data in (25), as investigated above, presents evidence for an innovated *-h-* as a thematic consonant in the passive forms of the corresponding Māori reflexes. The forms of these verbs, except for *kōwhiti*, are such that they invalidate my conditioning environment hypothesis for /-ria/ verbs since these words also end in a round vowel preceded by another vowel of the same quality. Therefore, I suggest that the preliminary patterns proposed do not have enough explanatory power for the *r* or *h* thematic consonant innovations. A more detailed phonological analysis in search of a robust pattern is presented in §3.

2.3.4 Diachronic Analysis Discussion

§2 set out to examine the historical origins of the thematic consonants surfacing in the passive verbs in Māori. This is done to determine whether they are results of vertical inheritance from POc, or whether they are innovations in the language. While the view of the Māori passive suffixes as unpredictable morphemes remains unchallenged in the former case, innovations, especially those that are local in the language, are regarded as occurrences that are promising for the current study since they point to possible patterns predictive of the suffixes under scrutiny.

12 instances of the thematic consonants *-m-*, *-r-*, and *-h-* appearing in the passive forms of a total of distinct 86 Māori verbs are found to be innovations in the language: 1 in the /-mia/ class, 7 in the /-ria/ class, and 4 in the /-hia/ class.

(29)

Māori Base	Passive	Gloss	POc	Gloss
ngaro	ngaromia	’out of sight’	m^{w}aloq	’submerged rock’
kākahu	kahuria	‘clothe’	*kaput	‘wrap’
takoto	takotoria	‘lie down’	*keteŋ	‘straighten’
tohu	tohuria	‘sign’	*tusuq	‘point at’
tupu	tupuria	‘grow’	*tubuq	‘grow’
whakaako	whakaakoria	‘educate’	*akop	‘learn’
whakamataku	whakamatakuria	‘fright’	*matakut	‘be afraid’
whatu	whaturia	‘weave’	*patuR	‘weave’
kōwhiti	kōwhitihia	‘twitch’	*pidik	‘to fillip’
rongo	rongohia	‘hear, feel’	*roŋoR-	‘hear s.t.’
whakapono	whakaponohia	‘belief, faith’	*bonor	‘true, correct’
whakapoto	whakapotohia	‘shorten’	*botoŋ	‘short’

Of these 12, 10 seem to be innovations that appeared within the Māori speech community without any external affects (e.g., borrowing) while 2, shown below, are suspected of being due to innovations in other languages.

(30)

Māori Base	Passive	Gloss	POc	Gloss
whakamataku	whakamatakuria	‘fright’	*matakut	‘be afraid’
rongo	rongohia	‘hear, feel’	*roŋoR-	‘hear s.t.’

It was emphasized previously that multiple innovations involving the same thematic consonant is a relatively strong indication of a pattern recognized as being associated with that sound by speakers of the language since human beings are very sensitive to linguistic associations. Thus, such local innovations are considered as indicative of the possibility that once the pattern is recognized, one can indeed predict which consonant is to be found in the passive forms of Māori verbs. The findings in this section are such that two out of three classes of verbs[36] have multiple members (6 local in /-ria/ and 3 local in /-hia/) with innovated thematic consonants. As such, I tentatively propose that this is an indication of the possible predictability of the given Māori passive suffixes. In return, however, this preliminary assertion prompts an inquiry into

[36] Though not necessary, a positive correlation between the number of total verbs in a class and the number of innovations is not surprising. /-mia/ is the smallest class with only 6 distinct verb roots as opposed to 26 in /-ria/ and 54 in /-hia/; therefore, only 1 local innovation in the /-mia/ class is not yet alarming.

the discernible patterns that underlie the patterns associated with each thematic consonant. The following section attempts to find those patterns mainly from a phonological perspective.

3 Synchronic Analysis

The findings of the previous section, §2, forms the foundation of the current one. The local innovations found are suggestive of patterns recognized as being associated with certain thematic consonants by Māori speakers. Consequently, the term "patterns" brings to the mind the question of what kind of patterns in what domain may they be? This section investigates the data to provide answers from one domain.

Human beings who are able to speak or sign at least one language are exposed to all the layers of that language from its phonetic properties to pragmatic features found within the speech community. Therefore, when humans acquire or learn a linguistic item, they do not do so in an isolated fashion; the item in question has a phonological shape, a lexical meaning or a grammatical function, a syntactic category, and so on. That is, every linguistic expression has a bundle of features associated with itself, and similar expressions are deemed similar because they share common features with each other, and humans are great at detecting those similarities—what are called "patterns" in this context. Moreover, while even young children demonstrate exceptional abilities in grasping such linguistic patterns and resolving linguistic enigmas in their native languages within very short periods of time, these skills may not be as readily accessible to the conscious mind of a linguist engaged in language study. Certainly, a linguist studying a language other than their native language(s) does not have access to all the features in the bundle making up every linguistic expression in the same way a native speaker of that language does. Therefore, the patterns sought after in this study are not easy to find, especially considering that they have features from different domains. As a result, the current analysis is focused on only one aspect of the Māori verbs: their phonological shapes and features.

Parker Jones (2008) underscores the advantages of employing computational methods in linguistic research, particularly in facilitating the execution of intricate statistical analyses over extensive data sets. Furthermore, Parker Jones makes the suggestions of a more comprehensive and meticulous examination of the phonological qualities inherent in Māori verb roots, with the aim of uncovering potential regularities (i.e., patterns) underlying the choice of thematic consonants in passive constructions.

Following Parker Jones (2008), in this section, I examine the Māori verb roots in the data set in order to figure out whether the predictive patterns, which are pointed to in §2, reveal themselves in the phonological aspects of active Māori verbs. To this end, exhaustive analyses of the surface forms of the Māori verbs in the corpus were conducted. One primary category of analyses is based on the phonological segments of the verb root, and the other is based on their phonological features. If any of these analyses find strong associations between any pattern and its corresponding thematic consonant, then the suggestion of predictability of the Māori passive suffixes will be strengthened further since such a finding will serve as an example of explicit demonstration of what Māori speakers might have detected in the active verbs when they made innovations with the thematic consonants.

3.1 Linguistic Feature-Based Analysis

A program is developed to conduct the feature-based analyses. The input file of the program consists of two columns, where the first column lists the Māori verb roots, and the second column lists their corresponding passive suffixes.

Table 8: Examples from the Māori data set used in the phonotactics-based analysis.

Māori Verb Root	**Passive Suffix**
ā	ia
epa	ina
hahau	tia

As mentioned before (§1.1), each verb root is associated with only one passive suffix in the data set of 889 total verbs. The phonological assumptions made in the presentation of the data are consistent with those given in §2.2.

