

TE TAURA WHIRI  
I TE REO MĀORI

MĀORI LANGUAGE COMMISSION



# International good practice for developing and disseminating lexicon

A Literature Review

Sarah Brown

Te Pūtahi Rangahau

Te Taura Whiri i te Reo Māori

June 2022



## Key Objective and Scope

The key objective of this project is to produce a literature review of international good practice literature around lexicon development and dissemination where the information has been sourced from online databases as well as internally available documents.

The review will focus on literature from the 1980s through to present, based on the Indigenous languages of Wales, Hawai'i, Israel, Australia, Spain, France (specifically Basque) and Aotearoa.

The literature will be used to provide Te Taura Whiri i te Reo Māori with an insight into international good practice for lexicon development and dissemination. The examples given in the literature review will then provide Te Taura Whiri i te Reo Māori with ideas on how to improve their lexical practices.

# 01

What does international good practice around lexicon development and dissemination look like?

# 02

How can examples of international good practice around lexicon development and dissemination help Te Taura Whiri i te Reo Māori improve its lexical practices?

**These Key Questions will be the focus of this literature review;**

- › What is a lexicon?
- › According to Singleton (2016), lexicon 'is the term used by linguists to refer to those aspects of a language which relate to words,' otherwise expressed as the 'vocabulary' of a language.





Ježek (2016) describes lexicon in a deeper sense as ‘the set of words of a language,’ organised in word groupings whose members share similarities from the point of view of their form and/or their meaning. The groupings are usually based on meaning associations, and groups of words with similar syntactic behaviour, for example nouns, verbs, or adjectives.

The words within word groupings share similarities from the point of view of their form and/or their meaning from. The following are examples from ‘The lexical work of Te Taura Whiri’ (n.d.):

### kiri meaning ‘skin’

#### Established word meaning:

kiriwai ‘inner skin’ (wai ‘moist’)

kirikā ‘fever’ (kā ‘take fire, burn’)

#### Commission (Te Taura Whiri i te Reo Māori) word meaning:

kirihuna ‘camouflage’ (huna ‘concealed, hidden’)

kiritai ‘wetsuit’ (tai ‘of the sea, water’)

kirikamo ‘conjunctiva’ (kamo ‘eye’)

Example of a word grouping whose members share similar syntactic behaviour (using the suffix -nga to turn a verb into a noun) from ‘The lexical work of Te Taura Whiri’ (n.d.):

#### Long-established word form:

hokinga ‘return’ hoki ‘to return’ + -nga

whakaaturanga ‘show, exhibition’ whakaatu ‘show’ + -ranga

tangihanga ‘(period/act of) grieving’ tangi ‘weep over, mourn’ +-hanga

#### Commission (Te Taura Whiri i te Reo Māori) word form:

ranunga ‘mixture’ ranu ‘mix’

whāinga ‘objective, aim’ whai ‘pursue’

ngahunga ‘focus’ ngahu ‘look intently’

### ▶ What is lexicon development?

Lexicon development incorporates both lexicography and lexicology.

### ▶ Lexicography

According to Ježek (2016), lexicography can be described as “compiling dictionaries for human users”. Its main goal is to identify word properties and the most effective ways to disseminate words according to specific criteria such as the type of dictionary, media type, the intended user group, etc.

An exemplar product of lexicography is “Te reo pāngarau: a Māori language dictionary of mathematics,” (Ministry of Education, 2004; updated 2011). This dictionary includes many excellent examples of te reo Māori maths terms being used in sentences along with good, detailed illustrations. The reo Māori mathematics lexicon was disseminated by means of a dictionary. Where its intended audience were Māori-medium education students and their teachers, learning mathematics through the medium of Māori (Neologisms Briefing, 2014).

### ▶ Lexicology

Ježek (2016) explains that lexicology is the study of the lexicon of a language. The main goal of lexicology is to identify word properties and how they relate to one another. Lexicology highlights that the construction of and the semantics (meanings) of words are highly structured.

Lexicology can be seen in The Lexical Work of Te Taura Whiri i te Reo Māori (n.d.), where already formed words are used to create new words.

For example, with the use of the word ‘kore’. It can be combined with a noun or verb to indicate an absence or lack of whatever is denoted by the term it is combined with, as the suffix –‘less and prefix ;’non’- do in English. For instance, the meaning of ngoikore, ‘weak, listless’, stems from the meaning of ‘ngoi’ ‘energy, strength’ and the suffix – ‘kore’ ‘without, lacking’. Similarly, ‘mutukore’ or ‘mutunga kore’ (mutu ‘end, cease’) is ‘endless, infinite’.

