Department of Electrical, Computer, and Software Engineering

Part IV Research Project

Final Report

Te reo Māori Pronunciation Dictionary Building Tool

James Coppard

Brendon Joe

Supervisors:

Catherine Watson

Peter Keegan

14/10/2022

Declaration of Originality

This report is my own unaided work and was not copied from nor written in collaboration with any other person.

Name: James Coppard

ABSTRACT: Despite being one of the official languages of New Zealand, te reo Māori is an endangered language, and has been the subject of revitalization efforts by the government. As society shifts towards reliance on modern technology, it is important that lower resourced languages such as Māori do not get left behind. However, building tools for lower resourced languages is inherently difficult. This paper follows the development of a GUI pronunciation building tool for te reo Māori which is an extension of a previously developed command line tool. The GUI combines all features of the previous tools into a portable web application that can be accessed anywhere without the need to setup an environment which significantly lowers the barrier to entry. A new pronunciation generation technique has been implemented with multi-level stress application for long and compound Māori words and achieves an accuracy of 95%. Moreover, our results show that our tool increases the efficiency of pronunciation transcription and verification.

Table of Contents

List of Tables

List of Figures

1. Introduction

Recent modeling shows that te reo Māori could be headed towards extinction if nothing more is done to encourage people to use it [1]. Therefore, as the world continues to move into a digital era, it is vitally important that te reo Māori users and learners can use their language when interacting with modern technology. Speech synthesis tools such as text-to-speech and speech-to-text are widely available for commonly used languages, however, creating these systems for low-resource languages is an uphill battle as speech processing requires a large amount of specific data [2]. These tools use pronunciation dictionaries in order to translate between speech and text. A pronunciation dictionary is a mapping of words in a language to their respective pronunciations, using a machine translatable format such as IPA or SAMPA. In 2020, a te reo Māori pronunciation dictionary was created [3] and we have been provided with the tools used to create it. This study aims to produce a single tool that ties these tools together in a way that is friendly to non-technical users thereby increasing the efficiency by which the dictionary can be added to and edited. We also aim to improve the stress markup of multi-morphemic and compound words in the current dictionary.

2. Literature Review

This literature review will investigate the current knowledge of the building of pronunciation dictionaries for low-resource languages. There is a lack of specific information regarding te reo Māori pronunciation dictionaries, so this study will analyze literature from other languages that are similarly resourced. We also need to investigate the literature on stress rules in te reo Māori and any measures we can use to determine if the changes we make to the current stress markup are an improvement.

2.0. Te Reo Māori Linguistics

This study aims to produce a product that improves the phonetic accuracy of the current te reo Māori pronunciation dictionary, so it is important to first understand the basics of its linguistics. This

knowledge could also be relevant to use as teaching material for users of our product so that they can understand the lexical markup of the language.

2.0.1. Characters

Te reo Māori was originally only a spoken language and was translated into writing when British missionaries arrived in the early 1800s. As such, the language uses the Roman alphabet and consists of five vowels, but only ten consonants. There is also the concept of double vowel sounds, which are dictated by macrons. Two of the consonants are consonant pairs: *ng* (pronounced *ŋ*) and *wh* (pronounced *f*).

Consonants	h	k	m	n	ng	p	r	t	w	wh
Vowels	a	e	i	o	u	ā	ē	ī	ō	ū

Table 1. Māori Characters

2.0.2. Structure

Syllables usually take the form of (C)V(V), and there is also the concept of *mora* which take the shape (C)V [4]. All syllables, therefore, end in a vowel, making it an open syllable language. Any words that are borrowed from English also follow this rule.

Pairs of vowels produce diphthongs, which are often counted as separate vowels in addition to the monophthong vowels [4]. It is also important to make a distinction between vowel sequences and diphthongs; some vowel pairs such as *ua* are pronounced as a sequence of vowels, while *au* is pronounced as a "glide" between the two vowels which is also referred to as diphthongization [5].

2.0.3. Stress Rules

All Māori words and most particles contain a stressed syllable [5]. Te reo Māori follows clear rules for applying stress.

