

20 years of reflections using technology to compile Māori language dictionaries

Karaitiana Taiuru

Introduction

Personal reflections of compiling multiple digital and Information Technology genre based te reo Māori word lists/dictionaries, over the past 20 years. In addition, numerous te reo Māori software, has given me a great insight into how technology can assist te reo Māori revitalisation/ retention and how technology could be used to create a new comprehensive and authoritative Māori dictionary. It is important to clarify that compiling dictionaries does not require language expertise, rather it merely requires collating the works of language experts. In particular, those works that were sparsely published and hard to access, thus applying technical skill sets to bring the works together.

Personal reflections

Presently, there are between 60 and 180 te reo Māori dictionaries available in academic libraries (many of which have previous editions) to cater to the estimated 50,000 to 125,000 te reo Māori speakers.¹ In addition to this, there is a plethora of Māori lesson books and grammar books continually being published. The ease and use of technology is attributed to these numerous dictionaries. Surprisingly, social issues, and not technical issues, have created barriers, preventing collaborative projects among all te reo Māori creators of terminology to create a comprehensive dictionary, similar to what we would expect from an Oxford or Collins English language dictionary.

New terminology that is created in te reo Māori has traditionally been kept private and had restrictions applied to it. In 2002, Te Taura Whiri created and coined new terminology that greatly enhanced te reo Māori revitalisation by making te reo Māori relevant in a modern and evolving society. Though this took place within the commission, the public did not have access to the database of words until they were published in a commercial dictionary, Te Matatiki. Likewise, the large volumes of new terminology that was created for the information technology sector, including by Huawei smart phone makers and up until recently the 2Degrees corpus, are kept private and not shared in public.

It has become usual practice for translators to protect their creations, especially in their industry where automation is more and more common, thus removing the need for many human translators. This exclusivity makes it impossible to learn new words and to look up new words. It also restricts rangatiratanga of our language, our taonga by allowing commercial entities to withhold our taonga from the public. The commercial argument is that the entity invests a large sum of money to create the words and seeks to protect their investment. The flip side to the argument is that if more people could access and learn the terminology, then the number of users would increase, generating demand and profit.

A student of the English language requires one dictionary and maybe a thesaurus. A student of te reo Māori requires multiple dictionaries by different vendors. Many

Māori dictionaries are genre specific such as modern Māori, classic, loan words and other specialised topics such as Health, ICT and plant names. One possible reason for this silo approach to te reo Māori terminology development and dissemination is that there is no collaborative entity with an interest in te reo Māori. If there was, the entity could merge all of the published dictionaries together, and access the plethora of specialised word lists that are created by organisations and individuals, to build one large database with free access on the Internet. This could also be replicated to create one authoritative paper based te reo Māori dictionary. The nearest resource to this was the CD-Rom *Te Reo Tupu* Māori dictionary published in 1997. The CD-Rom resource was a compilation of multiple dictionaries that could be searched simultaneously. Seeking cooperation and agreement from the dictionary owners was time intensive and extremely challenging. The royalty costs to use each dictionary was also prohibitive.

The act of commercialising te reo Māori is holding all speakers of te reo Māori and the tax payer to ransom for what is our birth right. Our language is a recognised taonga under Te Tiriti o Waitangi. The trend in the late 1990's and early 2000's was to commercialise Māori language via paid software that would spell check in te reo Māori. *Te Reo Tupu* evolved into a paid subscription online te reo Māori dictionary endorsed by the kaitiaki of te reo Māori at Te Taura Whiri. A user will pay from \$960 to almost \$10,000.00 plus GST per annum to access the dictionary, which is simply a collection of other works that can largely be searched, separately, online. This in stark contrast an original online te reo Māori dictionary that is free to use at <http://www.maoridictionary.co.nz>.

Te reo Māori spell checkers only required a text based file of te reo Māori words to be installed into a specific folder for it to work. Despite this, the commercialisation of spell checkers was becoming dominant. In 1998, I developed a unique Ngāi Tahu spell checker called *Te Aua Kupu* that would spell check any standard te reo Māori word that replaced a ng with a k. It also replaced ng with an underlined k. This was the first free Māori spell checker produced. Then in 2001, I developed a free standard te reo Māori spell checker called *Te Ngutu Kura* which matured into three versions. These three iterations were, and remain, the most comprehensive spell checker on the market. It grew to accommodate most software editing applications. There were no commercial restrictions and no one to justify the time to create and enhance it. Hence, why it was so widely used. Today, popular Windows based word processing software has te reo Māori built in as a default, removing much of the need for such add-ons.

Compiling multiple te reo Māori word lists and dictionaries over a 20 year period, to produce and maintain the Māori spell checkers, provided a database of te reo Māori with over 480,000 unique entries. The words were quickly extracted out of the database for use in *Te Ngutu Kura*. Many of the words were tagged by genre in the database. This allowed the ability to quickly extract datasets and create specialised te reo Māori dictionaries for genres such as personal names. If an authoritative entity was established, the database could then be used to create a large authoritative Māori Language Dictionary similar to an English Language Dictionary.