The data set includes verbs derived by the causative prefix *whaka-* and also verbs that contain partial or total reduplication. Since this analysis is intended to examine the phonological properties of verb roots, the derived verbs were excluded, and the analyses were run on the remaining 734 verb root-suffix pairs. This has left us with 7 /-mia/, 28 /-ria/, and 54 /-hia/ verbs analyzed.

In this analysis, 16 feature-based formal qualities, which encompass stem segments, segment features, segment sequences, and the stem syllable counts of the active Māori words, were extracted. This is a much larger number compared to the three coding schemes Parker Jones explicitly encodes in his study (2008).

The following subsections offer a presentation of the extracted features, along with a detailed examination of their distributions per suffix. For the sake of coherence, certain features that undergo similar analyses have been grouped together. An illustrative example is the treatment of root-final vowels and their distinctive feature, discussed jointly in §3.1.1. Finally, a concluding discussion is presented.

The tables provided for each main feature adhere to a specific format: patterns that exhibit noteworthy relative frequency are listed in the first column. For example, out of 7 /-mia/ verbs in total, 3 end in /o/ and the remaining 4 end in /u/, which makes both root-final vowels prominent enough to be included in table 9 given their coverage of the available data. The second column indicates the total number of verbs in the data set that display the specific pattern. For example, a value of "90" in the second column of table 9 indicates that out of the 734 verbs in the data set, 90 end with the vowel /o/. Finally, the last column depicts the probability of the given pattern being associated with the corresponding suffix, relative to its occurrence with all suffixes in the data set. This probability is computed by dividing the total count of associations between the specific pattern and the relevant suffix by the total number of occurrences of the pattern with any suffix.

$$P(\text{suffix}|\text{feature}) = \frac{\text{suffix count per given feature}}{\text{total count of verbs with that feature}} \tag{1}$$

To illustrate, the value ".0333" displayed in the fourth column of table 9 denotes the probability of the co-occurrence of root-final /o/ and the /-mia/ suffix. This probability is calculated by dividing the count of occurrences where both elements coincide (3 instances) by the total number of verbs that end with /o/ (90 occurrences).[37]

$$P(\text{/-mia/}|\text{/-o/}) = \frac{3}{90} = .0333 \tag{2}$$

The patterns selected as prominent and subsequently presented in the tables are based on their high frequency counts relative to other patterns associated with each suffix for every feature. For instance, the inclusion of the /u/ sound in table 10 is due to the fact that 19 out of 28 /-ria/ verbs end with that specific vowel.

Finally, in order to facilitate the evaluation of how informative each quality analyzed is, the probabilities of the corresponding suffixes in the data are also provided in the table captions. In order for a property to be informative enough about a suffix, its conditional probability with the suffix must be considerably higher than the probability of the suffix alone since co-occurrence would indicate association. However, if the probabilities are approximately the same, then we could conclude that the quality tested is not informative enough.

3.1.1 Final Vowels and Final Vowel Features

3.1.1.1 /-mia/

Of a total of 7 /-mia/ verbs, 3 of them ends in /o/ and the remaining 4 ends in /u/, which are [$-$high, $-$low, $+$back, $+$round, $-$long] and [$+$high, $+$back, $+$round, $-$long] respectively. As mentioned before, they are both [$+$back, $+$round] vowels. However, there are 90 verb roots ending in /o/ and 135 verbs roots ending in /u/ in the data set. Therefore, the probabilities we get from the root-final /o/ and /u/ and /-mia/ association are quite low as presented.

[37] For the sake of readability and interpretability, all probabilities have been rounded to four decimal points.

Table 9: Final vowels of /-mia/ verbs; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Final	o	3	90	.0333
Vowel	u	4	135	.0296

There may be a unidirectional association between the [+back, +round] vowels and the passive suffix /-mia/ in Māori: although /-mia/ attaches only to those verb roots ending in /o/ or /u/, a root-final /o/ or /u/ cannot predict the selection of /-mia/.

3.1.1.2 /-ria/

Among the five distinct vowels observed root-finally showing in 28 /-ria/ verbs, merely two of them exhibit a probability exceeding .1 of co-occurring with this suffix.

Table 10: Final vowels of /-ria/ verbs; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Final	ū	1	6	.1667
Vowel	u	19	135	.1407

Again, even though /-ria/ seems to have a tendency to attach to [+high, +back] vowels (20 out of 28 verb endings), it is hard to propose that these vowels condition the /-ria/ attachment given the total number of verbs taking other suffix despite ending in the same two vowels.

3.1.1.3 /-hia/

Each of the 54 /-hia/ verbs end in one of the 7 distinct vowels extracted; however, only one of them, /o/ as previously observed, has the highest count of these 7 divisions. The three main root-final vowels /-hia/ seems to like are given in table 11.

Table 11: Final vowels of /-hia/ verbs; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Final Vowel	o	23	90	.2556
	a	8	149	.0537
	i	11	220	.0500

Despite the relatively high co-occurrence instances, it is difficult to propose that any of these final vowels are predictive of /-hia/. Furthermore, besides [−long], there is no other feature that all three vowels share.

3.1.2 Vowel Sequences and Vowel Sequence Features

3.1.2.1 /-mia/

Within the 7 /-mia/ verbs under examination, we identify a total of five distinct linear (left-to-right) combinations of vowels. Two prominent patterns are provided below.

Table 12: Vowel sequences of /-mia/ verbs; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Vowel Sequence	ao	2	23	.0870
	au	2	31	.0645

3.1.2.2 /-ria/

28 /-ria/ verbs exhibit 17 vowel sequence patterns, the most populous of which are given in table 13.

Table 13: Vowel sequences of /-ria/ verbs; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Vowel Sequence	aau	3	6	.5000
	au	9	31	.2903

The co-occurrence counts seen here are in line with the preliminary observation of a pattern showing up in /-ria/ verbs, presented in §2.3.2, namely /...*a(C)u(C)(a)(u)*/. Even though the probabilities are substantial, a larger data set with higher number of /-ria/ verbs would make it easier to confirm or reject this tentative association.

3.1.2.3 /-hia/

33 vowel sequences are observable for the total of 54 /-hia/ verbs. Table 14 presents two relatively notable patterns.

Table 14: Vowel sequences of /-hia/ verbs; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Vowel	āao	2	3	.6667
Sequence	ao	20	23	.4348

Likewise, while this pattern bears resemblance to the initial observation stated in §2.3.3, namely /...*(C)a(C)o*/, the probabilities associated with these patterns do not attain a sufficiently high level to confidently posit a robust predictability governing the selection of /-hia/ as a passive suffix. Even though all possible sequences of two short vowels are attested in Māori (Bauer, 1993, p. 534), the dominant pattern common to all three suffixes is /a/, a [−round] sound, followed by one of the [+round] vowels, /o, u/ as observed in tables 10-12. While this seems to be a coincidence born out of the choice of suffixes in this study, it is also not surprising to see that the phoneme /a/ occurs often considering that it is the most common phoneme in the language (Harlow, 2007, p. 68).

