Phonotactic Probability and the Māori Passive: A Computational Approach

‘Ōiwi Parker Jones
Oxford University Phonetics Laboratory
41 Wellington Square
Oxford, OX1 2JF, UK
oiwi.parkerjones@phon.ox.ac.uk

Abstract

Two analyses of Māori passives and gerunds have been debated in the literature. Both assume that the thematic consonants in these forms are unpredictable. This paper reports on three computational experiments designed to test whether this assumption is sound. The results suggest that thematic consonants are predictable from the phonotactic probabilities of their active counterparts. This study has potential implications for allomorphy in other Polynesian languages. It also exemplifies the benefits of using computational methods in linguistic analyses.

Active	Passive	Gloss
/ɸera/	/ɸerahia/	‘to spread’
/oma/	/omakia/	‘to run’
/inu/	/inumia/	‘to drink’
/eke/	/ekeŋia/	‘to climb’
/tupu/	/tupuria/	‘to grow’
/aɸi/	/aɸitia/	‘to embrace’
/huna/	/hunaia/	‘to conceal’
/kata/	/kataina/	‘to laugh’
/ako/	/akona/	‘to teach’
/heke/	/hekea/	‘to descend’

Table 1: Examples of active and passive verbs in Māori.

1 Introduction

The Māori passive is perhaps the most famous problem in Polynesian linguistics. It has received attention from Williams (1971, first published in 1844), Biggs (1961), Hohepa (1967), Hale (1968; 1973; 1991), Kiparsky (1971), Kaye (1975), Kenstowicz and Kisseberth (1979), McCarthy (1981), Moorfield (1988), Sanders (1990; 1991), Harlow (1991; 2001; 2007), Bauer (1993), Blevins (1994), Kibre (1998), de Lacy (2004), and Boyce (2006). Some representative examples of active and passive verbs are given in Table 1 (Ryan, 1989).

Two types of analysis have been proposed for these data (Hale, 1968). These are known as the ‘morphological’ and ‘phonological’ analyses. For the subset of passives with thematic consonants, the analyses parse the data differently into stems and suffixes. To illustrate this, the examples from Table 1 have been parsed in Table 2 with hyphens inserted between the stems and suffixes. The thematic consonants have also been flagged.

In both types of analysis, the qualities of the thematic consonants are assumed to be unpredictable and are therefore lexicalized. To cite just one example, Blevins writes that “a consonant of *unpredictable* quality appears in the passive and gerundial forms, but this consonant is absent when the verb occurs unsuffixed” (Blevins, 1994, p. 29, my emphasis).

In the phonological analysis, the thematic consonants are lexicalized with the rest of the stem. The active forms are derived by a rule that deletes stem-final consonants. Although less obvious, the morphological analysis also lexicalizes the thematic consonants by allowing stems to be stored with ‘diacritic features’. The reason for the diacritic features

Proceedings of the Tenth Meeting of the ACL Special Interest Group on Computational Morphology and Phonology, pages 39–48,
Columbus, Ohio, USA June 2008.

MOR	PHON	THEME
/ɸera-hia/	/ɸerah-ia/	/h/
/oma-kia/	/omak-ia/	/k/
/inu-mia/	/inum-ia/	/m/
/eke-ŋia/	/ekeŋ-ia/	/ŋ/
/tupu-ria/	/tupur-ia/	/r/
/aɸi-tia/	/aɸit-ia/	/t/
/huna-ia/	/huna-ia/	none
/kata-ina/	/kata-ina/	none
/ako-na/	/ako-na/	none
/heke-a/	/heke-a/	none

Table 2: Morphological analyses (MOR), phonological analyses (PHON), and thematic consonants (THEME).

is to constrain the free combination of stems and suffixes, which, if unconstrained, would over-generate unattested passive forms. As an illustration, if we assume the exhaustive association of /inu/ with the diacritic feature [+m], then the stem would be allowed to combine with /-mia/, but not with any other suffixes. (In short, to store the diacritic feature is to lexicalize the quality of the thematic consonant.) Although lack of a diacritic feature is allowed for stems that take 'default' suffixes (/-tia/, /-ia/, or /-a/, depending on stem's size and composition), this would only be one thematic consonant (out of six) that would not be lexicalized; the phonological analysis could still be seen as lexicalizing the majority of the thematic consonants. Furthermore, a case could be made that the contrastive absence of a [+t] diacritic feature effectively lexicalizes /-tia/, too. Finally, it is worth noting that the purpose of the default suffixes is to provide analyses for previously unseen stems, such as nonce words or borrowings; in other words, the purpose of defaults is not to make /-tia/ non-lexical.