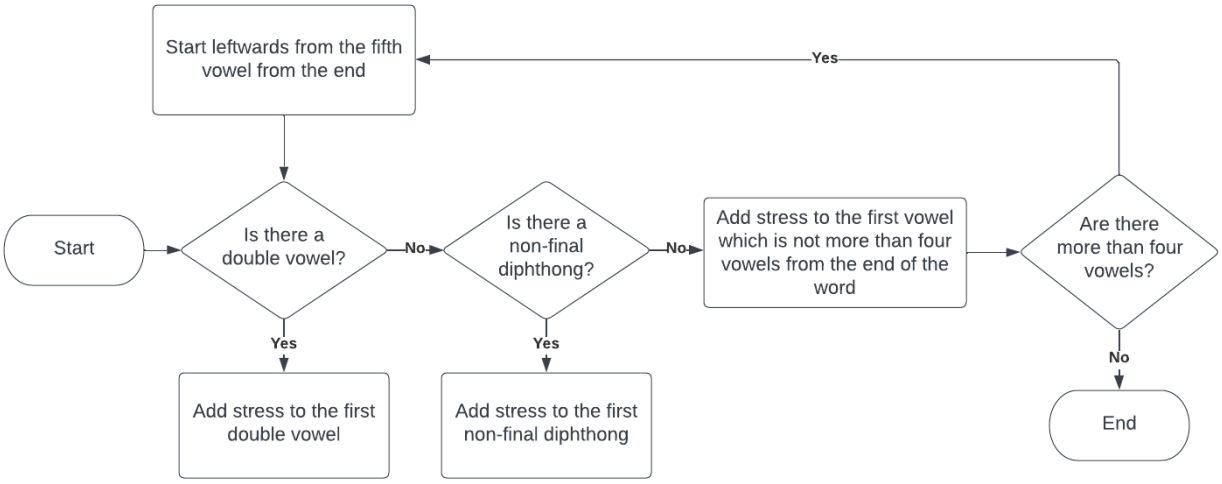

Figure 1. Biggs' Stress Rules Flowchart

2.0.4. Regional Dialects

One requirement for the final product is to support different dialects in text-to-speech. There are four common alternative te reo Māori dialects spoken in New Zealand, each being associated with a different region and iwi [4]. Each dialect includes a clean mapping of one consonant or consonant pair to another, shown in figure 2. There are also non-trivial differences between these dialects such as changes in grammar, lexicon, and phonemes, however, a pronunciation dictionary will only be concerned with grapheme changes.

Dialect	Original Maori Consonant	Replacement Consonant
Ngāpuhi	wh	h
Tūhoe	ng	n
Ngāi Tahu (Kai Tahu)	ng	k
Whanganui	wh	w' (glottal stop)

Table 2. Differences between common dialects

2.1. Current Te Reo Māori Pronunciation Dictionary Building Tools

A te reo Māori pronunciation dictionary and its respective toolkit was created in 2020 (under the same supervision as this study) [10]. The dictionary achieved a lexicon size of over 10,000 words with a reported accuracy of 95%. The tools provided were written in python and include functions for converting word inputs into machine readable pronunciations (IPA and SAMPA). Currently, the dictionary is stored in dislocated text files with no centralized database available. A list of 959 words with associated pronunciations transcribed manually are available to verify word conversion accuracy. A speech model was built using a custom recorded speech corpus of 1030 sentences which was used to create a MaryTTS synthesizer. This allows machine readable Māori pronunciation inputs to be transformed into synthesized waveforms. A user-friendly end to end system was also created, allowing a user to input a Māori word and receive its pronunciation as an audio waveform. As this is an emerging field, there are no further specific resources available for te reo Māori pronunciation dictionaries.

2.2. Existing Solutions for Improving Pronunciation Dictionary Tooling

Languages evolve over time through phonetic, lexical, spelling, semantic and syntactic changes, which means that any given pronunciation dictionary will have to adapt to language changes. Specifically, te reo Māori has been heavily influenced by the presence of the much more popular New Zealand English which has brought phonetic changes to monophthongs and has also introduced many borrowed words [15]. As previously stated, the current tooling was built for the full creation of a pronunciation dictionary, not its editing. If we want to keep up with changes in a language, then we need to adapt this suite of tooling to allow adding new words and editing existing words.

2.2.1. LEXiTRON-Pro Editor

The Thai language pronunciation dictionary faced very similar issues; they had a suite of 3 dislocated tools with 5 steps being required to make changes to their dictionary [16]. They devised a single tool, named the *LEXiTRON-Pro Editor* to combine their toolkit into a single tool, with a single step needed to modify the dictionary. The system works as follows: First, a user submits a word or list of words to the interface via a text file, and the software strips all punctuation, numbers, and whitespace, just leaving words. Then, the words are displayed in a table, selecting a word causes editing tools to appear. Automatically generated pronunciations are presented to the user who can select them if they are correct. If the user wants to make a change to this pronunciation, then they can use the user interface to denote the correct pronunciation which is submitted. Any word that is submitted is added to the dictionary immediately. This tool was shown to work tremendously well, reducing time consumption for pronunciation generation from 5 minutes to 45 seconds, and increasing accuracy from 18.1% to 73.6%.

3. Māori Pronunciation Dictionary Building Tool

3.0. Database

A key part of our solution lies with our centralized pronunciation database, which we have developed using MongoDB.

3.0.1. Word Data Schema

Previously, phonetic data was encoded in a series of text files (e.g., syllables.txt, which contains the syllable breaks for each word in the dictionary). We have created a new schema for word data storage in JSON format, which allows it to be machine readable and more flexible to change in the future.