3.1.3 Final Consonant and Final Consonant Features

Prior to delving into the specifics, it is crucial to clarify that, in the context of this analysis, what I refer to by the term "final consonant" is the consonant preceding the true root-final vowel, taking into account the open syllable structure and absence of root-final consonants in Māori.

3.1.3.1 /-mia/

There are only two patterns observed with an almost-even distribution: /n/ and /r/ as shows in table 15. However, since these final consonants seem to be common among all the verbs, the probabilities of these sounds hinting at /-mia/ suffixation are quite low.

Table 15: Final consonants of /-mia/ verbs; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Final Consonant	n	4	65	.0615
	r	3	164	.0183

3.1.3.2 /-ria/

/-ria/ verbs exhibit five different final consonants, and 12 /-ria/ verbs, the largest number in this group in terms of final consonant patterns, end in one of them.

Table 16: Final consonants of /-ria/ verbs; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Final Consonant	t	12	96	.125

However, the probability is still too low to be indicative of a /-ria/ selection following a final /t/.

3.1.3.3 /-hia/

There are eight distinct consonants /-hia/ verbs have as their final consonantal segment, and two prominent tendencies.

Table 17: Final consonants of /-hia/ verbs; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Final Consonant	r	18	164	.1098
	t	9	96	.0998

As is seen in table 17, /-hia/ verbs seem to have an abundance of /r/ and /t/ as their final consonant; however, these two sounds seem to be popular among all the verbs in the data set, resulting in probabilities of negligible magnitude that do not hold substantial predictive capacity for the /-hia/ suffix.

A general observation is that all consonants for /-mia, -ria, -hia/ verbs are [+coronal]; however, this is not very informative for our purposes of finding predictive patterns for individual suffixes.

3.1.4 Consonant Sequences and Consonant Sequence Features

3.1.4.1 /-mia/

7 /-mia/ verbs show five distinct patterns in terms of the sequences of consonants, with 3 of them preferring one single consonant.

Table 18: Consonant sequences of /-mia/ verbs; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	n	3	9	.3333

Despite the relatively high probability, /-mia/ is certainly not in the dominant suffix Māori verb roots with a single /n/ consonant seem to prefer. Note that the reason why only *n* is presented despite the "consonant sequence" label is because of the dominance of verb roots with only one consonant in the /-mia/ category, e.g., *inu* 'drink' and *nao* 'handle, lay hold of.'

3.1.4.2 /-ria/

There is a wide range of variation in consonant sequences of /-ria/ verbs, namely 21 for 28 total /-ria/ verbs, hence the distribution is shallow. Three main tendencies are shown below.

Table 19: Consonant sequences of /-ria/ verbs; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant Sequence	tp	3	16	.1875
	k	3	20	.1500
	t	3	22	.1354

3.1.4.3 /-hia/

54 total /-hia/ verbs exhibit 37 patterns of consonant sequences resulting in low numbers of frequency count per pattern.

Table 20: Consonant sequences of /-hia/ verbs; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	rng	3	6	.5000
	hr	3	15	.2000
	r	3	18	.1667

The three patterns with the highest frequency count of 3 are *...r...n...g...*, *...h...r...*, and *...r...* among /-hia/ verbs. Although half of *...r...n...g...* verbs are associated with /-hia/, the total number of that pattern is too low for its probability to represent a meaningful prediction.

3.1.5 Consonant Sequence Features: [±CONSONANTAL]

In consideration of spatial constraints and formatting limitations, the abbreviation "[±cs]" is employed as an abbreviation for [±consonantal] in this context.

3.1.5.1 /-mia/

Following are the 3 patterns /-mia/ verbs represent.

Table 21: Consonant sequences of /-mia/ verbs: [±consonantal]; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	[+cs]	4	109	.0367
	[−cs][+cs]	1	55	.0182
	[+cs][+cs]	2	278	.0072

The low probabilities are an indication of lack of predictive power of the [±consonantal] quality of each consonant in consonant sequences for /-mia/.

3.1.5.2 /-ria/

6 distinct patterns extracted show two prominent tendencies.

Table 22: Consonant sequences of /-ria/ verbs: [±consonantal]; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant Sequence	[+cs]	8	109	.0734
	[+cs][+cs]	9	278	.0324

Likewise, the probabilities show that [±consonantal] feature is not informative for /-ria/ selection, either.

3.1.5.3 /-hia/

Out of 7 distinct patterns, two merit exhibition.

Table 23: Consonant sequences of /-hia/ verbs: [±consonantal]; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	[+cs][+cs]	27	278	.0971
	[+cs]	9	109	.0826

Despite the relatively high frequency counts for /-hia/ verbs, the probabilities are still too low to be indicative of a regularity of /-hia/ selection.

3.1.6 Consonant Sequence Features: [±SONORANT]

Likewise, I employ "[±sn]" as an abbreviation of [±sonorant] in this section for the same spatial reasons.

3.1.6.1 /-mia/

Two out of three distinct patterns stand out.

Table 24: Consonant sequences of /-mia/ verbs: [±sonorant]; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant	[+sn]	4	43	.0930
Sequence	[−sn][+sn]	2	122	.0164

3.1.6.2 /-ria/

Of 7 distinct patterns, one shows up in 8 and another represents 10 other /-ria/ verbs.

Table 25: Consonant sequences of /-ria/ verbs: [±sonorant]; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant	[−sn]	8	87	.0920
Sequence	[−sn][−sn]	10	151	.0662

Even though there seems to be a tendency for [−sonorant] consonants, the low probabilities indicate a lack of predictive power toward the /-ria/ suffix in these patterns.

3.1.6.3 /-hia/

in table 26, I present the most frequent two patterns out of 14 for the /-hia/ verbs.

Table 26: Consonant sequences of /-hia/ verbs: [±sonorant]; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	[+sn][+sn]	8	60	.1333
	[−sn][+sn]	10	122	.0820

Although these two patterns have relatively higher frequency counts, they are still uninformative.

3.1.7 Consonant Sequence Features: [±NASAL]

I use "[±N]" in place of [±nasal] for spatial reasons.

3.1.7.1 /-mia

One pattern, out of five, is presented below.

Table 27: Consonant sequences of /-mia/ verbs: [±nasal]; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	[+N]	3	22	.1364

Despite a slight apparent tendency, the probability is again too low to be informative.

3.1.7.2 /-ria/

Below, the most frequent two, out of seven, patterns are presented.