In this paper, I want to question the assumption that thematic consonants are unpredictable in Māori passives. To do so, I will focus on the phonotactic probabilities of active verbs as predictive of their passive and gerundial forms. I implemented the analysis as an artificial neural network, which I describe below. This follows from a rich tradition of using neural networks in phonology and morphology, as exemplified by the English past tense models of Rumelhart and McClelland (1987) and Plunkett and Marchman (1991). Incidentally, I chose neural networks to implement my analysis because of their computational properties, not because of an argument for the biological plausibility of my analysis. I suspect that similar results could have been obtained from another statistical formalism, like the k-nearest neighbor approach of TiMBL (Daelemans and van den Bosch, 2005).

The paper is laid out as follows. The network is described in section 2, the data and experimental methodology are presented in section 3, and the experimental results are reported in section 4. The discussion and conclusion follow in sections 5 and 6, respectively.

2 Network architecture and settings

The network I used in this study was designed to model a function from the representation of an active verb in Māori (alternatively, from a verb stem in the morphological analysis) to a set of output categories corresponding to passive formations (i.e., to a set of passive suffixes in the morphological analysis).

For the simulations in this study, I used a 3-layer feed-forward architecture with 199 input units, 100 hidden units, and 10 output units. The connectivity between adjacent layers was all-to-all. One fully activated bias unit was connected to every unit in the hidden and output layers (to model a thresholding effect and to aid learning). Figure 1 provides a rough blueprint of the network in 'slab' notation.

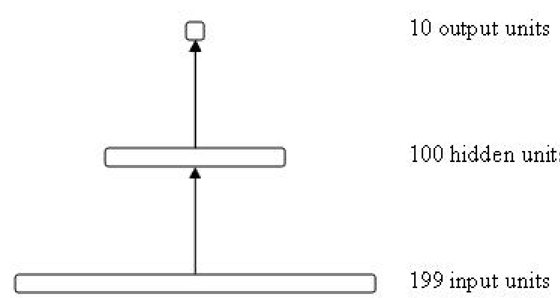

Figure 1: Network architecture; all-to-all connections between units in adjacent layers; bias unit not shown.

To calculate the output or activation of a node i in the network, I used a sigmoid function

$$a_i = \frac{1}{1 + e^{-net_i}}, \tag{1}$$

where e is the exponential and net_i is the net input to node i. As usual, the net input to node i was defined as

$$net_i = \sum_j w_{ij} a_j, \tag{2}$$

where w_{ij} refers to the weights on the connections from nodes j to node i, and where a_j refers to the activations of nodes j (Plunkett and Elman, 1997). Learning was achieved using back-propagation and a learning rate of 0.1 (Werbos, 1974). No momentum was used. Let us turn now to the design of each layer in the network's architecture.

2.1 The input layer

There were 199 input units, where the number of input units was chosen to allow up to 18 segments in the input. Each segment was transformed into an 11-bit vector according to the feature encodings in Table 3. The unaccounted-for unit was used to tell the model if it was learning a passive or a gerund function; it can be thought of as specifying the semantic value PASS or NMLZ.

	vocalic	*long*	*round*	*high*	*low*	*bilabial*	*alveolar*	*velar*	*plosive*	*fricative*	*nasal*
/a/	1	0	0	0	1	0	0	0	0	0	0
/aː/	1	1	0	0	1	0	0	0	0	0	0
/e/	1	0	0	0	0	0	0	0	0	0	0
/eː/	1	1	0	0	0	0	0	0	0	0	0
/i/	1	0	0	1	0	0	0	0	0	0	0
/iː/	1	1	0	1	0	0	0	0	0	0	0
/o/	1	0	1	0	0	0	0	0	0	0	0
/oː/	1	1	1	0	0	0	0	0	0	0	0
/u/	1	0	1	1	0	0	0	0	0	0	0
/uː/	1	1	1	1	0	0	0	0	0	0	0
/p/	0	0	0	0	0	1	0	0	1	0	0
/t/	0	0	0	0	0	0	1	0	1	0	0
/k/	0	0	0	0	0	0	0	1	1	0	0
/ɸ/	0	0	0	0	0	1	0	0	0	1	0
/h/	0	0	0	0	0	0	0	0	0	1	0
/m/	0	0	0	0	0	1	0	0	0	0	1
/n/	0	0	0	0	0	0	1	0	0	0	1
/ŋ/	0	0	0	0	0	0	0	1	0	0	1
/ɾ/	0	0	0	0	0	0	1	0	0	0	0
/w/	0	0	0	0	0	1	0	0	0	0	0
//	0	0	0	0	0	0	0	0	0	0	0

Table 3: Māori phonemes and feature encodings.