Figure 2. Word Schema

We have taken advantage of this new flexibility by adding more data fields to each word object in the database. The source field stores a string value of where the word originated from. When a user adds a word, this would be that user's name, if a link was scraped, this would be that link. Approved shows if the pronunciation has been verified by an approved user.

Pronunciation is stored as an array of syllables, with each syllable containing a list of phonemes and an associated stress. Stress is realised as a number value representing the level of stress. In Fig.1, the first syllable has a primary stress, and the second has none. Storing the phones and stress in this way allows us to trivially convert the database to the festival pronunciation format, which was previously complex.

3.1. TTS

In order to play pronunciation audio waveforms in our client, we have created a simple HTTP TTS microservice that uses festival to do the conversion. Festival is only available for Linux, so we deployed this as an external service to our main server in order to have access to the waveforms on any platform while developing.

An issue arose with the conversion of arbitrary pronunciations as our Māori voice pack did not support the 'SayPhones' command. To get around this, a pronunciation could be added as a lexical entry, tied to a specific string x, which could then be converted using the command 'SayText x'.

3.2. Pronunciation Generation

The old system for pronunciation generation did not consider multiple stress levels and was tied closely to the legacy text file format that the dictionary was stored in. This system was rewritten to utilize the new word object format. To make sure that the new system was correct, we compared the outputs of both generation techniques to make sure that they matched. During this testing only one stress level was considered in the new generation.

To add multiple stress levels, Biggs' fourth stress rule [9] was followed by splitting each word into groups of four mora (if applicable) and then applying stress on these groups individually in reverse order. There is some contention over the number of possible stress levels in Māori pronunciations; our system is flexible in the number of stress levels that can be applied.

3.2.1. Display

To display the pronunciation of each word in the user interface we show the syllable boundaries and syllable stress levels.

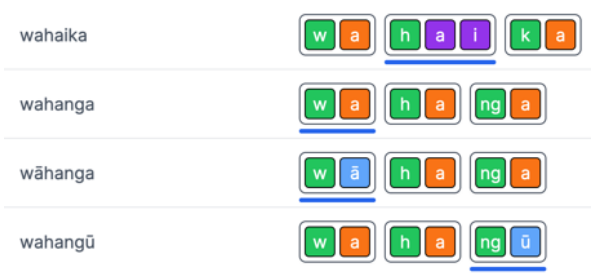

Figure 3. Pronunciation View

Colours are used to highlight the differences between monophthongs, diphthongs, and long vowels.

Stress is shown below each syllable as a coloured cell.

3.3. Pronunciation Editor

A drag-and-drop based pronunciation editor has been implemented in order to allow full granular control over individual word pronunciations.

Figure 4. Editing pronunciation

The example in Fig.2 shows the process of splitting a diphthong into monophthongs. When the editor is shown, a list of possible pronunciation mutations is calculated using a list of valid diphthongs [10]. If moving an individual letter is found to be possible, the user will be allowed

One point to note is that long vowels cannot be split or combined using this editor as it would result in the word's orthography changing. A separate spelling altering input is given where necessary.

The following rules are given for reference and are all derived from the open syllable nature of Māori:

- In order to not form a closed syllable:
    - A consonant cannot be moved
    - A single vowel (monophthong) cannot be moved if the current syllable has a consonant
    - A single vowel (monophthong) cannot be moved to the next syllable if the next syllable has a consonant
- Remaining valid moves:
    - A single vowel (monophthong) can be moved to a previous syllable if a valid diphthong would be formed
    - A vowel in a diphthong can be moved to a new previous syllable if it is the first vowel in the diphthong
    - A vowel in a diphthong can be moved to a next syllable if that syllable has no consonant, a valid diphthong is formed, and it is the final vowel in the diphthong
    - A vowel in a diphthong can be moved to a previous syllable if the current syllable does not contain a consonant and it is the first vowel in the diphthong

3.4. Bulk Editing

The editing workflow has been implemented as described in 2.4.1. with a few extra features.

The user is first presented with a text input for bulk text and a text input for overriding the source. Any text can be entered to be processed, including links. By default, the source of all words will be attributed to the current user, and the contents of any links to the link. The user can override the source if they want to attribute bulk text to some external source such as a spreadsheet.

3.4.1. Processing

First, all links are stripped from the text and all text from the linked page will be added to the user inputted text (if applicable). Then, processing is done to remove all non-alphabetic characters including extra whitespaces and newlines and the words are lowercased and placed into a set to ensure no repetitions. The result is a list of words which are each checked against a simple Māori word filter and then passed back to the user.