Table 28: Consonant sequences of /-ria/ verbs: [±nasal]; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant Sequence	[−N]	8	108	.0741
	[−N][−N]	10	279	.0358

Even though both patterns exhibit only [−nasal] consonants, no discernible predictive power is evident from this analysis.

3.1.7.3 /-hia/

Three most frequent patterns, out of 12, are shown.

Table 29: Consonant sequences of /-hia/ verbs: [±nasal]; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	[−N][+N]	7	67	.1045
	[−N][−N]	20	279	.0717
	[−N]	7	108	.0648

Likewise, despite the tendency for sequences of mostly [−nasal] consonants, none of these patterns holds predictive power for /-hia/.

3.1.8 Consonant Sequence Features: Place of Articulation

There are only [+labial], [+coronal], and [+dorsal] consonants in Māori, and for the same formatting reasons, they are abbreviated as [L], [C], and [D] respectively in this section.

3.1.8.1 /-mia/

There seems to be a slight tendency for [+coronal] sounds in the /-mia/ class.

Table 30: Consonant sequences of /-mia/ verbs: place of articulation; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	[C]	4	49	.0816

However, this pattern is certainly not predictive of /-mia/.

3.1.8.2 /-ria/

Table 31 presents the two most frequent patterns out of 14 total.

Table 31: Consonant sequences of /-ria/ verbs: place of articulation; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant	[D]	4	41	.0976
Sequence	[L][C]	4	91	.0440

Despite having the highest frequency, these two patterns are not informative enough about the selection behind /-ria/.

3.1.8.3 /-hia/

/-hia/ verbs also exhibit two prominent patterns out of 16.

Table 32: Consonant sequences of /-hia/ verbs: place of articulation; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant	[C][D]	9	59	.1525
Sequence	[L][C]	9	91	.0989

In addition to the lack of any specific discernible sequencing, these patterns do not exhibit predictive power for /-hia/.

3.1.9 Consonant Sequence Features: [±CONTINUANT]

The same spatial reasons call for an abbreviation for this feature as well; I use "[±CT]" for [±continuant].

3.1.9.1 /-mia/

One pattern stands out among five total.

Table 33: Consonant sequences of /-mia/ verbs: [±continuant]; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	[−CT]	3	83	.0361

Although three out of seven /-mia/ verbs make this selection, both the numbers of verbs and the probabilities are too low to be informative.

3.1.9.2 /-ria/

Two out of eight total patterns are worth taking a look at below.

Table 34: Consonant sequences of /-ria/ verbs: [±continuant]; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant	[−CT]	8	883	.0964
Sequence	[−CT][−CT]	8	157	.0510

The tendency to have [−continuant] consonants are still not indicative of the /-ria/ ending given the low probabilities.

3.1.9.3 /-hia/

15 /-hia/ verbs, the most crowded group for the [±continaut] feature, select the following pattern.

Table 35: Consonant sequences of /-hia/ verbs: [±continuant]; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	[−CT][−CT]	15	157	.0955

Similarly, the relative dominance of [−continuant] segments are not powerful enough to prompt the selection of /-hia/.

3.1.10 Consonant Sequence Features: [±SPREAD GLOTTIS]

I adopt the commonly used [±SG] abbreviation for [±spread glottis].

3.1.10.1 /-mia/

The following three patterns are observed for seven /-mia/ verbs.

Table 36: Consonant sequences of /-mia/ verbs: [±spread glottis]; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	[−SG]	4	112	.0357
	[+SG][−SG]	1	46	.0217
	[−SG][−SG]	2	297	.0067

Given the pattern distribution and low probabilities, there can be no prediction based on the [spread glottis] feature of the consonants in /-mia/ verbs.

3.1.10.2 /-ria/

Below, I present the three most frequent patterns in /-ria/ verbs.

Table 37: Consonant sequences of /-ria/ verbs: [±spread glottis]; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant Sequence	[−SG]	8	112	.0714
	[−SG][−SG][−SG]	7	126	.0556
	[−SG][−SG]	9	297	.0303

We observe a sequence of consonants that are [−spread glottis], but that is not a predictive feature of /-ria/ verbs given the low probabilities.

3.1.10.3 /-hia/

Among six patterns exhibited by /-hia/ verbs, three most frequent are provided below.

Table 38: Consonant sequences of /-hia/ verbs: [±spread glottis]; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	[−SG][−SG]	29	297	.0976
	[−SG]	9	112	.0804
	[−SG][−SG][−SG]	10	126	.0794

Once more, notwithstanding the inclination towards phonemes with the [−spread glottis] feature, the low probabilities indicate the lack of predictive power behind these patterns.

In fact, 9 out of 10 consonantal phonemes of Māori are [−spread glottis], and the only [+spread glottis] sound, /h/, has the lowest frequency percentage of 4.5 among all the vocalic and consonantal segments in the language according to Harlow (2007, p. 68). Hence, the [−spread glottis] dominancy in the observed patterns for each of the three suffixes is to be anticipated.

3.1.11 Consonant Sequence Features: [±voiced]

[±V] for [±voiced] is used in this section.

3.1.11.1 /-mia/

Table 39 shows the consonant sequences of /-mia/ verbs based on their voicing features.

Table 39: Consonant sequences of /-mia/ verbs: [±voiced]; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant Sequence	[+V]	4	43	.0930
	[+V][+V]	1	60	.0167
	[−V][+V]	2	122	.0164

[+voiced] sounds seem to dominate the patterns, but their probabilities are too low to be indicative of a /-mia/ selection.

3.1.11.2 /-ria/

Of the seven patterns, the most frequent two are presented below.

Table 40: Consonant sequences of /-ria/ verbs: [±voiced]; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant Sequence	[−V]	8	87	.0920
	[−V][−V]	10	151	.0662

/-ria/ verbs tend to be composed of voiceless consonantal segments; however, the probabilities are again too low for this to be a feature exclusively associated with /-ria/.

3.1.11.3 /-hia/

The two most frequent patterns, among fourteen in total, are shown in table 41 below.

Table 41: Consonant sequences of /-hia/ verbs: [±voiced]; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant	[+V][+V]	8	60	.1333
Sequence	[−V][+V]	10	122	.0820

Voiced consonant seem to dominate the patterns for /-hia/ verbs, but as in the cases of other suffixes, voicing is not a predictive feature of /-hia/ verbs, either.

3.1.12 Syllable Counts

Finally, the syllable count frequencies per suffix are presented in the following sections. As is the practice, I use lowercase sigma, σ, per syllable. In this section, it is important to point out that the basic root shape in Māori is bimoraic with the root morpheme shape of (C)V(C)V (Harlow, 2007, p. 127).

