



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://researchcommons.waikato.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

No explanation required:
Entrenchment and perception of
Māori loanwords
in a diachronic newspaper corpus
of New Zealand English

A thesis
submitted in fulfilment
of the requirements for the degree
of
Master of Arts (Applied) in Applied Linguistics
at
The University of Waikato
by
KATHARINE JOY LEVENDIS



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2018

Abstract

Māori loanwords are uncontestedly the most notable feature of New Zealand English (Gordon, 2005; Macalister, 2006b). Although frequencies of Māori loanwords are reported to be increasing in recent years (Macalister, 2006b), and analyses strongly indicate a skew in loanwords across certain topics (Davies & Maclagan, 2006; de Bres, 2006), no studies have yet addressed whether Māori loanword behaviour differs in a corpus restricted to a single topic (with the exception of Calude, Miller, Harper, & Whaanga, Forthcoming). Nor has a thorough diachronic analysis been made on the subject since Davies and Maclagan (2006) and Macalister (2006b). This thesis is concerned with frequency of use and perception of Māori loanwords in a diachronic corpus restricted to a singular theme: *te Wiki o te Reo Māori*, Māori Language Week. Using a quantitative methodology in a corpus constructed from New Zealand regional and national newspaper articles spanning 2008-2017, average Māori loanword frequencies are found to be almost 5 times that of the most recent comparable diachronic study (Macalister, 2006b), and demonstrate a statistically significant diachronic increase in use. In contrast, the markedness (or translation) of Māori loanwords shows a statistically significant decreasing diachronic trend over the 10-year period. However, practice of marking appears to be affected by user perceptions and intention to educate, and therefore cannot be relied upon as a gauge of loanword entrenchment. Other conventional measures of loanword entrenchment such as frequency, listedness and morphological assimilation are also shown to be unreliable when considered independently from one another, and in cases such as that of New Zealand English where a change from above influenced by social and cultural factors (Māori loanwords functioning as an expression of political attitude and stance) appears to be interfering in loanword use. Measurement of Māori loanword entrenchment cannot be made using the same criteria as loanwords from non-threatened languages, and suggestions are made for the revised criteria for studying loanword entrenchment which takes into account the sociolinguistic context in which the loanwords are used.

Acknowledgements

This my moment to finally and officially thank in print everyone who has contributed to the completion of this thesis. There is no way I could have done it alone. First of all, I want to acknowledge Dr. Andreea Calude, my wonderful supervisor. Not only for proposing the possibility of this topic as a Masters, for providing the opportunity to take my work to Japan, and for enabling me to talk about it in the media – thus perpetuating the topic of Māori loanwords in newspapers! – but also for being so supportive, kind and encouraging at every step of the project. I am very lucky to have had such a supervisor. Thanks also go to Dr. Hēmi Whaanga for providing help and advice on particular Māori loanwords, and to Anne Ferrier-Watson, my incredible subject librarian. Nor can I forget the Royal Society Marsden Grant Fund, or the Waikato Graduate Women Educational Trust, without whose financial support I would not have been able to undertake this opportunity.

I would love to be able to say that all of this work was done on my own strength, but as a Christian I cannot deny that, without God’s help, I would have crashed and burned many months ago! The same goes for the support of family and friends, especially my parents, constant encouragers. For all the pep talks, every second night it seemed at times, thanks Mum! For having confidence that I’d get through it, and for taking phone calls at random hours, thanks Dad! To all my friends who listened to my agonies and more than their fair share of often unsolicited linguistic diatribes, thank you for sticking with me! It’s finally done!

Tēnā ko te hunga e tatari ana ki a Ihowā, puta hou ana he kaha mō rātou;
kake ana rātou ki runga; ko ngā parirau, koia ānō kei o ngā ēkara; ka rere
rātou, ā, e kore e māuiui; ka haere, ā, e kore e ngenge.

(Ihāia 40:31 Paipera Tapu 2012)

Table of contents

Abstract	ii
Acknowledgements	iii
Table of contents	iv
List of figures	vii
List of tables	ix
1. Introduction	1
2. Literature review	7
2.1. Lexical borrowing and loanwords	7
2.1.1. Defining loanwords.....	7
2.2. Previous research on Māori loanwords in New Zealand English	9
2.2.1. Māori loanwords in newspapers	9
2.2.2. Māori loanwords in the television media.....	19
2.2.3. Māori loanwords in picture books	21
2.2.4. Māori loanwords in other studies.....	23
2.3. Summary	24
3. Methodology	26
3.1. Background	26
3.2. The search parameters.....	27
3.3. The search results.....	28
3.3.1. TW sub-corpus search results	28
3.3.2. MLW sub-corpus search results.....	29
3.4. Article exclusions.....	30
3.5. Codeswitching exclusions.....	30
3.6. Irregular article retentions.....	31
3.7. Coding the data	31
3.7.1. Coding the metadata	32
3.7.2. Coding the loanword data	33
3.7.2.1. Distinguishing between loanword types	33
3.7.2.2. Variants of loanword types	34
3.7.2.3. Marked loanwords	35
3.7.2.4. Semantic coding.....	35
3.7.2.5. Loanwords in article titles.....	35
3.7.2.6. Loanword exclusions	36