By using ‘kore’ in this way, the Commission has been able to develop the following new vocabulary items:

#### New word gloss form

huakore ‘futile’ hua ‘benefit, outcome (positive)’ + kore> ‘without positive outcome’

hōnea kore ‘the black hole’ hōnea ‘escape’ + kore ‘lack of, no’> ‘(place) of no

## ▶ What is lexicon dissemination?

Lexicon dissemination is the process of how lexicon is dispersed, by what means it is dispersed (resources) and to whom the lexicon is intended for (audience).

There has been a push to disseminate lexicon through mediums other than hard-copy dictionaries through the likes of Wordstream and their web-based version of the CD-ROM Te Reo Tupu termed Wakareo-a-Ipurangi or simply Wakareo as well as the He Pātaka Kupu online database for reo Māori lexicon by Te Taura Whiri i te Reo Māori which in monolingual, provided only in Te Reo Māori (Neologisms Briefings 2014).

## ▶ International good practice on lexicon development and dissemination

The literature review will summarise international literature and give examples of good practice around the development and dissemination of lexicon.

The information presented, draws on international good practice in lexicon development and dissemination, to provide insight to Te Taura Whiri i te Reo Māori in how it can continue to improve its lexical practices.



## Wales – Purism

The Celtic Languages (2009) states that Gwalarn<sup>1</sup> was launched in 1925, where writers devoted themselves to producing literature using the new language. As a result, the structure of the Celtic-Welsh language was modified in 1941 so its dialect could be distinguished from other languages of the time. The goal when modifying the language was to achieve a purer 'Celticity' in syntax and lexicon, standardising and establishing it as a national language.

Further into the 20th Century the purist lexicon coined in the Gwalarn literature was prominent. According to Hawke (2018), "the promotion of puristic terms can indeed alter the course of a lesser-used language's development and increase the confidence of its speakers and their appreciation of the language."

An example of a Welsh, purist term would be "milfeddyg (literally 'animal-doctor')" instead of using the English words "veterinarian or veterinary surgeon (which are based on the Latin word for cattle), (Hawke, 2018).

<sup>1</sup> Gwalarn is a literary review that was founded in 1925 by young activists involved in the revival of the Breton language (including Welsh). The two main topics of their struggle were 1. to gain official recognition of the language, and 2. the establishment of a unitary writing system (Denez, G., & Hupel, E, 2009).



# Hawai'i – The Hawai'ian Language Lexicon Committee

The University of Hawai'i, Hilo has their own College of Hawai'ian Linguistics, Ka Haka 'Ula O Ke'elikōlani. Within this sits the Hawaiian Language Center, the Hale Kuamo'o which supports and encourages the expansion of the Hawaiian language as a medium of communication. And within the Hale Kuamo'o is the Hawai'ian language lexicon committee, Kōmike Hua'ōlelo established in 1987.

The Kōmike Hua'ōlelo wrote the dictionary Māmaka Kaiao, publicly disseminating new words that have been added to their lexicon. Māmaka Kaiao also presents a detailed breakdown of the committee's guidelines for lexicon development and dissemination. The Kōmike (committee) provides a space for the public to submit word suggestions and word requests on the University of Hawai'i, Hilo website (2022).

## The 10 guidelines are summarised below:

1. Make minor changes to a word which already appears in the dictionary. The most common changes are to either insert or delete a kahakō (*macron*), or to join or separate parts of a word or term, 'hāpaina' (*carrier*) and 'kāka'ikahi' (*few*).
2. Record a word which is used by native speakers but is not found in the dictionary, or one which appears in the dictionary but is used by native speakers with a meaning which is different from that listed in the dictionary, e.g. 'kāka'ahi' (*deal, as cards*), and; 'alo'ahia' (*stress*) have been used by native speakers but are not found in the dictionary.
3. Use reduplication of an existing word in order to alter or extend the meaning, e.g. 'ūlialia' (*coincidence*) from 'ulia', 'hohoki' (*neutral*) from 'hoki'.
4. Add either a prefix or a suffix to an existing word, e.g. The word 'kālai' (*intellectual policy*) has been transformed into a prefix meaning "-ology, the scientific study of." With this meaning, it has been used to form new words such as 'kālaiaopaku' (*physical science*) and 'kālaianiau' (*climatology*).
5. Explain the meaning of a word or term by using Hawaiian words, e.g. 'ala mōlehu' (*crepuscular*) and "ōlelo kuhi lima 'Amelika' (*American Sign Language*).
6. Combine Hawaiian words to create a new word, e.g. hamulau (*herbivore*), ka'a'ike (*communication*).
7. Combine Hawaiian words while shortening at least one of the words, e.g. 'ikehu' (*energy*) which was formed by combining 'ika' and 'ehu; lāhulu' (*species*) which comes from a shortened 'lāhui' plus "hulu; and 'mo'olako' (*inventory*) which comes from "mo'olelo and 'lako'.
8. Extend the meaning of a word which is already found in the dictionary, or give an existing word a new meaning, e.g. haumia (*pollution*) and kaulua (*double, in math*).
9. Use a word or part of a word from another Polynesian language with its meaning intact or slightly changed, e.g. such as kōkaha (*condensation*) from the Māori word tōtā, and ha'uki (*sport*) from the Tahitian word ha'uti.
10. Hawaiianize the orthography of a word or term from a non-Polynesian language, e.g. naelona (*nylon*) and 'okikene (*oxygen*).