3.1.12.1 /-mia/

There are only two patterns observed in /-mia/ verbs.

Table 42: Syllable counts of /-mia/ verbs; P(/-mia/) = .0095.

		/-mia/ Verb Count	Total Verb Count	P(/-mia/ | ...)
Consonant	σσ	6	410	.0146
Sequence	σ	1	72	.0139

Given the basic root type of Māori, the dominance of disyllabic pattern is not very informative for /-mia/ verbs alone.

3.1.12.2 /-ria/

/-ria/ verbs show the following three patterns.

Table 43: Syllable counts of /-ria/ verbs; P(/-ria/) = .0381.

		/-ria/ Verb Count	Total Verb Count	P(/-ria/ | ...)
Consonant Sequence	σ	5	72	.0694
	σσσ	9	193	.0466
	σσ	11	410	.0268

We also observe trisyllabic verbs in the /-ria/ class, but all the probabilities are too low for any of these three patterns to be informative enough about the /-ria/ suffix.

3.1.12.3 /-hia/

Lastly, /-hia/ verbs exhibit the following four patterns.

Table 44: Syllable counts of /-hia/ verbs; P(/-hia/) = .0736.

		/-hia/ Verb Count	Total Verb Count	P(/-hia/ | ...)
Consonant Sequence	σσ	40	410	.0976
	σσσσ	1	18	.0556
	σσσ	10	193	.0518
	σ	2	72	.0278

A range of syllable counts are exhibited in verb roots taking the /-hia/ suffix, but the probabilities are again too low for any of these patterns to be exclusively associated with /-hia/.

3.1.13 Nearest Phonological Neighbors

In the course of the diachronic analysis of the Māori verb roots taking one of the three passive suffixes under examination, /-mia, -ria, -hia/, one goal was to ascertain the origins of the thematic consonants *-m-*, *-r-*, and *-h-* in cases of innovations. In an endeavor to explore potential explanations for these innovations, a preliminary examination was undertaken to determine whether lexical factors, situated within the phonological and semantic context of neighboring verbs sharing the same thematic consonant in their passive forms, could have contributed to their emergence. It is crucial to note that I adopted a broader scope of comparison in my previous endeavor to identify nearest phonological and semantic neighbors. For instance, in addition to examining other verb roots within each relevant verb class, /-mia, -ria, -hia/, I also conducted a comprehensive examination of all POc reconstructions and Māori words featuring final /m, r, h/ available on ACD with careful consideration of their semantic associations to identify potential similarities to discern formal and lexical resemblances to the roots displaying innovations. However, the current analysis takes a more focused approach, limiting the comparison solely to the verb roots exhibiting innovations and those within the same class for the sake of increased manageability and facilitation of the analytical process. That is, all the verbs with innovations individually examined below, and their corresponding neighbors found inflect with the same passive suffix carrying the same thematic consonant.

Even though a wide-scope semantic analysis is not possible at this time, as mentioned previously, I was able to conduct a more systematic analysis for phonological neighbors using a BK-tree[38] (Burkhard and Keller, 1973) structure employing the Levenshtein distance metric since the type of data being compared is string of characters.

The diachronic analysis, presented in §2, identified 12 innovations within the data set. The subsequent sections present the phonological neighbors of each verb root exhibiting innovation. It is important to highlight that, in order to maintain focus on the examined verb roots, the causative versions are ignored, if present in the data set, as they are derived from the verb roots under investigation. Furthermore, in cases where the verb exhibiting innovation solely exists in its causative form within the data set, a distinct approach is adopted. Specifically, the causative prefix (i.e., *whaka-*) is removed, and the nearest phonological search is conducted for the remainder of the verb, which is the verb root in each verb examined. For instance, *whakapono* 'belief, faith' is present in the data set while the verb root *pono* is absent.[39] Nevertheless, a search is conducted for phonological neighbors associated with *pono*, considering that *whakapono* is derived from *pono* through affixation.

[38] A BK-tree is a data structure designed to efficiently search for elements based on their similarity to a given reference element. It offers a practical and efficient way to find elements that are similar to a given reference element. The "given element" is my analysis is each verb root whose thematic consonant is an innovation.

[39] This is because the two main dictionaries, Biggs (1990) and Ryan (1989), that the corpus is built on do not list *pono* with a passive suffix, hence the exclusion from the corpus.

The purpose behind looking up the nearest phonological neighbors of the Māori roots with innovated thematic consonants is to evaluate the possibility of innovation due to lexical analogy within the phonological and semantic vicinity of a verb. Clearly, phonological similarity is only one half of the puzzle, and it is difficult to reach firm conclusions without an investigations of the lexical content. The following subsections include discussions of semantic similarity of the Māori verbs and their phonological neighbors; however, it is important to note that this is a highly informal approach devoid of systematic evaluation of the lexical contents. As previously mentioned, systematic semantic analysis will be undertaken in future research.

3.1.13.1 /-mia/

The current investigation has yielded only one Māori verb root with an innovated thematic consonant in this class: *ngaro* 'out of sight, absent, destroyed, disappear, disappeared, hidden, lost, missing.' The nearest phonological neighbors, also inflected with /-mia/ in the passive, are *aru* 'chase, follow, pursue; woo, court; interrupt' and *nao* 'handle.' There is no obvious semantic relation between *ngaro* and its phonological neighbors.

3.1.13.2 /-ria/

According to the analysis, there are seven cases of innovation in the data set, and the nearest phonological neighbors of each are given below.

kākahu **'cloak, garment; clothe, dress'**

There are two phonological neighbors suggested; one is *ahu* 'tend, foster', and the other one, quite expectedly, is *kahu* 'cloak' since the two verbs are related to each other by virtue of sharing the same root.[40] Thus, the innovation is clearly shared by *kākahu* and *kahu*, and it is not a matter of one affecting the other. As for *ahu*, it does not appear to be semantically close to *kākahu*.

[40] See footnote 22 for why *kākahu* is treated as the verb root in this paper.

***takoto* 'lie down'**

The only verb root found is *tū* 'stand, stop.' The meanings of these two verbs are considered to be distant from each other.

***tohu* 'sign; signal'**

There are four verb roots found to be phonologically nearest to *tohu*: *ahu* 'tend, foster'; *kahu* 'cloak'; *tupu* 'grow, sprout; growing shoot of a plant or tuber; real, genuine'; *tau* 'count, calculate, figure.' None of the neighbors seem to be semantically close to *tohu*.

***tupu* 'grow, sprout; growing shoot of a plant or tuber; real, genuine'**

Five verb roots are detected to be the nearest: *apu* 'heap, heaped up, heap up'; *tohu* 'sign; signal'; *tau* 'count, calculate, figure'; *whakatapu*[41] 'sanctify, consecrate, hallow, cherish'. One may suspect some kind of semantic similarity between *tupu* and *apu*, but it is not possible to measure yet. It does not seem to be semantically close to the other neighbors.