3.7.2.6.1.	Proper nouns	36
3.7.2.6.2.	Search terms.....	36
3.7.2.6.3.	Māori Language Week year themes	36
3.7.2.6.4.	Unclear meanings and other exclusions.....	37
4.	The data	38
4.1.	Overview of the corpus.....	38
4.2.	The articles.....	38
4.3.	The newspapers.....	39
4.3.1.	The newspaper authors	41
4.4.	The search term data	41
4.5.	The Māori Language Week year theme data	42
5.	Loanword use	43
5.1.	Linguistic characteristics	43
5.1.1.	Loanword frequencies.....	43
5.1.2.	Loanword frequencies in article titles.....	47
5.1.3.	Semantic field	49
5.1.4.	Pluralisation by suffix	53
5.2.	Extra-linguistic patterns	57
5.2.1.	Loanword use in newspapers	57
5.2.1.1.	Most frequent loanwords in newspapers.....	59
5.2.1.2.	Average loanword frequencies in newspapers.....	61
5.2.1.3.	Diachronic effect of loanword use in newspapers	63
5.2.1.4.	Effect of ethnicity on loanword use in newspapers	65
5.2.2.	Diachronic use of loanwords in the whole corpus	67
5.2.3.	Selection of individual loanwords for diachronic analysis	68
5.3.	Case study: <i>te reo Māori</i>	71
5.3.1.	Frequencies of language-referent loanwords	72
5.3.2.	Sub-types of Māori language-referent loanwords.....	73
5.3.3.	Sub-types of Māori language-referent loanwords in newspapers.....	75
5.3.4.	Diachronic use of sub-types of Māori language-referent loanwords.....	77
5.4.	Case study: <i>Marae</i>	81
5.4.1.	<i>Marae</i> diachronic results.....	81
5.4.2.	<i>Marae</i> in newspapers	83
5.4.3.	Semantic analysis of <i>marae</i>	84

5.4.4.	Collocates of <i>marae</i>	87
5.4.4.1.	<i>Marae</i> and <i>manaakitanga</i>	88
5.4.4.2.	Other loanword collocates with <i>marae</i>	90
5.4.5.	<i>Marae</i> summary	94
6.	Loanword entrenchment and perception	96
6.1.	Listedness in loanwords	96
6.1.1.	Listedness in newspapers	99
6.1.2.	Diachronic effect of listedness	101
6.2.	Marked loanwords	102
6.2.1.	Summary of terms	102
6.2.2.	Mode of markedness	104
6.2.3.	Language order in marked loanwords	106
6.2.3.1.	Translations of marked loanwords	106
6.2.3.2.	Translation of corpus search terms	109
6.2.4.	Marked loanwords in newspapers	111
6.2.4.1.	Diachronic marked loanword use in newspapers	114
6.3.	Marked loanword frequencies in the whole corpus	116
6.3.1.	Diachronic frequencies of all marked loanwords	121
6.4.	Author perception of loanword entrenchment	124
6.4.1.	Author perception of loanword entrenchment: Diachronic effect ...	127
6.4.2.	Continuum of perceived loanword entrenchment	129
7.	Discussion	132
7.1.	Māori loanword use as an expression of political attitude	132
7.2.	Measuring entrenchment of loanwords from Māori as a threatened language	138
7.3.	Māori loanwords in NZE cannot be evaluated by the same criteria as loanwords originating in non-threatened languages	143
8.	Conclusion	149
8.1.	Limitations and future suggestions	149
8.2.	Concluding remarks	152
9.	References	154
	Appendix A: A complete list of the loanwords	161

List of figures

Figure 1: Initial search parameters in Newztext (TW sub-corpus).....	27
Figure 2: Secondary search parameters in Newztext.....	28
Figure 3: Summary of article results by year and source newspaper in Newztext (TW sub-corpus).....	28
Figure 4: Summary of article results by year and source newspaper in Newztext (MLW sub-corpus). Image reproduced with permission.....	29
Figure 5: Map of newspaper article distribution (percentage of corpus).....	40
Figure 6: Loanword types by frequency band	43
Figure 7: Frequency of most common loanword types in article titles.....	48
Figure 8: Percentage of loanword types by semantic field.....	52
Figure 9: Percentage of loanword tokens by semantic field.....	52
Figure 10: Loanword token frequencies per 1,000 words by newspaper (including proper nouns).....	62
Figure 11: Loanword token frequencies per 1,000 words by newspaper (excluding proper nouns).....	62
Figure 12: Diachronic loanword frequencies per 1,000 words per year, by newspaper	64
Figure 13: Percentage of Māori and non-Māori populations by region.....	66
Figure 14: Comparison of regional Māori population data and newspaper loanword frequencies per 1,000.....	67
Figure 15: Diachronic loanword token frequencies per 1,000 words.....	68
Figure 16: Maori token frequencies by semantic reference (excluding ‘te reo Maori’).....	72
Figure 17: Percentage of tokens of Māori language-referent sub-types.....	75
Figure 18: Frequency of Māori language-referent sub-types by newspaper	76
Figure 19: Comparison of Māori language-referent token frequencies to all other token frequencies by newspaper.....	77
Figure 20: Diachronic token frequencies for Māori language-referent sub-types.....	78
Figure 21: Diachronic frequency distribution of <i>marae</i> non-proper noun tokens.....	82
Figure 22: Diachronic frequencies of <i>marae</i> tokens per 1,000 words.....	83
Figure 23: Semantic categorisation of <i>marae</i> by newspaper.....	86
Figure 24: Semantic categorisation of <i>marae</i> by year	86
Figure 25: Percentages of types and tokens by listedness	97