The filter removes words that:

- Contain characters not found in the Māori alphabet
- Do not follow an open syllable structure
- Contain a mismatching number of ‘ng’ and ‘g’ characters (a ‘g’ can only be present in Māori text if it is paired with a ‘n’)

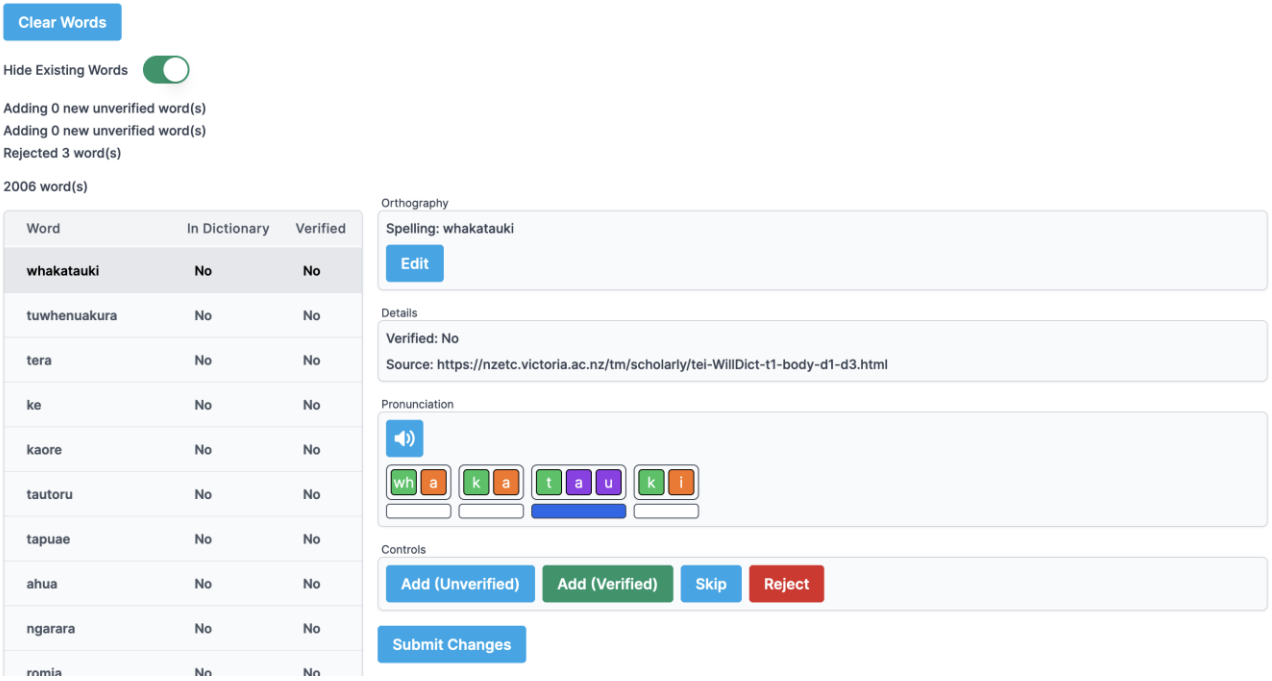

Figure 5. User interface for mass word verification

Inside the editor, the user is presented with a table which contains all processed words from the input. By default, all words that already exist in the dictionary are hidden, but the user may decide to show these so they can be edited. The highlighted word in the table is the word that is currently being checked and can be changed by either clicking on another word in the table or by finishing the check for this word.

If the user would like to alter the orthography (e.g., changing a single vowel to a double vowel) of the current word, they can use the edit button which opens a prompt to change the text.

The user can then look at and listen to the generated pronunciation in the pronunciation section. If they find that the pronunciation is wrong, they can edit the stress and the syllable boundaries as described in section 3.6.

Finally, the word may be added, skipped, or rejected. Words that are skipped or rejected will not be added to the dictionary.

At any point the user may submit the changes they have made to the dictionary which will also remove those words from the current edit table as a convenience.

3.5. GUI Features

3.5.1. Browse

The browse page allows users to search through the words in the dictionary in the form of a table. Each word is shown with its spelling and pronunciation. There are also a series of quick actions next to each word (depending on the permission level of the logged in user). The search bar allows for prefix, contents, and suffix searching.