## Israel – Evolving lexicography

In 1938, the Millon ivri 'Hebrew dictionary' of Yehudah Goor (Garsovski) was published. Its second edition (1946) was, in its day, probably the most favoured dictionary of Hebrew (Elsevier, 2006). This dictionary played a major role in publicly disseminating, articulating and implementing Hebrew as a “modern spoken language of the post-renaissance world” (Fellman, 2011).

Millon h.adash 'A new dictionary,' by Avraham Even-Shoshan, whose first edition (5 vols, 1948–1952) and in particular the second edition, ha-Millon he-h.adash 'The new dictionary' (1966), dominated the dictionary scene in Israel for the second half of the 20th century.

As well as the inclusion of the entire lexicon and newly coined words, this was the first Hebrew dictionary to include loan words to reflect the rapidly changing world and the influences of foreign languages and cultures upon Israel. The latest version (edited by Moshe Azar, 2003) has been updated mainly with new entries, but the basic architecture and structure are not affected in any way (Elsevier, 2006).

“LRav-Milim: the complete dictionary of Modern Hebrew” by Yaacov Choueka and the Rav-Milim team (6 vols, 1997) has very different lexicographic principles, design and structure in comparison to previous dictionaries. The lexicon covers all registers<sup>2</sup> of the language, and every entry is supplemented by the word's 'family.' As a user-friendly dictionary, it has been computerised and made easily accessible to the public through the internet (Elsevier, 2006).

<sup>2</sup> Registers are the linguistic features of words associated with the context and situation in which they are used (Halladay & Hassan 1976).



## Australia - Lexicography in a multilingual Aboriginal Australia

Australian Aboriginal lexicography is unique as there are many Aboriginal languages within Australia.

Below are some examples taken from Papers in Australian Linguistics No. 15: Australian Aboriginal Lexicography (1983):

The lexicon in the Meriam Mir Dictionary were collected by the following methods

- (a) by the sound of the word;
- (b) based on the meaning of the word;
- (c) picking stories from texts.

It is good practice in the compilation of Walpiri language lexicon that “each entry must specify, in addition to the part of speech of the item, the syntactic environment in which the item occurs,” or that each entry must be complimented by an example of how the word is placed in the sentence according to the structure of the language.

For example:

Ngarrka-ngku wawirri pantu-rnu (kurlarda-rlu)

Man kangaroo spear (spear)

'The man speared the kangaroo (with a spear).'

For southern Pilbara languages, the compilation of a bilingual dictionary was divided into four tasks

- (a) preparation of a text collection from which words and examples of their use may be drawn;
- (b) preparation of the main lexical body of the dictionary in which Aboriginal language words with their definitions are given;
- (c) writing an introduction to the dictionary, including instructions on its use;
- (d) product ion of an English-Aboriginal language finder list so that equivalents of selected English words may be found in the main body of the dictionary.



## Spain – Lexicography in tertiary education

In the Spanish-speaking world, lexicography is becoming a discipline of its own in research centres and universities. This is shown in the increasing number of conferences and publications as well as the establishment of postgraduate programmes and courses on lexicography offered by several universities in the Hispanic world, especially in Spain.

It is growing popular to adopt a pan-Hispanic perspective in these types of educational programmes. A good example of this is the Master's Degree in Hispanic Lexicography offered by the ULE (Universidad de León) in collaboration with the School of Hispanic Lexicography (Escuela de Lexicografía Hispánica, ELH) (Lacorte, 2015).