Figure 26: Percentage of listed tokens by newspaper	100
Figure 27: Listed and unlisted loanword token frequencies per year	101
Figure 28: Marked type frequencies by mode of markedness	105
Figure 29: Diachronic frequencies of marked tokens by mode of markedness...	105
Figure 30: Token frequency of language order of translations/definitions in marked loanwords.....	107
Figure 31: Comparison of article frequencies and search term frequencies.....	109
Figure 32: Diachronic percentages of translated (marked) search terms.....	110
Figure 33: Marked token and type frequencies, by newspaper	112
Figure 34: Comparison of marked and unmarked type frequencies, by newspaper	114
Figure 35: Diachronic marked token frequencies by newspaper.....	115
Figure 36: Percentages of marked and unmarked loanwords	118
Figure 37: Marked loanword tokens by year	122
Figure 38: Percentage of marked loanwords by year.....	123
Figure 39: Continuum of anecdotes of author perceptions of loanword entrenchment.....	130

List of tables

Table 1: Loanword categories by form and meaning (adapted from Zenner & Kristiansen, 2014).....	8
Table 2: Example of article file name coding.....	32
Table 3: Distinction of loanword types in <i>kura</i> , <i>kaupapa</i> , <i>Maori</i>	34
Table 4: Examples of loanword type variants.....	35
Table 5: Summary of words excluded as codeswitching.....	37
Table 6: Types and tokens in the corpus.....	38
Table 7: Diachronic article, type and token data	38
Table 8: Newspaper of publication article data	39
Table 9: Search term frequencies.....	41
Table 10: Year themes and frequencies	42
Table 11: Loanwords with a frequency of 10 or greater.....	44
Table 12: Comparison of loanword frequency data with Kennedy (2001) and Macalister (2006b).....	46
Table 13: Title loanword data per article.....	48
Table 14: Type and token frequencies by semantic field	49
Table 15: English suffix pluralisation in the tokens	54
Table 16: Summary of newspaper data by article, type and token	58
Table 17: Token frequencies of the five most frequent loanwords, by newspaper.....	60
Table 18: Average loanword frequencies per 1,000 words, by newspaper	60
Table 19: Loanword token and type article averages by newspaper	61
Table 20: Diachronic loanword frequencies per 1,000 words, by newspaper.....	64
Table 21: Type frequency distribution across 2008-2017	69
Table 22: Diachronic distribution of most frequent loanword tokens.....	70
Table 23: Frequencies of Māori language-referent loanword sub-types	74
Table 24: Percentage of Māori language-referent loanword tokens by newspaper	77
Table 25: Diachronic frequencies for all Māori language-referent sub-types	79
Table 26: Diachronic token frequencies of all Māori language and all non-Māori language-referent loanwords.....	79
Table 27: Kendall Tau results for diachronic Māori language-referent loanword types	80
Table 28: Diachronic <i>marae</i> tokens by newspaper.....	84

Table 29: Semantic analysis of <i>marae</i> (terms adapted from Degani, 2010).....	85
Table 30: Most frequent loanwords to co-occur with <i>marae</i> in the corpus.....	88
Table 31: Comparison of <i>marae</i> and <i>manaakitanga</i> co-occurrences in 2011	89
Table 32: Semantic field frequencies of loanword collocates with <i>marae</i>	90
Table 33: Comparison of <i>marae</i> tokens and education collocates	92
Table 34: Types and token frequencies by listedness.....	96
Table 35: Listed and unlisted types and tokens by frequency band	97
Table 36: Distribution of unlisted tokens by newspaper	100
Table 37: Listed and unlisted loanword token frequencies by newspaper	100
Table 38: Typology of loanword marking in spoken and written language	103
Table 39: Examples of markedness by mode	104
Table 40: Summary of marked loanword data (English to Māori directional translations).....	108
Table 41: Marked and unmarked token frequencies, by newspaper.....	112
Table 42: Summary of frequencies of markedness in all loanword types	116
Table 43: Summary of diachronic marked loanword data.....	122
Table 44: Comparison of explicitly mentioned loanwords with corpus frequencies, listedness, and markedness	125
Table 45: Comparison of marked and listed types explicitly mentioned by authors	127
Table 46: Explicitly mentioned loanwords by year	128

1. Introduction

It is beyond contention that Māori loanwords account for the most distinct difference between New Zealand English and its counterparts in the Anglophone world (Gordon, 2005; Macalister, 2006b). This subject has been treated extensively from the time English speakers first arrived on the shores of Aotearoa New Zealand, and with renewed interest over the last several decades. Such recent studies have variously considered the use of Māori loanwords in newspapers (Calude, Miller, Harper, & Whaanga, Forthcoming; Davies & Maclagan, 2006; Degani, 2010), the television media (de Bres, 2006), picture books (Daly, 2007, 2009, 2013) and multi-genre corpora, both written and spoken (Kennedy, 2001; Macalister, 2006b; Onysko & Calude, 2014). Cultural conceptions and motivations of loanwords have been examined (Degani, 2017; Grant, 2012); as has familiarity with loanwords over time (Macalister, 2007; Macalister, 2008) and shifts in semantic connotations (Macalister, 2001; Metge, 2009). Sociolinguistic approaches have been taken (Calude, Miller, & Pagel, 2017). Loanwords have been considered individually (de Bres, 2006; Degani, 2017; Kennedy, 2001; Macalister, 2001) and in hybrid compounds (Degani & Onysko, 2010). More recently, networking approaches in a topic-bound corpus have been considered (Calude et al., Forthcoming). Much of this research has also been concerned with the influence of ethnicity on loanword use (e.g., Calude et al., Forthcoming; de Bres, 2006; Degani, 2017; Kennedy, 2001; Onysko & Calude, 2014). The studies mentioned here are only a small representation of work done in the field, and give an indication of just how pervasive Māori loanwords are at all levels of New Zealand society.