***whakaako* 'educate'**

As explained above in §3.1.13, the neighbors of the expected verb root, *ako*, were searched for. There are three neighbors found: *ahu* 'tend, foster'; *apu* 'heap, heaped up, heap up'; *whakamataku* 'fright, frighten, frightened, scare, terrify'. Out of the three, *ahu* may be considered as being semantically closest to *ako*.

***whakamataku* 'fright, frighten, frightened, scare, terrify'**

Likewise, a search for the phonological neighbors of *mataku* was undertaken. The five phonological neighbors found are *ahu* 'tend, foster', *apu* 'heap, heaped up, heap up', *kau* 'swim', *mātau* 'adept, conscious, intelligent, know, skill, skilful, understand, wise', and *tau* 'count, calculate, figure'. None of the phonological neighbors seem to present lexical similarity to *mataku*.

[41] The main source, Biggs (1990, p. 26), that are used to build the data set only lists *whakatapu-ria* and not *tapu-ria*.

whatu 'weave'

Similarly, there are four neighbors found all of which were mentioned for other verbs above: *ahu* 'tend, foster'; *apu* 'head, heaped up, heap up'; *tū* 'stand, stop'; *tau* 'count, calculate, figure'. *Whatu* does not seem to be semantically close to any of its phonological neighbors presented here.

3.1.13.3 /-hia/

Four Māori verb roots that are associated with /-hia/ as a result of innovation have been detected. The nearest phonological neighbor(s) of each is provided below.

kōwhiti 'sort by size, twitch, turn inside out, scoop from shell, rise of new moon'

Four neighbors found are *ao* 'gather up, scoop up', *huri* 'convert, discuss, invert, overthrow, overturn, turn inside out, turn the back, turn upside down; grind (flour), turn; trump (in playing cards)', *ato* 'thatch, to thatch', and *kimi* 'look for, search, seek'. *Ao* and *huri* appear to be semantically much closer to *kōwhiti* than the other two.

rongo 'feel, sense, hear, obey, smell, taste v.'

Seven verb roots are detected as the nearest phonological neighbors of rongo: *ao* 'gather up, scoop up'; *amo* 'bear, carry on a litter, carry on the shoulder; bier, carrier, litter, stretcher'; *apo* 'gather together, grasp, take, heap up; avaricious'; *hou* 'bind'; *karanga* 'call out, hail, shout, summon; ceremonial call of welcome onto a marae'; *ato* 'thatch, to thatch'; *oro* 'grind on a stone'. *Rongo* has gone through a range of semantic extensions in many Oceanic languages (Ross et al., 2016, pp. 499-503); however, none of its phonological neighbors in Māori seem to have any lexical similarity to it.

whakapono 'belief, faith, religion, trust, credence, ritual over victim'

As explained above in §3.1.13, I searched for neighbors of *pono*, which yielded three in total: *ao* 'gather up, scoop up', *kaiponu* 'avaricious, stingy, withhold', and *oro* 'grind on a stone'. None of the three neighbor exhibit semantic resemblance to *pono*.

***whakapoto* 'abbreviate, abridge, shorten'**

Likewise, the nearest phonological neighbors presented here are extracted for *poto* even though it does not exist without the causative prefix in the data set. There is seven neighbors found: *ao* 'gather up, scoop up'; *motu* 'snap, break (as a stick, bone), severed'; *ato* 'thatch, to thatch'; *kato* 'pick flowers, pluck'; *patō* 'crack, cracked'; *kote* 'squeeze, squash, form of spell to bewitch'; *oro* 'grind on a stone'. No clear neighbor seem to also be in the semantic vicinity of *poto*.

3.1.14 Feature-Based Analysis and Nearest Phonological Neighbor Discussion

This section presents a comprehensive analysis of the /-mia, -ria, -hia/ verbs, encompassing 16 distinct formal qualities. Despite observing relatively high frequencies in certain patterns, the overall probabilities are so low that no robust characteristics have been identified that definitively determine the selection of any of the three suffixes.

This finding diverges from Parker Jones's suggestion of possible inherent formal subregularity within Māori verb roots, where specific features might act as predictive indicators for individual passive suffixes.

Regarding the nearest phonological neighbors, 11 out of 12 verbs that exhibit clear innovations (the exception is *takoto* with only one neighbor, *tū*) possess two or more neighbors within their phonological vicinity, all belonging to the same verb class denoted by the /-mia, -ria, -hia/ suffixes in this study. A lack of evident semantic correspondence between most verbs and their respective neighbors has been observed. Despite this semantic disparity, I propose that these innovations might have resulted from lexical analogy based on phonological neighbors within the same verb class. This preliminary assertion is stronger in the cases of *whakaako* and *kōwhiti* since they appear to have semantic relations with some of their phonological neighbors.

4 Discussion

This study has taken a comprehensive approach to one of the biggest puzzles coming from the Polynesian languages, the Māori passives. While the abundance of passive suffixes with various initial consonants available in the language presents a surprising case itself, it is also quite convincing to propose that, depending on the dialect, there is one passive suffix that has earned the default status in time, and that it is not possible to predict which one of the other suffixes is to be used with a novel word if not the default one. However, inspired by Moorfield (1988) and Harlow (2007), Parker Jones (2008) carries out a preliminary study that challenges this long-held belief with highly accurate correct predictions from a neural network model he builds that uses phonotactic information of active Māori verbs.

If one assumes that there are indeed patterns that lead to the selection of passive suffixes with certain consonants, it is a straightforward matter to wonder whether such a selection has taken place in the history of the language; that is, whether any of the morpheme-initial consonants in the passive suffixes of certain verbs was an innovation to replace another consonant or to insert a consonant when none was there. To this end, first, a diachronic approach was employed to examine the Māori verbs that select one of the three passive suffixes that were the focus of this study: /-mia, -ria, -hia/. Of a total of 114 verbs analyzed, 86 have distinct verb roots 12 of which acquired the consonant in their suffixes via innovation in place of another consonant that used to occupy the stem-final positions in those verb in pre-Polynesian times. It is important to reiterate that the present analysis refrained from delving deeper into verbs whose earlier reconstructed forms also do not have any root-final consonants. While there is no reason to assume that the consonants in their passive forms do not represent cases of innovations, the current study's temporal constraints require me to defer this aspect for future inquiry.