I approached the representation of active verbs empirically. Three coding schemes were considered, one of which was segment-based and two of which were syllable-based. Table 4 provides a handful of examples in the segmental coding scheme. Notice that each representation is right-aligned within the matrix and that there are no gaps between the segments. Null phonemes were used to fill the empty cells so that each input vector would be exactly 199 bits long.

		6	5	4	3	2	1
/aː/	↦						aː
/uhi/	↦				u	h	i
/waiho/	↦		w	a	i	h	o
/inoi/	↦			i	n	o	i
/tia/	↦				t	i	a

Table 4: Examples of segmental coding.

For both syllabic coding schemes, I used a 3-cell sequence to represent a CVV syllable template. To illustrate this, the examples from Table 4 have been reanalyzed in Table 5 to be consistent with both syllabic coding schemes.

		Syll			Syll		
		C	V	V	C	V	V
/aː/	↦					aː	
/uhi/	↦		u		h	i	
/waiho/	↦	w	a	i	h	o	
/inoi/	↦		i		n	o	i
/tia/	↦	t	i			a	

Table 5: Examples of syllabic coding.

Within each syllable sequence (Syll) in Table 5, the first position (C) was reserved for an onset, the second position (V) was reserved for the primary vowel, and the third position (V) was reserved for the second vowel of a diphthong. Again, every representation was right-aligned. Any sequence of short vowels in an active verb was treated as a diphthong, unless the vowels were equal in quality or the second vowel was lower than the first. For example, /ei/ and /eo/ would be diphthongs, but /ee/ and /ea/ would be analyzed as hiatus.

The syllabic coding schemes differed in their treatment of a long vowel followed by a short vowel, where the two vowels had non-identical qualities and the second was not lower than the first (i.e., where they would be diphthongs if both were phonemically short). The first coding treated these sequences as diphthongs (Coding 1); the second did not (Coding 2). Table 6 contrasts the two syllabic schemes for the word /taːoro/ 'to break down'. Since de Lacy (2004) advanced the analysis on which I based Coding 2, I shall sometimes distinguish these schemes by referring to Coding 2 as 'the de Lacy analysis'.

	Syll			Syll			Syll		
	C	V	V	C	V	V	C	V	V
Coding 1				t	aː	o	ɾ	o	
Coding 2	t	aː			o		ɾ	o	

Table 6: Two syllabic codings for /taːoro/.

In the section on experiments below, I report on which of these three schemes worked best. For now, my aim has been to motivate the network's input layer. Notice that 199 input units provides space for input representations of up to 6 syllables (6 syllables $\times$ 3 prosodic positions $\times$ 11 features = 198), with room for the semantic unit mentioned above. None of the active verbs in the passive dataset required more than 6 syllables in any of the coding schemes.

2.2 The output layer

Since there were 10 passive categories in my dataset (corresponding to the passive suffixes in the morphological analysis, illustrated in Figure 2), 10 output units were employed in the network. It was considered appropriate to model membership to each category independently, as many verbs show multiple passive forms (as /motu/ 'to separate, wound' does in its passive forms /motu-hia/ and /motu-kia/). The key to reading the model's passive output can be given as the vector [/-hia/, /-kia/, /-mia/, /-ŋia/, /-ɾia/, /-tia/, /-ia/, /-ina/, /-na/, /-a/]. Although the model represents its outputs as bits, they can be interpreted by reference to this key. For example, the passive output for /tapa/ 'to name' should be [1, 0, 0, 0, 0, 0, 1, 1, 0, 0], since Ryan's dictionary attests /tapa-hia/, /tapa-ia/, and /tapa-ina/. Note that these alternative outputs are taken to represent 'free' variation within a single speaker, rather than dialectical variation between speakers.

While the main focus of the model is the Māori passive, the network can also be used to associate active verbs (alternatively, morphological verb stems) with their gerundial forms (i.e., gerund suffixes, in a morphological analysis). Although there are fewer gerund suffixes than passive suffixes, there is a well-known parallel between the existing gerund suffixes and the subset of passive suffixes with thematic consonants. Consider the vector [/-haŋa/, /-kaŋa/, /-maŋa/, /-ŋa/, /-raŋa/, /-taŋa/, /-aŋa/, N/A, N/A, N/A], which can be used as the key for interpreting gerund outputs in the network. Notice that the passive and gerund keys both order the thematic consonants as in the vector [/h/, /k/, /m/, /ŋ/, /ɾ/, /t/, //, N/A, N/A, N/A]. (Here, the null segment // has parallels in both keys.) So, interpretation of the gerundial output can also be performed by lookup. For example, the target output for /ɸiːtiki/ 'to tie up' on the gerund task is [0, 0, 0, 0, 1, 0, 0, 0, 0, 0], since the dataset from Ryan's dictionary attests /ɸiːtiki-raŋa/. Finally, output activations for the last three nodes are undefined in the gerund task. I would have interpreted a significant activation for any of them as a false prediction.