One issue of great importance to general studies in loanwords is loanword entrenchment (sometimes termed integration)¹. According to Backus (2014), each loanword or borrowing is a choice, and the more it is chosen, and therefore used, the more entrenched it becomes in a receptor language. Turpin (1998) agrees that loanword frequency is an indicator of entrenchment, writing that “the more a loanword is diffused (and accepted) in the linguistic community, the more it is integrated into the recipient language” (p. 231). Although levels of frequency are widely held to indicate a loanword’s place along a continuum of entrenchment

¹ The present research prefers the term ‘entrenchment’ to ‘integration’ for reasons of cultural sensitivity. For an expanded discussion, refer to Section 6.

(Davies & Maclagan, 2006; Degani, 2010; Stammers & Deuchar, 2012), other research has found that loanwords are assimilated into receptor languages suddenly and abruptly through the study of morphological inflection (Poplack & Dion, 2012). How well a loanword conforms to the grammatical and morphological rules of a receptor language has also been used as a measurement of loanword entrenchment (Poplack, Wheeler, & Westwood, 1989; Stammers & Deuchar, 2012; Turpin, 1998).

But frequency of use and morphological assimilation are not the only factors in loanword entrenchment. Loanwords which are listed in a dictionary are generally considered to be well established in that language, while unlisted loanwords are nonces (Muysken, 2000). Several studies agree that loanword listedness is a gauge of how well entrenched a loanword is (Poplack, Sankoff, & Miller, 1988; Stammers & Deuchar, 2012). Similarly, a lack of markedness² in loanwords indicates better levels of loanword entrenchment (Sharp, 2007). While marking generally indicates non-standard usage in a receptor language, it also “signals the writer’s awareness that the borrowing is a marked [deliberate] choice” (Grant-Russell & Beaudet, 1999, p. 26). Furthermore, correlations between loanword frequencies, listedness and markedness have been noted in Sharp (2007), in which topically unrestrained corpora have higher frequencies of listed than unlisted loanwords, and high levels of marking; while topically-constrained corpora have higher uses of unlisted loanwords and less marking.

Issues of entrenchment have not been completely ignored in research on Māori loanwords in New Zealand English. Although most are concerned with loanword frequency of use, some have touched upon listedness (Degani, 2010; Kennedy, 2001; Macalister, 2006b), markedness (Daly, 2007; Davies & Maclagan, 2006; Degani, 2010), and morphological entrenchment (Davies & Maclagan, 2006; de Bres, 2006; Degani, 2010; Onysko & Calude, 2014). However, while most authors agree that such factors indicate levels of loanword entrenchment and their acceptance in New Zealand English (hereafter ‘NZE’), none explore them in any great depth. This gap constitutes one main focus of this study.

² Markedness here refers to what is variously termed ‘marking’, ‘flagging’, or ‘glossing’ in the literature.

Macalister (2006b) has stated that “it is widely accepted that the most distinctive feature of New Zealand English is the presence of words of Maori [*sic*] origin”³ (p. 1). This is a sentiment echoed time and again in the literature. Māori loanwords have been shown to be a marker of bicultural identity (Davies & Maclagan, 2006), while the Māori language is a vital part of national identity and culture, regardless of ethnicity (Albury, 2016). In addition to making NZE unique (Gordon, 2005), the Māori language has also donated a large number of lexical items to the English language worldwide (Durkin, 2014). Māori is the fifteenth highest contributor of lexis to English (Durkin, 2014, p. 25). This in itself is remarkable, given that NZE is “the smallest of the regional varieties of English belonging to the ‘inner circle’ described by Kachru” (Kennedy, 2001, p. 60), and has the lowest current number of speakers of any country mentioned in Kachru’s (1988) three concentric circles of English⁴ (p. 5). It has been estimated that the average speaker of NZE knows approximately 70-80 Māori loanwords (Macalister, 2004, p. 73).

In order to understand what it is about Māori loanwords in NZE that is so special, and why they are important, it is first necessary to contextualise them within the history of contact and relationship that exists between the Māori language and NZE. There is certainly no shortage of literature on the topic of language contact in New Zealand, most of it explored in greater depth and more eloquence than may be achieved here. Comprehensive general histories of New Zealand can be found in Anderson, Binney and Harris (2014) and King (2007); while Moon (2013) covers certain issues such as the Māori renaissance in greater detail. Albury (2016) in particular provides a concise and engaging account of the Māori language since British colonisation. More loanword-centric histories of language contact in New Zealand are included in Degani (2010) Kennedy (2001), and Macalister (2006b) amongst others. Because of the wealth of information available on the subject, only a brief summary of contact between the Māori and English languages in New Zealand will be provided here, as it pertains to Māori loanwords in NZE.