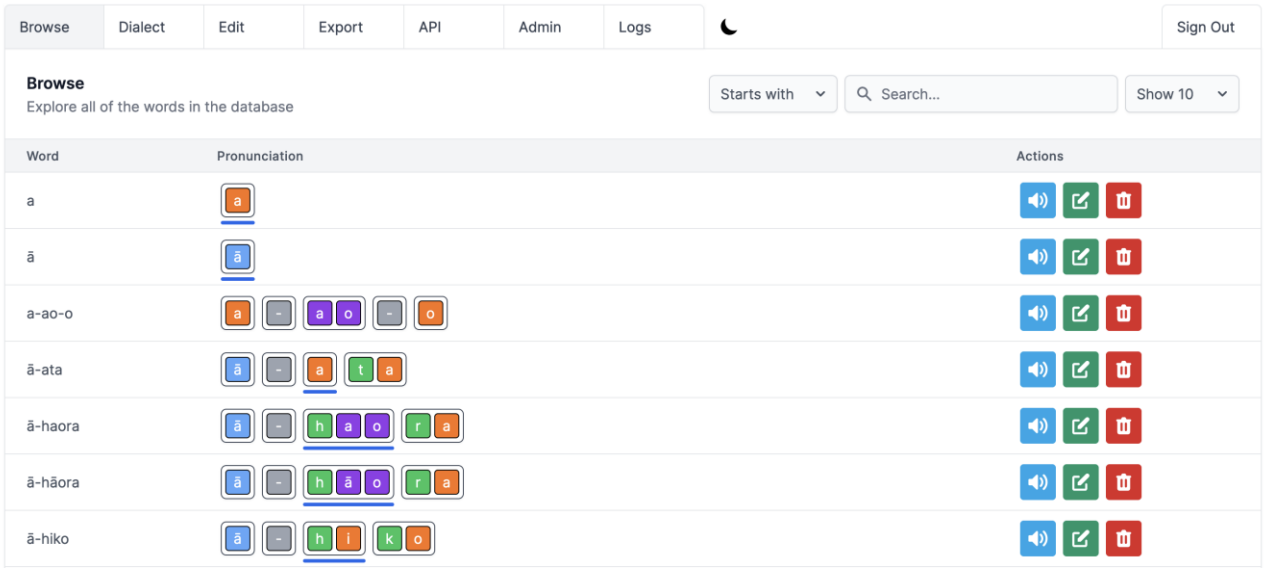

Figure 6. User interface for browsing words in the dictionary

3.5.2. API

Documentation of our externally accessible endpoints are shown here with all details about parameters, requests, and responses. Each user has an associated token which can be exchanged for a temporary access token which can then be used to access the given API endpoints.

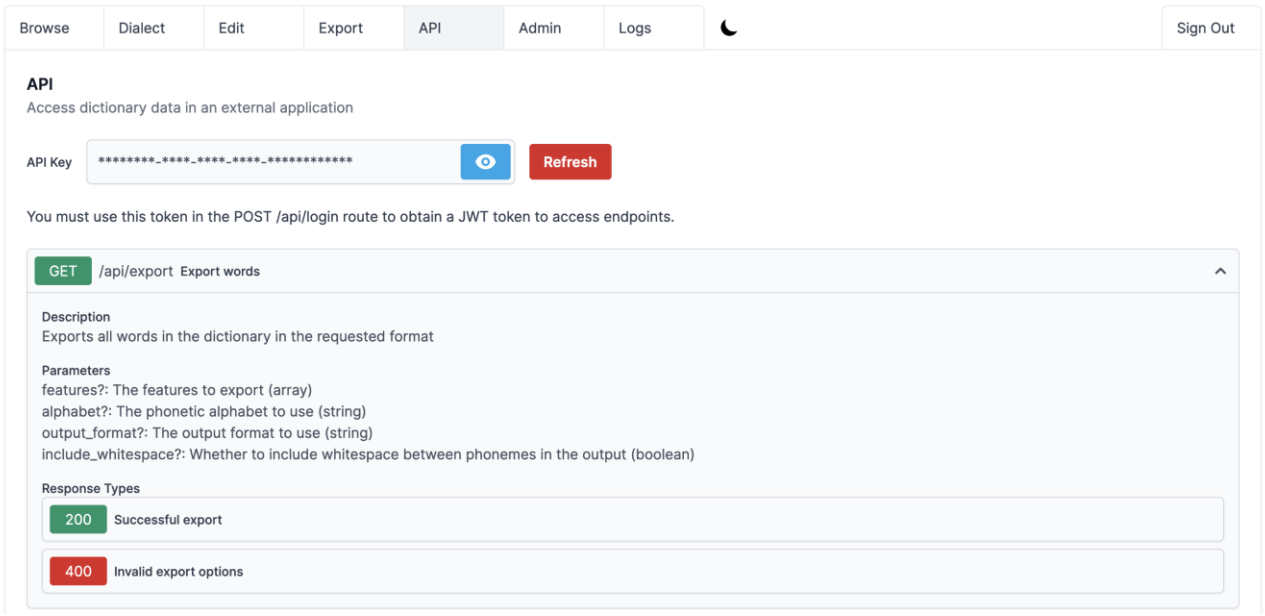

Figure 7. User interface for API documentation

3.5.3. Logs

All additions, edits, and removals are shown on the logs list and are associated with a user and time. The logs may be queried by word, user, and time range.