2.3 The hidden layer

In general, too few hidden units do not provide a network with enough computational power to learn a desired function; too many units will result in the network overfitting the data, in which case its ability to generalize will suffer. Given the dimensions of the input and output layers, I was able to estimate the required number of hidden units empirically. Starting with a conservatively small number of hidden units, I trained the network for 100 epochs on 371 patterns in the passive dataset (i.e., approximately 80% of 464 patterns, which did not contain any known loanwords), and then froze the network's weights and tested its predictions on 46 of the withheld patterns (i.e., approximately 10% of the passive dataset), measuring the mean squared error. I repeated this procedure for increasingly populated hidden layers, until a trend emerged suggesting an optimum number of hidden units to minimize the mean squared error on the test set. For this task, 100 hidden units seemed to work well. The results for the estimation of hidden units have been graphed in Figure 2.

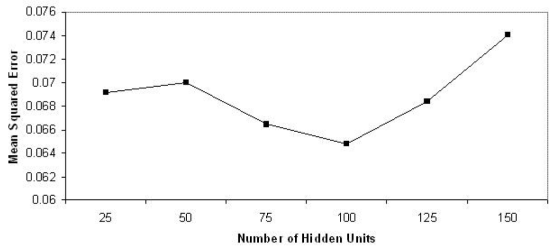

Figure 2: Minimizing error in the network.

3 Methodology

3.1 Data

The passive data in this study were drawn from the Māori-English section of *The Revised Dictionary of Modern Māori* (Ryan, 1989). This provided 476 passive patterns, 12 of which were flagged as English borrowings.

Active	Passive	Gloss
/taraiwa/	/taraiwa-tia/	'drive'
/raka/	/raka-ina/	'lock'
/paera/	/paera-tia/	'boil'
/wepu/	/wepu-a/	'whip'
/perehi/	/perehi-tia/	'press, print'
/paura/	/paura-tia/	'powder'
/ɸaka-hoːnore/	/ɸaka-hoːnore-tia/	'honor'
/pauna/	/pauna-tia/	'to weigh' (< *pound*)
/parau/	/parau-tia/	'plough'
/minita/	/minita-tia/	'minister'
/ɸaka-rapihi/	/ɸaka-rapihi-tia/	'to make rubbish of'
/parai/	/parai-tia/	'fry'

Table 7: 12 English borrowings with their passive forms.

Since I only found two gerund patterns in Ryan's dictionary (viz. /huːpana-taŋa/ and /ɸiːtiki-raŋa/), I also searched the Māori Broadcast Corpus (MBC) for words ending as if they had gerundial suffixes (Boyce, 2006). This turned up 1537 gerund-like tokens, which reduced to 139 gerund-like types.

An overview of the data is provided in Tables 8 and 9. Table 8 shows that 464 passive patterns map to 28 output categories, the most populous of which contains 188 members. In other words, 188 verb stems take the passive suffix /-a/ and no other. By contrast, only one verb stem takes either /-tia/ or /-na/ as its passive suffixes. Similarly, Table 9 shows that 233 gerund-like patterns map to 16 output categories. For example, 120 (presumed) verb stems take the gerund suffix /-taŋa/.

Category	Members	Category	Members
{/-a/}	188	{/-ŋia/, /-a/}	2
{/-tia/}	112	{/-ria/, /-tia/}	2
{/-hia/}	33	{/-hia/, /-ia/, /-ina/}	1
{/-na/}	27	{/-hia/, /-kia/}	1
{/-ŋia/}	19	{/-hia/, /-mia/}	1
{/-ia/}	17	{/-ia/, /-ina/, /-a/}	1
{/-ria/}	16	{/-ina/, /-a/}	1
{/-ina/}	13	{/-ŋia/, /-ia/}	1
{/-kia/}	6	{/-ŋia/, /-ria/}	1
{/-tia/, /-a/}	6	{/-ŋia/, /-tia/}	1
{/-mia/}	4	{/-ŋia/, /-tia/, /-a/}	1
{/-ia/, /-a/}	3	{/-ria/, /-ia/}	1
{/-hia/, /-a/}	2	{/-tia/, /-ina/}	1
{/-hia/, /-tia/}	2	{/-tia/, /-na/}	1

Table 8: 464 passive patterns map to 28 output categories.