## Basque – Preserving an ancient language on the French-Spanish border

In early studies of the indigenous language of the Basque, it was initially believed that the language was of Indo-European, Afro-asiatic, old-European, Iberian and/or Caucasian origins, however no substantial evidence has been found to prove this certain.

The Basque language has intrigued linguists for years as very minimal connections have been made as to where the language was derived (Trask, 1997). According to Basque language (2013), it has preserved its distinctive characteristics despite the overwhelming pressure to which it has been subjected over a period of at least 2,000 years. “The earliest known glossary of Basque is a list of a dozen words recorded by the French pilgrim Aimery Picaud in the 12th century,” and Basque language has been well documented since then (Trask, 2008).

Alongside the preservation of the language's characteristics and its original lexicon, it has evolved with the times. In order to expand its lexicon, Basque language has turned to borrowing and using loanwords from the neighbouring languages, such as 'hira' 'anger' from the Latin or Romance term ira (Trask, 2008).



## Aotearoa – Looking at new approaches to lexicon development and dissemination

According to Neologisms<sup>3</sup> Briefing (2014), Aotearoa has a plethora of dictionaries, wordlists and thesauri both available physically and/or online. Over the years, neologisms have been coined by Te Taura Whiri i te Reo Māori e.g. Te Matatiki to compliment new corpora<sup>4</sup>.

Wider Aotearoa has also compiled well-used dictionaries such as H.W. Williams's A dictionary of the Māori language (1971), Ryan's (2012) The Raupo dictionary of modern Māori (4th edition), H.M. Ngata's English-Māori dictionary in 1993 (hardcopy and online) and John Moorfield's Te Aka Māori-English, English-Māori Dictionary (2011) (3rd edition) (available in various multimedia formats) (Neologisms Briefing, 2014).

Through the lexicographical work of Te Taura Whiri i te Reo Māori (n.d.) and reo Māori linguists across Aotearoa, lexical extension has become a regular practice. Extending the lexicon of te reo Māori has allowed for the language to keep up with new words of the time (e.g. Kowheori-19, Covid-19).

Neologisms Briefing (2014) acknowledges the efforts and successes of Aotearoa lexicography and lexicology. However, it does highlight areas for improvement, specifically how the lexicon is presented and disseminated. It states a need for

1. a dictionary that provides etymology (word history),
2. acknowledgement of word productivity and frequency (how often a word is used),
3. acknowledgement of words being used in spoken language (as current dictionaries only draw from text examples),
4. information on word pronunciation (specifically syllabification and stress placement).

<sup>3</sup> A neologism is the name for a newly coined term, word, or phrase that may be in the process of entering common use but that has not yet been accepted into mainstream language. or a new usage of an existing word (Neologisms Briefing, 2014).

<sup>4</sup> Corpora is the plural of corpus, meaning collection of texts (Merriam Webster, n.d.).

## ▶ Key Findings

The literature has highlighted key findings and themes relative to each of the countries researched

- 
- |                    |   |
|--------------------|---|
| ▶ <b>Wales</b>     | The promotion of the use of Purisic terms has altered the use of their language and increased the confidence and language appreciation within its speakers.                       |
| ▶ <b>Hawai'i</b>   | The establishment of the Hawai'ian language lexicon committee, Kōmike Hua'ōlelo, has helped to develop lexicographical guidelines for lexicon development.                        |
| ▶ <b>Israel</b>    | Hebrew Lexicography has continued to evolve with the influences of foreign languages and cultures and the needs of the time.  |
| ▶ <b>Australia</b> | Australian Aboriginal Lexicography caters to multiple peoples and multiple languages according to their unique individual contexts.   |
| ▶ <b>Spain</b>     | Lexicography has become increasingly popular as its own discipline across Spanish-speaking countries with the establishment of postgraduate programmes and courses.               |
| ▶ <b>Basque</b>    | The Basque continues to preserve its over 2,000-year-old language by expanding its lexicon using loan words from neighbouring countries.  |
| ▶ <b>Aotearoa</b>  | As a stalwart within endangered language lexicography, Aotearoa looks to improve its lexical resources to cater to the present day by looking at international lexical practices. |
- 

## ▶ Conclusion

From the examples of good practice given in the international literature, how can Te Taura Whiri i te Reo Māori improve its lexical practices?

The literature on the international development and dissemination of lexicon emphasises good practices that could be implemented by Te Taura Whiri i te Reo Māori and wider Aotearoa, to help improve lexical development and dissemination within Aotearoa.