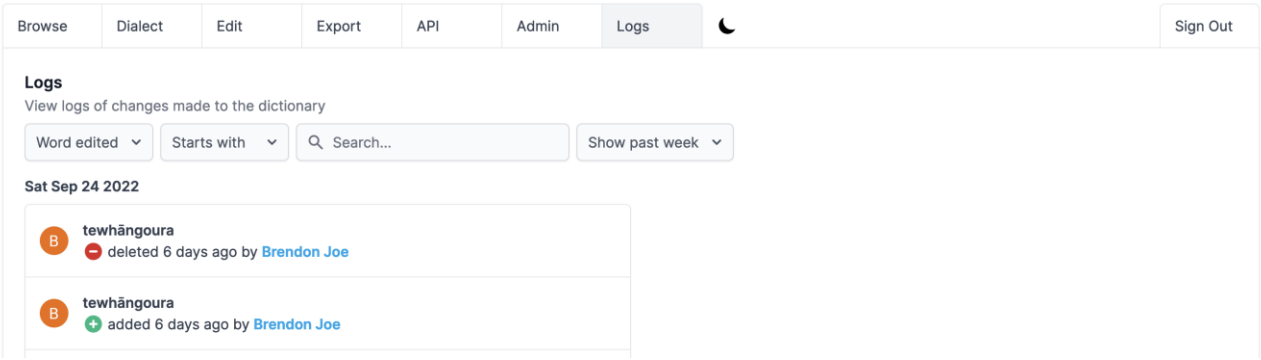

Figure 8. User interface for logged changes

3.5.4. Dialects

To support the various dialects found across New Zealand, the user is given an option to choose which dialect they would like to view the dictionary in. When the user selects a non-standard dialect, the given phonetic swaps are executed whenever a request is made for a pronunciation.

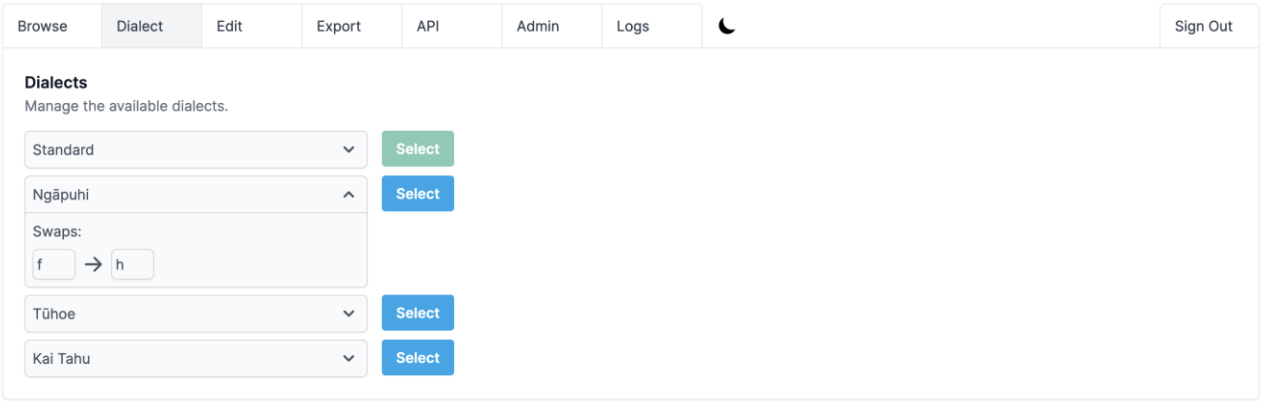

Figure 9. User interface for selecting a dialect

3.5.5. Export

To provide interoperability with text to speech systems, we provide an export system where the user can select from a list of file formats, phonetic formats, and dialects. The application currently supports MaryTTS, Festival, and JSON formats. A live preview of the export is also shown to the user.