³ While current convention dictates that macrons are used where appropriate in Māori spelling, this is not always the case in the literature. Any quotes in which *Māori* does not have a macron are left unchanged in this work, and marked as such by [*sic*]. Likewise, loanwords in this or previous studies which do not have macrons represent the original form of the word in the data, and not the correct form.

⁴ Population statistics are taken from the *World Population Prospects: The 2017 Revision* (United Nations, 2017).

The indigenous people of New Zealand, the Māori, arrived from Polynesia between approximately 1210 and 1385AD (Moon, 2013). Putting aside a brief encounter between Abel Tasman in 1642, Māori did not have official contact with Europeans until Captain James Cook arrived in 1769 (Moon, 2013).

Since the arrival of the British in New Zealand, there have been three stages of colonisation: colonisation (prior to the 1880s), recolonisation (from approximately 1880 to the 1970s) and decolonisation (from the 1970s) (Belich, 2001, as cited in Macalister, 2006b). Similarly, Māori language policy has progressed through three stages since contact with the English language: colonial tolerance, language shift, and language revitalisation (Albury, 2016). The revitalisation period, sometimes called the ‘Māori renaissance’, is thus named as it follows the middle period of recolonisation, in which the Māori language was almost extinguished by colonial actions such as the Native Schools Act of 1867, which effectively forbade the use of the Māori language in the classroom (Pawley, 1989). This had a near catastrophic impact on the language, and by the 1970s, only two percent of Māori children spoke Māori as a first language (Pawley, 1989). Since then, concerted efforts have been made to revive the language. Such efforts included the establishment of *kōhanga reo* (Māori language preschools; literally ‘language nests’); the establishment of Te Taura Whiri i te Reo Māori (the Māori Language Commission) to promote the well-being of the language; the recognition of Māori as an official language of New Zealand in 1987; and in 1989 and 1990 respectively, the establishment of *kura kaupapa* (Māori language primary schools) and *wānanga* (tertiary institutions that cater for Māori learning needs) (Macalister, 2006b). According to Macalister (2006b), the events of the 1970s-1990s cultural renaissance triggered a shift in the relationship between the Māori language and NZE.

In addition to this, *te Wiki o te Reo Māori*⁵ (Māori Language Week) has been established in New Zealand for almost forty years. Originating in 1972 as Māori Language Day, it attempted to bring the Māori language to the attention of all New Zealanders, and to promote and revitalise it (Pawley, 1989). In 1975, it became a week-long event (Ministry for Culture and Heritage, 2017). Recent years of Māori Language Week have focussed on teaching Māori vocabulary including

⁵ Capitalisation conventions for the spelling of *te Wiki o te Reo Māori* in this thesis follow those prescribed in Te Taura Whiri i te Reo Māori (2012, p. 17), unless otherwise stated.

some loanwords; accessible resources to do so can be found on the website for Te Taura Whiri i te Reo Māori (<http://www.tetaurawhiri.govt.nz>). Themes for the event have existed since 2007⁶, and the first ever Māori Language Week parade was held in Wellington in 2016 (Gardiner, 2016). Because of the growing social awareness surrounding it, and its inherent purpose of aiding the revitalisation of the Māori language, the theme of Māori Language Week is particularly well suited to the construction of a topically-constrained newspaper corpus of Māori loanwords.

In such a corpus, it might be expected that Māori loanword frequencies would be higher than those of previous years, owing to the effects of topical attraction (see de Bres, 2006; Sharp, 2007). However, any study of Māori loanword frequencies must also take into consideration a diachronic perspective. It has already been mentioned that there are three stages of colonisation in New Zealand's history. These stages resulted in the borrowing of Māori loanwords into NZE in "two distinct waves" (Macalister, 2006b, p. 10). The first of these was during the initial colonisation period, as new objects and concepts were encountered and naming needs were met; the second, in the still-ongoing decolonisation period, has mainly been comprised of incoming social culture loanwords (Macalister, 2006b)⁷. Since then, indications have also been that frequencies of Māori loanwords used in NZE are on the rise. Increases in loanword use have been documented in diachronic studies conducted in newspapers (Davies & Maclagan, 2006), television media (de Bres, 2006), and multi-genre written corpora (Macalister, 2006b). Increases have also been found in studies comparing frequencies with earlier findings not mentioned here (Kennedy, 2001). Speakers of NZE are reportedly becoming more familiar with words of Māori origin (Macalister, 2009), and Māori loanword use may be "motivated by the expression of solidarity with the Māori perspective and the desire of aligning their identity within a Māori background" (Calude et al., 2017, p. 24).

In light of this, several questions arise. Has the historical increase in frequency of Māori loanwords in NZE continued into the present? In particular, are

⁶ Insofar as information was available to be found; see Te Taura Whiri i te Reo Māori (2007) for the first mention of a Māori Language Week theme in an official document: *tāpoi* (tourism), and Section 4.5 for further treatment of this issue.