Category	Members	Category	Members
{/-taŋa/}	120	{/-ŋa/, /-taŋa/}	2
{/-haŋa/}	35	{/-raŋa/, /-taŋa/}	2
{/-ŋa/}	21	{/-haŋa/, /-aŋa/}	1
{/-aŋa/}	20	{/-haŋa/, /-kaŋa/}	1
{/-raŋa/}	16	{/-haŋa/, /-maŋa/}	1
{/-kaŋa/}	6	{/-ŋa/, /-aŋa/}	1
{/-maŋa/}	3	{/-ŋa/, /-raŋa/}	1
{/-haŋa/, /-taŋa/}	2	{/-raŋa/, /-aŋa/}	1

Table 9: 233 gerund-like patterns map to 16 categories.

3.2 Procedure

For the various experiments conducted, different subsets of the collected corpus were employed. In general, a sub-corpus was selected and then (randomly) split into training and testing sets. The size of these sets differed for the different experiments, since different amounts of relevant data were available. In every case, the stimuli consisted of input vectors and their corresponding target vectors.

Before training, the weights in the network were initialized using a random seed. Stimuli from the training set were then presented to the network randomly without replacement, so that each stimulus was seen once per epoch. Training lasted for 100 epochs. The weights were then frozen before each of the training stimuli were presented to the network again in order to validate the network's performance. The validated network was then presented with the test stimuli and its predictions were compared with the activations of the targets. In every experiment, the networks were run 5 times using 5 different random seeds to initialize the weights. I did this so that the results would be a little more robust. Performance was evaluated by taking the average percent correct over the 5 runs and variability was measured by calculating the standard deviation of the 5 runs.

Outputs were evaluated by fist rounding their activations to 0 or 1, before comparing them to the target patterns. It should be noted that this is a relatively liberal measure of the network's performance, given such alternatives as measuring the distance from output to target using a deviation < 0.1. Nonetheless, evaluation by rounding was justified on grounds that the only meaningful output patterns for the network were the non-negative integers 0 and 1.

I used chance as the null hypothesis when it was required for comparison with the network's performance, as chance represents the baseline for unpredictability. The chance of guessing the output activations correctly was calculated by assuming binary activations for the outputs (which is fair given the rounding of network outputs to 0 and 1). For 10 output nodes, $2^{10} = 1024$ possible guesses were possible. In such cases, the probability of guessing the correct output pattern for any stimulus was calculated as $\frac{1}{1024} \times 100 = 0.1\%$. Except where otherwise noted, the chance of guessing the right output patterns for n stimuli was calculated as $\frac{n}{1024} \times 100$.

In some cases, I used other calculations as comparisons against the network's performance. I will introduce these where applicable.

4 Experimental results

4.1 Segmental and syllabic representations

As mentioned above, the question of input representation is an empirical one. I introduced three coding schemes in section 2.1 (a segmental one and two syllabic ones). In order to compare the schemes' ability to predict the passive forms (including the thematic consonants), a sub-corpus of 464 patterns was selected (i.e., the full set of 476 passives found in Ryan's dictionary minus the 12 loanwords). Since these stimuli had already been randomly split into 80%-10%-10% subsets to estimate the number of hidden units in the network, I started by reusing this split. The 10% used as a test set for the hidden units task were then lumped back into the training set, resulting in a random 90%-10% split (i.e., 418 training patterns and 46 test patterns). Each coding scheme was then applied to the same training and test sets, and the network was run as described in the methods section.

The results are summarized numerically in Table 10 and graphically in Figure 3. They suggest that either syllabic coding scheme is better than the segmental one, and that the de Lacy analysis is better than the alternative syllabic coding scheme (i.e., Coding 2 beats Coding 1). This suggests that it is better to represent a long vowel followed by a short vowel in Māori as two syllables.

Coding Scheme	% Correct	Standard deviation
Segmental	90.43	2.92
Syllable 1	91.74	2.83
Syllable 2	93.91	1.82

Table 10: Representation experiment results, rounded to the nearest hundredth.

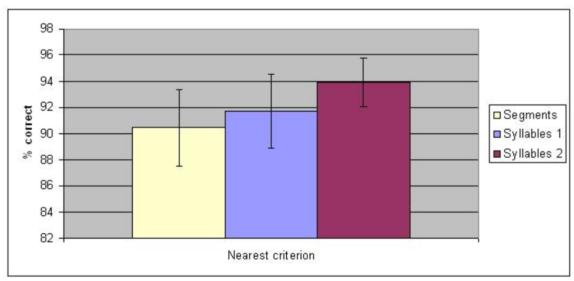

Figure 3: Test results for different representations of the stems, 5 runs apiece; error bars show standard deviations.