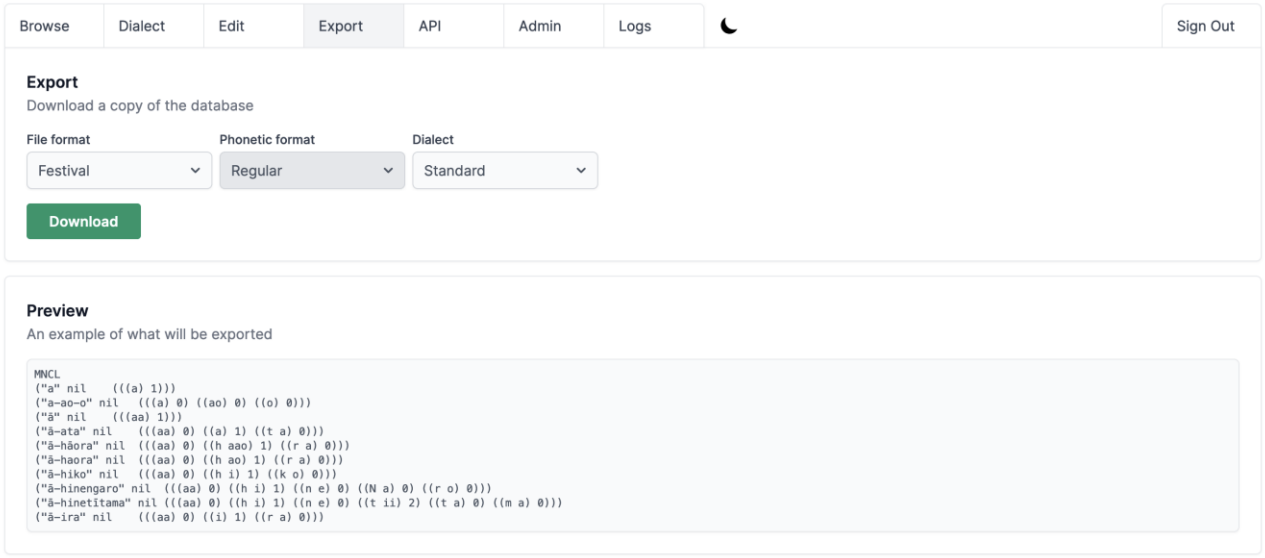

Figure 10. User interface for exporting dictionary data

3.6. Security

Given that the application is publicly facing, it is important that our endpoints are secure against actors who have not been given explicit permission to access or modify the data. All privileged requests require that an authorization header is attached with a valid JSON Web Token (JWT) that was issued by the server. JWTs are encrypted using a secret token which is not externally accessible, ensuring that the token can only be modified by the server.

Each user has an associated permission level dictating which endpoints they can access (see Table 1.), and which can only be set by a user with administrator access.

	None	View	View + Edit	Admin
Get Own User Info	Yes	Yes	Yes	Yes
Get Own API Key	Yes	Yes	Yes	Yes
Browse Words		Yes	Yes	Yes
Export Data		Yes	Yes	Yes

Browse Dialects		Yes	Yes	Yes
Add Words			Yes	Yes
Edit/Verify Words			Yes	Yes
Add User				Yes
Change User Permission				Yes
Remove User				Yes
View Logs				Yes

Table 3. User role permissions matrix

4. Results

4.0. Accuracy

A list of 957 manually transcribed pronunciations were available from the previous research group. It is important to note that only the primary stress of each word was considered in these transcriptions. The accuracy was measured at 95.4%, which is the same accuracy achieved by the previous researchers. This similarity was expected as we did not add any new generation functionality for a single primary stress point.

4.1. User Testing

User testing was conducted on select individuals who had an interest in Māori linguistics to evaluate if our application was intuitive to use and learn. Each user was given a series of tasks to complete inside the application without assistance. Our results show that the application was very user friendly and the users reported that they thought it would be easy to learn for someone who was interested or involved with Māori linguistics.

4.2. Input Processing

Through manual testing we calculated that our text input processing removed 97% of non-Māori words from the contents of a webpage. Additionally, we did not find any case where a valid Māori word was removed in this processing.

5. Discussion

5.0. User Testing

Through our user study we were able to highlight some key areas of improvement in the application. Many users did not discover our macron editor scheme, which let the user type a macron letter using a normal keyboard by adding a colon ‘:’ character after the vowel. When asked to try to type a macron, some users who had seen this notation before were able to complete the task, while others did not discover it themselves. This could be fixed by adding a tooltip or a help page somewhere in the app.

In one extreme case, a participant was running an adblocker script on their web browser which had blocked our sign in button from rendering. A remedy for this scenario was to include a message on the home page telling the user to either use a different browser, or to disable the script. Other than this case, the application was easily run by the testers without any setup required, which is a great leap from the existing tooling which required a large developer environment and technical knowledge.

Using the pronunciation was a slight learning curve for the users at the beginning, except for one user, who was able to pick up the drag and drop functionality very quickly without any prompting. After explaining the tool to the rest of the testers, all of them were able to make changes to a word’s pronunciation. The nature of the editor also meant that it was not possible to produce an invalid

pronunciation (meaning that a syllable could not be made closed, and two vowels could not form an invalid diphthong), which resulted in no room for error by the user.

Overall, the user testing results are very encouraging as everyone who was tested was able to quickly learn the different tools available in the application without much assistance.

6. Conclusion

We have succeeded in the future work goals set out by previous research through implementing a GUI, individual pronunciation editing, and multi-stress application. Our simple interface and centralized server enable linguists anywhere in the world to contribute to a unified Māori pronunciation dictionary. Through our verification tools, users can produce valuable data for future use in machine learning driven approaches to Māori pronunciation generation.

7. Suggestions for Future Work

There were a few features that we saw as nice to have's that we did not have the time to implement that could be completed as future work:

- Multiple pronunciations – There are some cases where a word's pronunciation could be interpreted in multiple ways. This feature would be integrated in the pronunciation editor of the GUI to allow a user to enter multiple pronunciations.
- Advanced affix support – In Māori, there are many types of affixes that directly affect the pronunciation of a word. In our project, we only considered a few suffixes and prefixes; it would be useful to research and find more of these types of affixes to add, as well as finding more types of affixes and showing them inside the UI.