⁷ Although there were some Māori loanwords borrowed into NZE during the recolonisation period (see Macalister, 2006b), these were relatively few in number and not enough to be considered as their own 'wave'.

frequencies in newspapers behaving as they did in Davies and Maclagan (2006), or is something different happening? Although topical effect has been noted to be a factor in loanword use, no study has yet explored Māori loanwords in a corpus bound to a single topic⁸. Constraining a loanword corpus to a particular topic such as Māori Language Week might be expected to draw out a greater number of loanwords than in previous studies. Exploring factors of loanword entrenchment beyond frequency, such as listedness, markedness and morphological assimilation, could provide a greater understanding of how loanwords function in general. Moreover, in a corpus constructed around the theme of the Māori language itself, it may be possible to make inferences regarding perceptions and attitudes to Māori loanword use and the Māori language. Such findings could potentially be applied to loanword studies around the world. The purpose of the present research, therefore, is to investigate frequencies, entrenchment and perceptions of Māori loanwords in a diachronic corpus constrained to the topic of Māori Language Week. With this in mind, and having established the linguistic and socio-cultural context of Māori loanwords in NZE, this issue will now be considered in a review of the literature.

⁸ With the exception of Calude et al. (Forthcoming).

2. Literature review

This literature review is structured with respect to the most notable studies that have been conducted to date on Māori loanwords in NZE. It will first seek to define loanwords, and then address previous research in Māori loanwords according to genre. As the present study concerns a diachronic, topic-bound newspaper corpus, precedence in this review will be given to work analysing newspaper language; followed by works in television media, picture books, and other studies. Where literature includes newspapers as one part of a multi-genre corpora, they are categorised as newspaper corpora; however, all relevant findings regardless of corpus genre are included in the same section.

A final note is that, over the course of this thesis, an issue which emerged as being far more pertinent than expected was that of loanword entrenchment. The marking of loanwords was particularly salient in the data. Due to the nature of entrenchment being a sizeable topic within the field of loanwords, and the fact that it did not emerge as such until late in the research process, entrenchment will only be mentioned in this review where specific issues of frequency, listedness, markedness, and morphological assimilation – usually regarding the pluralisation of loanwords – occurred in the literature. Due attention to this topic, including relevant literature, will instead be given in Section 6.

2.1. Lexical borrowing and loanwords

2.1.1. Defining loanwords

A loanword is the result of lexical material being transferred from a source language into a receptor language (Zenner & Kristiansen, 2014, p. 1). Loanwords are often called borrowings, although a borrowing may not necessarily be a loanword. In language contact situations, loanwords are almost always present to some degree. They are also one of the “prime manifestations of language change” (Backus, 2014, p. 21). There are numerous reasons a word might be borrowed into another language. These may include borrowing a word to fill a lexical gap; for reasons of prestige; to display values on and loyalty to a particular language; speaker perception of group identity; as a stylistic choice; or an expression of views on foreign interference in language (Degani, 2010).

Following a structuralist perspective, borrowing usually “focuses on the position of the borrowed items within the structure of the receptor language” (Zenner & Kristiansen, 2014, p. 2). The question of multiword phrases is usually excluded in such analysis (Backus, 2014). In contrast to codeswitching, where the switch takes place between grammatical systems, switches in borrowing occur morphologically and orthographically on loanwords (Zenner & Kristiansen, 2014). A hierarchy of the borrowability of loanwords exists, in which open class content items (nouns, verbs, adjectives) are more likely and easily borrowed than closed-class function words (pronouns, prepositions, conjunctions) (Degani, 2010).

Furthermore, not all loanwords are borrowed in the same manner. Following Zenner and Kristiansen (2014), loanwords are categorisable according to their form and meaning in both the source and the receptor languages. A summary of possible combinations, along with examples of Māori loanwords common in NZE where applicable, can be seen in Table 1 (adapted from Zenner & Kristiansen, 2014, p. 3).

Table 1: Loanword categories by form and meaning (adapted from Zenner & Kristiansen, 2014)

Type of loan	Form	Meaning	Example from Māori
Direct loan	Source language	Source language	<i>kauri</i> (large forest tree found only in the northern North Island)
Pseudo-loan	Source language	Receptor language	<i>utu</i> (source language meaning may include revenge, payment, reciprocity ⁹ ; receptor language meaning is generally reduced to ‘revenge’ alone)
Indirect loan	Receptor language	Source language	n/a
Language-internal loan	Receptor language	Receptor language	n/a
Hybrid loan	Mixed form	Mixed meaning	<i>Maoridom</i> (the Māori world and culture)

It should be noted that the above term ‘receptor language’ is sometimes known as the ‘receiving language’ (Kruger, 2012) or the ‘recipient language’ (Poplack & Dion, 2012; Turpin, 1998). For the purposes of the present research, as far as Māori loanwords in NZE are concerned, the Māori language will be referred to as the source language, and NZE as the receptor language. However, alternate terms may be used when quoting the literature.

⁹ See Moorfield (2018) for a more complete definition of the original Māori form of *utu*.

In the same way that loanwords types may vary, not all instances of lexical transference in language contact are the same. Historically, borrowings and codeswitches have been extremely difficult to differentiate. There is much debate over what distinguishes loanwords from codeswitches. In fact, Poplack and Dion (2012) describe it as “perhaps the thorniest issue in the field of contact linguistics today” (p. 311). While much research considers loanwords and codeswitches to be on a continuum of entrenchment into the receptor language (Degani, 2010; Stammers & Deuchar, 2012) not everyone agrees. Poplack and Dion (2012) argue that assimilation of foreign lexical items into a receptor language is not gradual but abrupt (p. 308). If this is true, it follows that loanwords in the present corpus will be more likely to suddenly occur in higher frequencies in a given year, rather than gradually increase in frequency over time.