Second, building upon the concept proposed by Parker Jones (2008), which is substantiated by the innovations discovered, an exhaustive examination based on formal features was undertaken. This analysis encompassed 16 distinct phonological features across all verb roots, with the objective of elucidating notable patterns within the subset of verb roots. The focus of this investigation was on those verbs that exhibit passive forms using one of the three specific suffixes in question within this research. Although certain overarching segmental patterns are discernible, namely the /*...a(C)u(C)(a)(u)*/ pattern for /-ria/ verbs and /*...(C)a(C)o*/ for /-hia/ verbs, the meticulous scrutiny of the examined features does not reveal any consistent patterns of sufficient strength to predict the occurrence of these specified suffixes.

Even though the Māori passive formation stands out with the large number of suffixes available and their assumed unpredictability, it is not difficult to see how it is reminiscent of more well-known cases like the English past tense debate and the German plurals especially given that the passivization process with the default suffixes in Māori is productive (Hale, 1973; Marcus et al., 1995; Yang, 2016; Gorman and Yang, 2019).

The current approach to the Māori passives is constructed mainly following the study by Marcus et al. (1995). In their paper where they advocate for a symbol-based, rule-governed mental processes of regular inflectional morphology as opposed to pattern associator memory hypotheses founded on similarity and frequency that view all linguistic productivity as generalization by similarity, Marcus et al. use two cases of German inflection, German verb conjugation and noun pluralization, to prove that it is not necessarily the overwhelming dominance of the number of words an affix is associated with that gives it the default status contrary to the claims of the connectionist accounts. It is emphasized in the paper that unlike their English counterparts, which constitute the majority, both the German participle *-t* and the German plural *-s* apply to a small number of verbs and nouns respectively, and they behave as defaults in the language as proven by empirical studies. In order to explain the phenomena in which regular inflection takes place as opposed to the majority hypothesis, Marcus et al. find 21 circumstances where the regular English past tense rule applies over irregulars, which would otherwise take place due to lexical analogies, and they claim that given the heterogeneous nature of these combined circumstances, it can only be a symbol such as *N* for "Noun" or *V* for "Verb" that encompass all of these various circumstances and inherent mental rules that operate on such symbols that can explain why an affix becomes the regular, default one, not the similarity between the lexical items.

Regarding the exceptional or "irregular" items whose inflected forms are created in idiosyncratic ways, as in the past tense forms of the strong verbs in English (e.g., *sing-sang* and *go-went*), Marcus et al. claim that they are canonical, uninflected stems committed to memory, and such retrieval of such irregular forms from the memory blocks the application of a mental rule-governed regular affixation such as the English past tense suffix *-ed*. In their discussion of partial structure of irregular morphology, Marcus et al. deny the arbitrary conception of the mental links between the stem and past forms of English irregular verbs based on three observations they make: the wide commonality of phonological material between the verb stems and their irregular past versions (e.g., *ring* and *rang*); the fact that there are far fewer unique changes between stems and their past forms than there are irregular verbs; that is, there are subregularities certain irregular verbs seem to follow (e.g., the alteration between *i* and *a* in pairs like *ring-rang*, *sing-sang*, and *drink-drank*); and finally the hypersimilarities (defined by sharing such common qualities as the same initial and/or final consonants as well as having the same vowels) among the verbs in the same clusters. While irregulars are not the main question examined in the study, Marcus et al. provide a few alternative solutions such as minor rules or the enrichment of the memory to handle this case of English irregular verbs that are neither fully idiosyncratic nor productive. In a similar fashion, Gorman and Yang (2019) mention the unpredictable nature of these irregulars while pointing out that, given the evidence for some kind of a serial processing, irregulars seem to be checked before the regulars (Yang, 2016), and thus, they are in competition with the productive, regular rule since the exceptional cases must be exhausted before the application of a regular inflection is approved by the language user.

Since, by definition, it would rule out the question of predictability of the three Māori passive suffixes examined in this paper, *irregular* naturally would not be the correct term to use for these suffixes; however, the approach of minor rules or the reinforced pattern associativity runs parallel to what this research investigates. In the current study, the main question is whether the Māori passive endings carrying one of the specific thematic consonants under question (i.e., /-mia, -ria,

-hia/) are predictable or not. If the diachronic analysis had found that the thematic consonants to be just results of vertical inheritance from Proto-Oceanic in all the lexical items examined, then the natural conclusion to be drawn would have been to state that there is little chance of predictability.[42] Nonetheless, the innovations found in each class of /-mia, -ria, -hia/ are reasons sufficient enough to implicate predictability. The question is are there any such minor rules or any patterns associated with these thematic consonants that determine the selection?

The feature-based synchronic analyses attempted to answer this question by examining mainly the phonological composition of each verb root associated with these three suffixes; however, no single pattern from the 16 features analyzed was found to be significantly associated with any of the suffixes under consideration, and no rules were detected. Evidently, this is at odds with Parker Jones's suggestion that the Māori passive endings appear to be predictable based on the phonotactic properties of verb roots; therefore, based on the evidence obtained from the analyses conducted in this study, I conclude that the phonological properties of active Māori roots do not exhibit any patterns that determine their corresponding passive shapes, that is, which thematic consonant to appear in their passive forms.

In summary, while the presence of at least one instance of innovation in which a root-final proto-consonant was replaced with either *m*, *r*, or *h* within each suffix hints at a potential recurring pattern discernible to generations of Māori speakers, the evidence based on the feature-based analyses conducted in this study support the long-held assumption of unpredictability underlying the Māori passives. However, this conclusion should remain tentative until all the other Māori passive suffixes are subject to at least the same level of diachronic and synchronic scrutiny that was pursued in this research. Furthermore, psycholinguistic studies to explore how children analyze Māori passives in the process of language acquisition and to understand the analysis that proficient adult users of Māori entertain would also provide invaluable empirical data to this field.

[42] That is, unless Proto-Oceanic had rule-governed processes that determined the selection of root-final consonants; however, besides being unlikely due to etymological reasons predating POc, this question falls outside the scope of the current study.

5 Conclusion

This study is the first step into a comprehensive examination of the Māori passive suffixes via a combination of diachronic and synchronic approaches using corpus data. It is also an example of the potential enhancement of human-led investigations through the integration of computational tools, effectively harmonizing the realms of historical linguistics and corpus linguistics. However, it has its limitations, as mentioned in the previous sections.

One of the primary contributions this study provides the field with is the historical developments that took place introducing *-m-*, *-r-*, and *-h-* to replace proto-sounds in passive forms of certain verbs (i.e., the 12 verbs acquiring these thematic consonants via innovation), which is suggestive of patterns that Māori speakers might have detected in certain verbs associated with either of these thematic consonants in the passive. Nonetheless, in its current state, the study is not able to reveal robust patterns associated with any of the three suffixes under question based on the corpus used. In other words, the outcomes of the feature-based analyses conducted in this study lead me to propose that it is not feasible to anticipate the specific thematic consonant that will appear in the passive form of a Māori word based solely on the phonological attributes of the corresponding active form.