The results also challenge the assumption that thematic consonants are strongly unpredictable (i.e., governed by chance). I note that 30 of the test patterns did not take a suffix with a thematic consonant, while 15 did. So, of the 15 relevant test cases, the null hypothesis would have guessed 23.44% correct (i.e., $\frac{15}{2^6} \times 100$); I adjusted the calculation of the null hypothesis here to reflect the focus on just 6 of the 10 output patterns (i.e., the ones with thematic consonants). Without adjusting the calculation, the null hypothesis would have done much worse (cf. $\frac{15}{2^{10}} \times 100$). By contrast, the network predicts 46.67% correct (i.e., $\frac{7}{15} \times 100$), since it correctly predicted 7 out of the 15 patterns. So, the network correctly predicted 23.23% more of the thematic consonant patterns than chance. This suggests that lexicalization is not the only way to address thematic consonants in Māori. Since the problem can be specified in terms of active and passive verbs (rather than in terms of stems and suffixes), this also suggests that the Māori passive need not be framed in terms of the 'morphological' and 'phonological' analyses of Hale (1968).

The model also does well predicting the passive form of a verb in general. Note that the null hypothesis would only get 4.49% of the 46 test stimuli correct (i.e., $\frac{46}{1024} \times 100$). Using the de Lacy analysis, the network correctly predicted 93.91% of the 46 test stimuli, which is a massive difference of 89.42%. Moreover, the network also outperforms a 'majority choice' strategy, whereby all verb stems take the most frequent output category (i.e., {/-a/}). Majority choice correctly predicts 40.52% of the 464 passive patterns (i.e., $\frac{188}{464}$), which is 53.39% less than the network's coverage.

4.2 Gerunds

To test beyond the passive dataset, two sets of gerunds were considered. The idea was to see if training a network on a dataset of passives would be able to predict the suffix patterns of gerunds.

By training the network on the entire passive dataset 5 different times, and then testing each one on the 2 gerunds found in Ryan's dictionary, the network predicted the 100% of the results correct for all 5 runs. (For 2 test items, the null hypothesis would have only guessed $\frac{2}{1024} \times 100 = 0.2\%$ correct.)

Using the same training set, but testing the network on the 139 gerund-like words in the MBC, the network correctly predicted an average of 90.36% correct (with a standard deviation of 0.82). For 139 test patterns, the null hypothesis would only predict 13.57% correct. In both cases, the model does noticeably better than chance.

4.3 Loanwords

When new verbs enter the Māori language, speakers generalize their knowledge about the passive endings to them. How well does the network do at modeling this ability? 12 loanwords were flagged in the passive dataset. By training the network on the 464 non-loanword passives and then testing it on the 12 loanwords, the network got 100% correct for all 5 runs. Chance would only predict 1.17% of this test set correctly (i.e., $\frac{12}{1024} \times 100$).

The network also outperforms majority choice on this task, since majority choice for the 12 loanwords predicts 83.33% (i.e., $\frac{10}{12}$). (The most common output category for the 12 loanwords is {/-tia/}.)

In this case, however, there is probably a more charitable null hypothesis against which to compare the network's performance. I refer to the default analysis, where verbs take /-tia/, /-ia/, or /-a/. On this analysis, any stem containing more than two morae takes /-tia/ as its default, any stem containing fewer than three morae and ending with /a/ takes /-ia/ as its default, and any other stem (i.e., one containing fewer than three morae and not ending with /a/) takes /-a/. (Incidentally, one single-mora stem exists in my database; it is /ko/ 'to dig', which takes /-ia/.)

So, how does the default analysis compare with the network's analysis? Of the 12 loanwords in the passive database, the default analysis gets 91.67% correct (i.e., $\frac{11}{12}$). Again, the network gets 100% correct every time, for 5 runs. Interestingly, all but one of the 12 loanwords takes /-tia/, /-ia/, or /-a/. Furthermore, the exception, /ɾaka-ina/ (< English *lock*), would appear to be a systematic hole in the default analysis, since analogous examples exist, such as /tia-ina/ (< English *steer*) (Paul de Lacy, personal communication). Since both of these stems consist of fewer than three morae and end with /a/, the default analysis incorrectly predicts that their passive forms should be */ɾaka-ia/ and */tia-ia/, respectively. In other words, while the network only outperformed the default analysis by one example from the dataset, that one example would appear to be representative of a class of stems that the default analysis necessarily gets wrong, but which the neural network analysis could possibly get right. However, since the network needs to be run in order to see what it actually predicts, additional work would be needed to address this further.