Another important term in the discussion of loanwords and borrowings is the ‘nonce’ loanword. Nonces are loanwords which occur once and once only in a corpus, and are unattested in any dictionary. According to Poplack and Dion (2012), “the overwhelming majority of other-language material [in a receptor language] consists of lone items”, and yet, nonce loanwords “do not generally become widespread” (pp. 307, 287). When considering frequencies, loanwords may also be idiosyncratic (used by only one person), widespread (used by more than 10 people), or recurrent (used more than 10 times in the corpus) (see Table 1 in Poplack, et al., 1988, p. 57).

2.2. Previous research on Māori loanwords in New Zealand English

2.2.1. Māori loanwords in newspapers

It has already been mentioned that a large number of studies of Māori loanwords in NZE have been conducted in newspaper corpora. The following selection of the most relevant literature in this genre will be addressed in chronological order of publication. It should be noted that while some studies are based exclusively on newspaper findings, others cited here are multi-genre corpora, and so relevant findings across genres in their works will be mentioned, while giving precedence to the newspaper aspects of their work.

The earliest piece of literature on Māori loanwords in NZE to be reviewed here is that of Kennedy (2001)¹⁰. The reason for this is that Kennedy (2001) gives a good summary of the different methodologies conducted in previous studies of Māori loanwords in NZE before centring his investigation in two previously existing corpora: the *Wellington Spoken Corpus of New Zealand English (WSC)*, which was compiled in 1998 by Holmes, Vine and Johnson (1998), and the *Wellington Written Corpus of New Zealand English (WWC)* constructed in 1993 by Bauer (1993b). Although only part of the *WWC* was comprised of newspapers, and the focus of Kennedy's (2001) study was on the spoken aspect of loanwords, by comparing loanwords frequencies in written and spoken NZE per 1,000 words, as well as semantic word category distribution, it was found that loanwords in the two corpora were behaving similarly. Spoken Māori loanwords occurred at a frequency of 4.83 per 1,000 words, while written loanwords had a frequency of 5.86 per 1,000 words (Kennedy, 2001, p. 68).

Furthermore, there was a clear effect of speaker ethnicity on spoken loanword use between Māori and Pākehā speakers (Kennedy, 2001). Māori speakers tended not only to use a higher frequency of Māori loanwords when speaking (17 loanwords per 1,000 words compared to Pākehā speakers, who used on average two per 1,000 words), but they also “[had] a wider Maori [*sic*] vocabulary” (Kennedy, 2001, p. 75). In contrast, Pākehā speakers used a higher proportion of Māori proper nouns than non-proper nouns in their speech (Kennedy, 2001, p. 75). Given the similarities found in comparing written and spoken corpora, it is likely that such effects of speaker ethnicity can be assumed to be the same in the written data. However, Kennedy (2001) does not explore this issue further.

Rank order of loanword frequency was also considered, and a list of the 100 most frequent written and spoken lemmatised loanwords (including proper nouns) was provided (Kennedy, 2001). In the *WWC*, *Maori* was overwhelmingly the most frequent loanword, with over 3.5 times more tokens than the next most frequent loanwords, *te*, and *Pakeha* (the Māori term for New Zealanders of European descent)¹¹. Interestingly, eight of the 10 most frequent loanwords were proper nouns, with the remaining two loanwords being particles (*te* and *tu*). It was also

¹⁰ For similar findings, see also Kennedy and Yamazaki (2000).

¹¹ See Table 4.6 in Kennedy (2001, p. 71) for a full list of the 100 most frequent loanwords in the *WWC*.

notable that loanword rank frequencies in the *WSC* varied not only according to ethnicity and gender, but also to genre of text (Kennedy, 2001). This has potential ramifications for differences in loanword frequencies between genres, and consequently, topic.

Kennedy (2001) concludes that Māori loanword assimilation into NZE is now being driven by Māori in terms of range and context, rather than by a need to fill a lexical gap by Pākehā. Such use is “a way of deliberately marking identity...of supporting language revival and, of course, of discussing topics that are culturally associated with being Māori and that are best expressed in Māori” (Kennedy, 2001, p. 77). Consequently, it may be expected that Māori loanword use increases in corpora which are bound to topics relevant to Māori. However, Kennedy’s (2001) statement is only made regarding spoken loanwords, and, due to the nature of the construction *WSC*, no such attention is able to be given as to whether the same is true of written loanword use. It is likely that such a claim may also be made about written loanwords, but that is not an issue which Kennedy (2001) explores.

The most comprehensive diachronic study into Māori loanwords in NZE to date was conducted by Macalister (2006b), who constructed a corpus of almost 5.5 million words from three different genres of written data: newspapers, *School Journals* and *Hansard* parliamentary reports. His research took a synchronic snapshot of Māori loanwords every 30 years (termed indicator years) between 1850-2000 to explore diachronic trends. Loanwords as proper nouns, non-proper nouns and hybrids were considered; as was the semantic category of a loanword. Importantly, and contrary to Kennedy’s (2001) lemmatisation of loanwords, Macalister (2006b) considered structural words as being part of the loanword in his frequency counts; for instance, *Te Aroha* was counted as one token. Further details on the methodology and selection of texts used in Macalister (2006b) can be found in Macalister (2006a).