What has been presented in this paper constitutes a relatively small portion (only 114 words have been analyzed) of a larger, ongoing research that attempts to analyze all 889 words present in the corpus. Subsequent phases of this research, marked by both historical, linguistic, and computational refinements, promise to shed a brighter light upon the history of te reo Māori, the probable synchronic analysis employed by humans, and the *black box* that is the neural network systems.

References

Arms, D. G. (1973). Whence the Fijian transitive endings? *Oceanic Linguistics*, 12(1/2):503–558.

Bauer, W. A. (1993). *Māori*. Routledge.

Besnier, N. (1992). Polynesian languages. In Bright, W., editor, *International Encyclopedia of Linguistics*, volume 3, pages 245–251. Oxford University Press.

Biggs, B. (1961). The structure of New Zealand Maaori. *Anthropological Linguistics*, 3(3):1–54.

Biggs, B. (1965). Direct and indirect inheritance in Rotuman. *Lingua*, 14:383–415.

Biggs, B. (1971). The languages of Polynesia. In Bowen, J. D., editor, *Linguistics in Oceania*, pages 466–506. De Gruyter Mouton.

Biggs, B. (1978). The history of Polynesian phonology. In *Second International Conference on Austronesian Linguistics: Proceedings*, volume 2, pages 691–716. Pacific Linguistics C61.

Biggs, B. (1990). *English–Maori, Maori–English Dictionary*. Auckland University Press.

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

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge University Press.

Blust, R. (1976). A third palatal reflex in Polynesian languages. *The Journal of the Polynesian Society*, 85(3):339–358.

Blust, R. A. and Trussel, S. (2020). Austronesian comparative dictionary. `http://www.trussel2.com/ACD`.

Burkhard, W. A. and Keller, R. M. (1973). Some approaches to best-match file searching. *Communications of the ACM*, 16(4):230–236.

Fox, J. (1971). Sister's child as plant: Metaphors in an idiom of consanguinity. In Needham, R., editor, *Rethinking Kinship and Marriage*, pages 219–252. Routledge.

Gatty, R. (2009). *Fijian-English Dictionary: With Notes on Fijian Culture and Natural History*. Southeast Asia Program Publications.

Gorman, K. and Yang, C. (2019). When nobody wins. In Rainer, F., Gardani, F., Dressler, W. U., and Luschützky, H. C., editors, *Competition in Inflection and Word-Formation*, pages 169–193. Springer International Publishing.

Greenhill, S. J. and Clark, R. (2011). POLLEX-Online: The Polynesian lexicon project online. `https://pollex.eva.mpg.de/`.

Hale, K. (1973). Deep-surface canonical disparities in relation to analysis and change: An Australian example. In *Diachronic, areal, and typological Linguistics*, pages 401–458. De Gruyter Mouton.

Hale, K. (1991). Remarks on G. Sanders "Levelling in the history of Polynesian passive formations". *The Journal of the Polynesian Society*, 100(1):99–101.

Harlow, R. (1996). *Māori*. LINCOM Europa.

Harlow, R. (2007). *Maori: A Linguistic Introduction*. Cambridge University Press.

Hohepa, P. W. (1967). *A Profile Generative Grammar of Maori. Indiana University Publications in Anthropology and Linguistics, Memoir 20*. Waverly Press.

Lichtenberk, F. (1978). Thematic consonants in Manam transitive verbs. *Anthropological Linguistics*, 20(5):185–193.

Lynch, J. (2001). Too much to swallow: On terms meaning 'swallow' in Oceanic languages. *Oceanic Linguistics*, 40(2):336–341.

Lynch, J., Ross, M., and Crowley, T. (2011). *The Oceanic Languages*. Routledge.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: The exception that proves the rule. *Cognitive Psychology*, 29(3):189–256.

McCarthy, J. J. (1981). A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry*, 12(3):373–418.

Mills, R. F. (1975). *Proto South Sulawesi and Proto Austronesian Phonology*, volume 1. University of Michigan.

Moorfield, J. C. (1988). *Whanake I Te Kākano*. Longman Paul.

Moorfield, J. C. (n.d.). Te aka Māori dictionary. `https://maoridictionary.co.nz/`.

Māori Language Commission (2012). Te taura whiri i te reo Māori guidelines for Māori language orthography. `https://tetaurawhiri.govt.nz`.

Parker Jones, Ō. (2008). Phonotactic probability and the Māori passive: A computational approach. In *Proceedings of the Tenth Meeting of ACL Special Interest Group on Computational Morphology and Phonology*, pages 39–48. Association for Computational Linguistics.

Pawley, A. (2001). Proto polynesian *-cia. *Issues in Austronesian morphology: A focusschrift for Byron W. Bender*, 519:193–216.

Pratt, G. and Newell, J. E. (1891). *Pratt's Grammar & dictionary of the Samoan language*. Malua Printing Press, 4 edition.

Ross, M., Pawley, A., and Osmond, M., editors (1998). *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*, volume 1: Material culture. Pacific Linguistics C-152.

Ross, M., Pawley, A., and Osmond, M., editors (2007). *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*, volume 2: The physical environment. ANU E Press.

Ross, M., Pawley, A., and Osmond, M., editors (2008). *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*, volume 3: Plants. Pacific Linguistics.

Ross, M., Pawley, A., and Osmond, M., editors (2011). *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*, volume 4: Animals. Pacific Linguistics.

Ross, M., Pawley, A., and Osmond, M., editors (2016). *The Lexicon of Proto Oceanic: The culture and environment of ancestral Oceanic society*, volume 5: People: body and mind. Asia-Pacific Linguistics.

Ryan, P. (1989). *The Revised Dictionary of Modern Māori.* Heinemann Educational.

Ryan, P. (2012). *The Raupō Dictionary of Modern Māori.* Penguin Random House, 4 edition.

Sanders, G. (1990). On the analysis and implications of Maori verb alternations. *Lingua*, 80(2):149–196.

Sanders, G. (1991). Levelling and reanalysis in the history of Polynesian passive formations. *The Journal of the Polynesian Society*, 100(1):71–90.

Tregear, E. (1891). Māori-Polynesian comparative dictionary. `https://nzetc.victoria.ac.nz/tm/scholarly/tei-TreMaor.html`.

Williams, H. W. (1988). A dictionary of the Maori language. `https://nzetc.victoria.ac.nz/tm/scholarly/tei-WillDict.html`.

Yang, C. (2005). On productivity. *Linguistic Variation Yearbook*, 5(1):265–302.

Yang, C. (2016). *The Price of Linguistic Productivity.* The MIT Press.