5 Discussion

Thus far, the model has been evaluated on its performance. But while a model that performs well on a task is valuable in its own right, one would also like to understand how the model is succeeding. Neu-

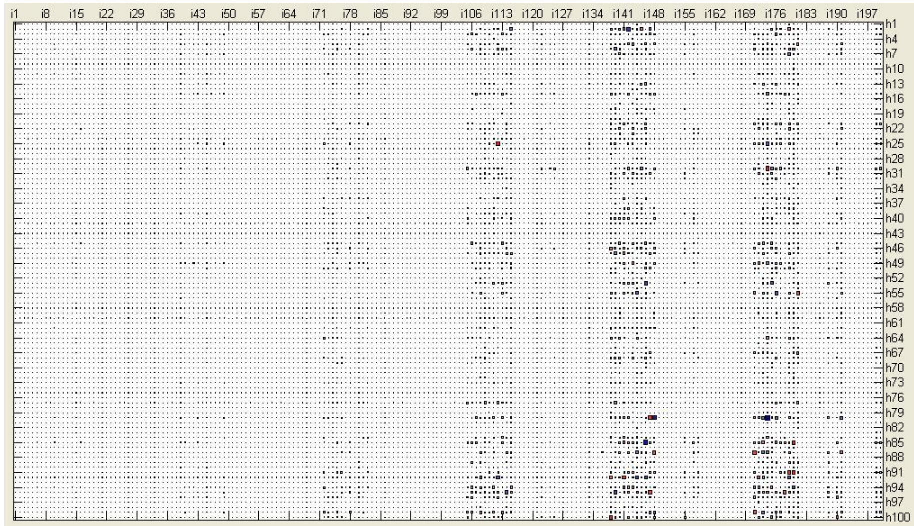

Figure 4: Hinton diagram for a typical weight matrix from input units (x-axis) to hidden units (y-axis).

ral network simulations are sometimes critiqued for being black box solutions, where a problem can be solved but the solution cannot be understood. Therefore, in this discussion section, I would like to begin to address the question of what properties in stem representations are responsible for the prediction of their output categories.

A few relevant sub-regularities have already been reported in the Māori literature, which are worth review. Citing Moorfield (1988, p. 66), Harlow reports that /-ina/ only occurs after words ending with /a/, while /-mia/ only occurs after words ending with /o/ and /u/ (i.e., the $[+round]$ vowels); his examples, with the stem-final vowels underlined, are /hua̲-ina/ 'be named', /aroha̲-ina/ 'be loved', /ɸaka-ŋaro̲-mia/ 'be made to disappear', and /inu̲-mia/ 'be drunk' (Harlow, 2007, p. 117). Although these observations provide necessary but not sufficient conditions for inferring a passive suffix from a verb stem, they exemplify the type of pattern that one might like to find. The problem is to find better patterns in the verb stems.

For ideas of what to investigate, we might look inside one of the trained networks. Figure 4 illustrates the weights from input units to hidden units in a network trained on the Māori passive data using the de Lacy coding scheme (from section 4.1). Notice the dark vertical bands around inputs 176, 141, and 113 (there are fainter bands around input 78 and 43, and faint and narrow bands around input 190 and 155). These bands represent stronger weights (both positive and negative) between the two layers in the network. In order to understand the network's performance, we might ask what these bands represent. Given that the syllabic coding scheme organizes the segments into vowels and consonants in a similar pattern, one hypothesis would be that the vertical bands represent vowels in the input; a complementary hypothesis would hold that they represent consonants in the input. While this is a rather crude distinction to make, it begins to narrow down the hypothesis space.

To test such hypotheses, we may use 'degraded' inputs. For example, to test one hypothesis, one might replace all consonants in the input representations with null phonemes; to test the other hypothesis, one might replace all vowels in the input representations with null phonemes. An example of these degraded input representations is given in Table 11 for the word /ɸera/ 'to spread'.

In a preliminary study (running the network just once), I found that the model with vowel-only input outperformed the model with consonant-only input by a slight margin. Further investigation is surely needed. But the methodological use of degraded inputs provides a way to probe which parts of these representations contribute most to the model's performance.

Additional studies might use degraded inputs with only the final syllables represented compared with

	Syll			Syll		
	C	V	V	C	V	V
All segments	ɸ	e		ɾ	a	
No consonants		e			a	
No vowels	ɸ			ɾ		

Table 11: Three representations of /ɸeɾa/ 'to spread'. The top one is an uncorrupted input using the de Lacy syllabic coding. The bottom two are degraded in different ways: one has no consonants, the other has no vowels.

ones in which only the penultimate or antipenultimate syllables are represented; they might even narrow down which phonetic features predict which passive and gerundial categories.