While creating free te reo Māori resources, I identified the need for a license that considered Traditional Knowledge/Indigenous ownership. No one person could own te reo Māori or a word unless a word is TradeMarked as was the case with the word 'Moana' in Germany. The solution was to take a freeware license and modify it to acknowledge that words cannot be owned and, if shared, an acknowledgement should be made. The Freeware community did not agree with this at the time and one academic stated that it was unethical to make such a statement.

Eventually the Creative Commons license was offered in New Zealand, which technically did not recognise traditional knowledge, but it offered enough protection with an Indigenous perspective to be useful. It was important for Creative Commons Aotearoa to have a Māori voice as non-Māori were making decisions that would impact upon Māori. At this time, I took up the opportunity to become a Creative Commons advocate. As a Creative Commons Advisory Board member, I implemented a strong recommendation for a Māori role on all future advisory boards.

The Creative Commons Aotearoa license was only in English, despite the name inferring it was bilingual and bicultural. It was my opinion at the time that there was a trend of te reo Māori only resources, both digital and hard copy, being produced and it would not be appealing to use a Creative Commons License in English when the rest of the resource was in te reo Māori. As an advisory member, I advocated for the Creative Commons Licence to be translated into te reo Māori. This was completed by Ian Cormack. The te reo Māori version is one of 15 licenses to be fully translated with another 16 languages yet to be completed.² A Traditional knowledge license was discussed as an add-on to the Creative Commons Aotearoa License. The project never eventuated in New Zealand due to a number of reasons. It is still a project that is urgently required, at both a national and international level, to promote and encourage more Indigenous resource development in order to have legal protection.

Conclusion

Commercialisation, gatekeepers of te reo Māori/knowledge and a fear of technology has seen te reo Māori development slow in the digital area. The lack of legal protections and hurdles for te reo Māori to be recognised in the technology area, by corporate service and resource providers, has also hindered this process. A major shift in attitude is required that allows non-Māori language experts to collate and produce digital products from established resources, which can be freely distributed for the benefit of all people to use the taonga Māori that we call te reo Māori. If government agencies, who use tax payer money to create and distribute te reo Māori resources, offer those resources freely to the community with licenses such as Creative Commons, it would likely see a flourish of innovative and creative resources that promote and make te reo Māori cool in the digital area.

Publications list

A list of te reo Māori publications created by the author:

1. A Dictionary of Māori and Social Media terms, English – Māori. Edition 3 (2016).
2. Word list and analysis of te reo Moriori (2016).
3. Baby names: Christian, Mormon and non religious Māori first names and their equivalent English name (2016).
4. Pahuwera Hapū Names. Data Set. (2017).
5. Ngā Puhi hapū names. Data Set. (2016).
6. Iwi Names. Data Set. (2016).
7. Place names of New Zealand with a Māori name. Data Set. (2016).
8. Street names of New Zealand with a Māori. Data Set. (2016).
9. Marae Names and locations. Data Set. (2015).
10. Diceware password security list in te reo Māori. Author Te Taipo using a word list from Te Ngutu Kura (2015).
11. Māori predicative Txt for Android: Adaptxt Keyboard (2014).
12. New Zealand School Names with a Māori translation of the primary name or a Māori as the primary name. Data Set. (2013).
13. Tertiary Institutions with bilingual Māori name. Data Set. (2013).
14. Māori corpus of unique words. Data Set. (2012).
15. New Zealand Government names with Māori equivalent. Data Set. (2012).
16. Dictionary of Māori Computer Related Terms Edition 2. ISBN: 0-9582621-1-X (2006).
17. Māori txtn glossary for mobile phones (2011).
18. Dictionary of Māori Computer Related Terms Edition 1. ISBN: 0-473-09356-1. (2003).
19. Te Ngutu Kura versions 1. (2001), 2. (2006), 3. (2011).
20. Open Office Māori Language Extension (2011).
21. Macron Converter: A browser based (on-line and off-line) and a Microsoft Office tool that converted double vowels (with the exception of genuine double vowel Māori words to Umlauts. (2003).
22. Unicode Macron Converter (2003).
23. Macron to double vowel converter (2001).
24. Te Reo Māori date for Office (2001).
25. Macron Tool Bar for Office (2001).
26. Te Aua Kupu Ngai Tahu spell checker (1999).
27. Moana Kupu version 1. (1999).
28. Te Reo Tupu Electronic Māori Dictionary (1998).

Endnotes

¹ http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/nz-social-indicators/Home/Culture%20and%20identity/Māori-lang-speakers.aspx

² https://wiki.creativecommons.org/wiki/Legal_Tools_Translation