I would also suggest to add automated tests for the pronunciation generation scripts against a list of

manually transcribed single and multi-level stress word pronunciations to make sure that changes made in the future do not result in dropped accuracy.

Acknowledgements

The student would like to thank the supervisors Catherine Watson and Peter Keegan for providing the support and guidelines to make this project successful.

References

[1] T. Barrett-Walker, M. J. Plank, R. Ka'ai-Mahuta, D. Hikuroa and A. James, "Kia kaua te reo e rite ki te moa, ka ngaro: do not let the language suffer the same fate as the moa," *Journal of the Royal Society Interface,* vol. 17, no. 162, 2020.

[2] J. Xu, X. Tan, Y. Ren, T. Qin, J. Li, S. Zhao and T. Y. Liu, "LRSpeech: Extremely Low-Resource Speech Synthesis and Recognition," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* vol. 11, no. 20, pp. 2802-2812, 2020.

[3] J. James, I. Shields, R. Berriman, P. J. Keegan and C. I. Watson, "Developing resources for te reo Māori text to speech synthesis system," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),* vol. 12284 LNAI, pp. 294-302, 2020.

[4] R. Harlow, Māori: A Linguistic Introduction, Cambridge: Cambridge University Press, 2007.

[5] R. Harlow, A Māori reference grammar, Wellington: Huia, 2015.

[6] B. Biggs, Let's Learn Māori (Revised Edition), Auckland: Auckland University Press, 2000.

[7] D. M. Eberhard, G. F. Simons and C. D. Fennig, "Ethnologue: Languages of the World. Twenty-fifth edition.," 2022. [Online]. Available: https://www.ethnologue.com/. [Accessed 11 4 2022].

[8] T. Schlippe, S. Ochs and T. Schultz, "Web-based tools and methods for rapid pronunciation dictionary creation," *Speech Communication,* vol. 56, no. 1, pp. 101-118, 2014.

[9] P. Godard, G. Adda, M. Adda-Decker, J. Benjumea, L. Besacier, J. Cooper-Leavitt, G.-N. Kouarata, L. Lamel, H. Maynard, M. Mueller, A. Rialland, S. Stueker, F. Yvon and M. Zanon-Bolto, "A Very Low Resource Language Speech Corpus for Computational Language Documentation Experiments," in *European Language Resources Association (ELRA)*, Miyazaki, Japan, 2018.

[10] J. James, I. Shields, R. Berriman, P. Keegan and C. Watson, "Developing Resources for Te Reo Māori Text To Speech Synthesis System," in *Text, Speech, and Dialogue: 23rd International Conference, TSD 2020, Brno, Czech Republic, September 8-11, 2020, Proceedings*, Brno, Czech Republic, Springer, 2020, pp. 294-302.

[11] D. Marelie and E. Barnard, "The efficient generation of pronunciation dictionaries: human factors during bootstrapping," in *Interspeech 2004*, Jeju Island, Korea, 2004.

[12] O. Adams, A. Makarucha, G. Neubig, S. Bird and T. Cohn, "Cross-Lingual Word Embeddings for Low-Resource Language Modeling," in *Proceedings of the 15th Conference of the {E}uropean Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Valencia, Spain, 2017.

[13] T. Kocmi and O. Bojar, "Trivial Transfer Learning for Low-Resource Neural Machine Translation," in *Proceedings of the Third Conference on Machine Translation: Research Papers*, Brussels, Belgium, 2018.

[14] M. d. Korte, K. Jaebok and K. Esther, "Efficient neural speech synthesis for low-resource languages through multilingual modeling," in *Interspeech 2020*, Shanghai, China, 2020.

[15] C. Watson, M. A. Maclagan, J. King, R. Harlow and P. J. Keegan, "Sound change in Māori and the influence of New Zealand English," *Journal of the International Phonetic Association*, vol. 46, no. 2, pp. 185-218, 2016.

[16] S. Klaithin, P. Chootrakool and K. Kosawat, "LEXiTRON-Pro Editor: an Integrated Tool for developing Thai Pronunciation Dictionary," in *Proceedings of the International Multiconference on Computer Science and Information Technology*, Wisła, 2010.

[17] M. Davel and E. Barnard, "Bootstrapping pronunciation dictionaries: practical issues," in *Interspecch 2005*, Lisbon, Portugal, 2005.

[18] J. Kominek and A. W. Black, "Learning Pronunciation Dictionaries: Language Complexity and Word Selection Strategies," in *Proceedings of the Human Language Technology Conference of the {NAACL}, Main Conference*, New York City, USA, 2006.

[19] J. King, "Maori: Revitalization of an Endangered Language," in *The Oxford Handbook of Endangered Languages*, Oxford Handbooks, 2018, pp. 591-612.