6 Conclusion

The work described here is clearly preliminary with respect to the problem of predicting passives and gerunds in Māori. But the experimental results are suggestive, especially as they challenge the long-held assumption that thematic consonants cannot be predicted. This research has implications for future investigations of allomorphy in Māori and other Polynesian languages, since Polynesian allomorphy has never before been explored using phonotactic probabilities (at least to the best of my knowledge).

In general, a computational approach makes it much easier to run complex statistical analyses over large datasets (compared with manual analyses using paper and pen). The success of utilizing statistics in this study exemplify the benefits of using computational methods in linguistics.

Acknowledgments

For various feedback, I am grateful to acknowledge John Coleman, Julien Mayor, Sharon Goldwater, Paul de Lacy, Doug Ball, Mary Boyce, and three anonymous reviewers. This research was supported by a Lamakū Scholarship. All imperfections are my own.

References

Winifred A. Bauer. 1993. *Maori*. Routledge, London and New York.

Bruce Biggs. 1961. The structure of New Zealand Maaori. *Anthropological Linguistics*, 3:1–53.

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

Mary Teresa Boyce. 2006. *A Corpus of Modern Spoken Māori*. Ph.D. thesis, Victoria University of Wellington.

Walter Daelemans and Antal van den Bosch. 2005. *Memory-Based Language Processing*. Cambridge University Press, Cambridge.

Paul de Lacy. 2004. Maximal words and the Maori passive. In John McCarthy, editor, *Optimality Theory in Phonology: A Reader*, pages 495–512. Blackwell, Oxford.

Kenneth Hale. 1968. Review of Hohepa (1967). *Journal of the Polynesian Society*, 77:83–99.

Kenneth Hale. 1973. Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas Sebeok, editor, *Current Trends in Linguistics 11*, pages 401–458. The Hague, Mouton.

Kenneth Hale. 1991. Remarks on G. Sanders' "Levelling in the history of Polynesian passive formations. *Journal of the Polynesian Society*, 100:99–101.

Ray Harlow. 1991. Consonant dissimilation in Maori. In *Currents in Pacific Linguistics: Papers in Austronesian Languages and Ethnolinguistics in honour of George W. Grace*, pages 117–128. Australian National Universiy, Canberra.

Ray Harlow. 2001. *A Māori Reference Grammar*. Pearson Education, Auckland.

Ray Harlow. 2007. *Māori: A Linguistic Introduction*. Cambridge University Press, Cambridge.

Patrick W. Hohepa. 1967. *A Profile Generative Grammar of Maori*. Indiana University Press, Bloomington, Indiana.

J. Kaye. 1975. A functional explanation for rule ordering in phonology. In R. Grossman, L.J. San, and T. Vance, editors, *Papers from the Parasession on Functionalism*. Chicago Linguistics Society, Chicago.

Michael Kenstowicz and Charles Kisseberth. 1979. *Generative Phonology: Description and Theory*. Academic Press, San Diego.

Nicholas Kibre. 1998. Formal property inheritance and consonant/zero alternations in Maori verbs. Rutgers Optimality Archive 285.

Paul Kiparsky. 1971. Historical linguistics. In William Dingwall, editor, *A Survey of Linguistic Science*, pages 577–649. University of Maryland Press, College Park, Maryland.

John J. McCarthy. 1981. The role of the evaluation metric in the acquisition of morphology. In C.L. Baker and John J. McCarthy, editors, *The Logical Problem of Language Acquisition*, pages 218–248. MIT Press, Cambridge, Massachusetts.

John C. Moorfield. 1988. *Whanake I Te Kākano*. Longman Paul, Auckland.

Kim Plunkett and Jeffrey L. Elman. 1997. *Exercises in Rethinking Innateness: A Handbook for Connectionist Simulations*. MIT Press, Cambridge, Massachusetts.

Kim Plunkett and Virginia Marchman. 1991. U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition*, 38:43–102.

David E. Rumelhart and James L. McClelland. 1987. Learning the past tenses of English verbs: Implicit rules or parallel distributed processing. In Brian MacWhinney, editor, *Mechanisms of Language Acquisition*, pages 194–248. Erlbaum, Mahwah, New Jersey.

P.M. Ryan. 1989. *The Revised Dictionary of Modern Māori*. Heinemann Education, Auckland, third edition.

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

G. Sanders. 1991. Levelling and reanalysis in the history of Polynesian passive formations. *Journal of the Polynesian Society*, 100:71–91.

Paul J. Werbos. 1974. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Ph.D. thesis, Harvard University.

H.W. Williams. 1971. *A Dictionary of the Maori Language*. Government Printer, Wellington, seventh edition. Originally published in 1844.