These practices could include:

1. Continuing the good practices that are already successfully implemented in Aotearoa, which are seen as good practice internationally
2. Further development and creation of resources to create and disseminate lexicon successfully
3. The establishment of a lexicon committee
4. The establishment of a lexicon community
5. The standardisation of reo Māori lexicography with acknowledgement of dialect and orthographic conventions
6. The development of lexicographical guidelines
7. The establishment of lexicon development, lexicography and/or lexicology as a tertiary discipline
8. Implementing the following when compiling dictionaries and disseminating lexicon
  - a) providing etymology (word history)
  - b) the acknowledgement of word productivity and frequency (how often a word is used)
  - c) the acknowledgement of words being used in spoken language (as current dictionaries only draw from text examples)
  - d) information on word pronunciation (specifically syllabification and stress placement)

## References

- 'Aha Pūnana Leo., Kōmike Hua'ōlelo., & University of Hawaii at Hilo. (2003). Māmaka kaiao: A modern Hawaiian vocabulary : a compilation of Hawaiian words that have been created, collected, and approved by the Hawaiian Lexicon Committee from 1987 through 2000. Honolulu: University of Hawai'i Press.
- Ball, & Muller, N. (2009). The Celtic Languages. (2nd ed.). Taylor & Francis. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/tetaur/detail.action?docID=448323>
- Denez, G., & Hupel, E. (2009). Gwalarn: An Attempt to Renew Breton Literature. Proceedings of the Harvard Celtic Colloquium, 29, 70–80. <http://www.jstor.org/stable/41219633>
- Elsevier. (2006). Hebrew Lexicography. (K. Brown, Ed.) Encyclopedia of Language & Linguistics (Second Edition), 268–272. doi:<https://doi.org/10.1016/B0-08-044854-2/04345-5>.
- Fellman, J. (2011). The role of Eliezer Ben Yehuda in the revival of the Hebrew Language: An Assessment. In J. Fishman (Ed.), Advances in language planning (pp. 427–456). Berlin, Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783111583600.427>
- Hale, K. (1983). A Lexicographic Study of Some Australian Languages: Project Description. Pacific Linguistics. Series A. Occasional Papers, 71–108.
- Halliday, M. A. K. & Hasan, R. (1976), Cohesion in English, London: Longman
- Hawke. (2018). Coping with an expanding vocabulary: the lexicographical contribution to Welsh. International Journal of Lexicography, 31(2), 229–248. <https://doi.org/10.1093/ijl/ecy004>
- Ježek, Elisabetta. The Lexicon : An Introduction, Oxford University Press, 2016. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/tetaur/detail.action?docID=4413991>
- Ka Haka 'Ula O Ke'elikōlani. (n.d.). Retrieved from Kualono: <http://www.olelo.hawaii.edu/en/hale-kuamoo/komike-huaolelo-hou>
- Lacorte. (2015). The Routledge handbook of Hispanic applied linguistics. Routledge. <https://doi.org/10.4324/9781315882727>
- Merriam-Webster. (n.d.). Corpora. In Merriam-Webster.com dictionary. Retrieved March 29, 2022, from <https://www.merriam-webster.com/dictionary/corpora>
- McConvell, P., Day, R., & Black, P. (1983). Making a meriam mir dictionary. Pacific Linguistics. Series A. Occasional Papers, 0(66), 19–30.
- Michelena, L. and Rijk, . Rudolf P.G. de (2013, February 25). Basque language. Encyclopedia Britannica. <https://www.britannica.com/topic/Basque-language>
- Singleton. (2016). Language and the lexicon : an introduction. Routledge. <https://doi.org/10.4324/9781315824796>
- Te Taura Whiri i te Reo Māori. Neologisms Briefing. (2014, June 25). Wellington: Te Taura Whiri i te Reo Māori.
- Te Taura Whiri i te Reo Māori. The Lexical Work of Te Taura Whiri i te Reo Māori. (n.d). Wellington: Te Taura Whiri i te Reo Māori.
- Trask, R. L. (2008) Etymological Dictionary of Basque, ed. for web publication by M. W. Wheeler. Brighton, University of Sussex. [https://www.academia.edu/2154989/Etymological\\_Dictionary\\_of\\_Basque\\_by\\_R\\_L\\_Trask\\_edited\\_for\\_web\\_publication\\_by\\_Max\\_W\\_Wheeler](https://www.academia.edu/2154989/Etymological_Dictionary_of_Basque_by_R_L_Trask_edited_for_web_publication_by_Max_W_Wheeler)
- Trask, R.L. (1997). The History of Basque (1st ed.). Routledge. <https://doi-org.ezproxy.waikato.ac.nz/10.4324/9781315004358>