Although unable to establish exactly when a loanword entered NZE (a potentially unfeasible task), Macalister (2006b) found that the majority of Māori loanwords entered NZE in two waves of borrowing. 72% of loanwords entered NZE either in the colonisation period (prior to 1880) or the decolonisation period (1970 onwards). Evidence was found of “a resistance to, or at least a stabilisation, of borrowing” during the recolonisation period of 1880-1970 (Macalister, 2006b, p. 18). Furthermore, these waves were differentiated by semantic category: flora

and fauna loanwords, as well as material culture loanwords were most common in the colonisation wave, while social culture loanwords dominated the decolonisation wave (Macalister, 2006b). In fact, “the peak for both the Flora and Fauna and Material Culture categories in the newspapers was in 1850, [but] the peak for Social Culture was 2000” (Macalister, 2006b, p. 15).

Over the course of the 150 years of data, there was an increase in Māori loanword frequencies in NZE. Average loanword frequencies in 1850 were only 3.29 loanwords per 1,000 words of English (Macalister, 2006b). In contrast, by 2000, on average six out of every 1,000 words was a Māori loanword. This figure varied slightly according to text type: newspapers had a higher frequency, with 7.7 loanwords per 1,000 words; while *School Journals* had the lowest, at 5.2 per 1,000 words (Macalister, 2006b). However, a figure for *Hansard* loanword frequencies was not given, “as the debates included were those with an identified Maori [*sic*] focus, and so contained a higher-than-average proportion of Maori [*sic*] words” (Macalister, 2006b, p. 12), rather than attributing this to an effect of topical attraction. There was also a “consistently insignificant proportion” (p. 14) of flora and fauna loanwords in *Hansard*; a fact which Macalister (2006b) claimed was disadvantageous to analysis. However, the variation of loanword frequencies between genres, as well as differences in semantic domains of loanwords therein, implies that certain loanwords are more likely to be present or absent according to the topic and genre of their texts.

Following Macalister’s (2006b) findings, it is expected that there should be a Māori loanword frequency per 1,000 words in the present study similar to that of his newspaper frequency, although there may also have been a slight increase in frequencies since 2000. It is also possible that the use of Māori Language Week as a topic around which to build a corpus may result in the attraction of a greater number of loanwords. Loanwords from the social domain are expected to be most frequent, as are proper nouns, with the type *Māori* dominating, as in Macalister (2006b).

Another study of Māori loanwords in a diachronic newspaper corpus is that of Davies and Maclagan (2006), who focussed on 13 individual Māori loanwords as they occurred between 1997-2004. Four newspapers from the Fairfax Media group were included in the corpus: *The Press*, the *Waikato Times*, *The Dominion* (including *The Dominion Post*), and *The Southland Times*. This selection gave a

balanced spread of newspapers from different geographical and demographic contexts.

Davies and Maclagan (2006) took a different approach to previous studies, in that they did not measure individual loanword frequencies. Consequently, no figures were available for frequencies of loanwords per 1,000 NZE words. Rather, any article with one or more loanwords in it was considered a single token. By measuring article frequencies rather than loanword frequencies, they were able to obtain “an indication of the status of Māori issues in a Pākehā-dominated industry” (Davies & Maclagan, 2006, p. 78). It was notable that frequencies of articles with loanwords in them were higher in newspapers from regions with higher Māori populations (Davies & Maclagan, 2006). Loanwords classified as belonging to the social culture semantic domain (see Macalister, 2006b) showed a diachronic increase in frequency of articles (Davies & Maclagan, 2006). This was attributed to “an increase in social and political status for Māori” (Davies & Maclagan, 2006, p. 96). However, one material culture loanword, *marae*¹², had only a steady frequency of use, with “its meaning...widening to include a social aspect as a community of interest” (Davies & Maclagan, 2006, p. 96). A topical effect was also observed, in which certain topics “affect[ed] the number of Māori words used”; this was termed the ‘news-driven effect’ (Davies & Maclagan, 2006, p. 77).

Additionally, a “trend away from glossing” (or translating) loanwords was found (Davies & Maclagan, 2006, p. 96). Although definitions in some glosses tended to vary, especially for words which have no exact equivalent in English, such as *whakapapa*¹³, in general glossing was shown to indicate whether readers are expected to recognise and understand a loanword or not (Davies & Maclagan, 2006, p. 80). By measuring the diachronic frequencies of individual loanwords in NZE and investigating their use in both Māori and non-Māori cultural contexts, as well as whether they were glossed, Davies and Maclagan (2006) were able to place loanwords on a continuum of acceptance into NZE. Loanwords with higher frequencies, and which were never glossed, were considered to be well established in NZE; while loanwords which were glossed earlier but dropped off in later years indicated increasing comprehension and therefore “acceptance as a borrowing”

¹² Degani (2010) defines *marae* as a “courtyard of a Maori [*sic*] meeting place” (p. 176). See Appendix A for an expanded definition.

¹³ See Appendix A for definition.

(Davies & Maclagan, 2006, p. 80). Thus, frequency alone is not the only factor in the entrenchment of a Māori loanword. Likewise, the zero-inflectional use of *hui* (meeting, gathering) and *Māori* in plural form, rather than *huis* and *Māoris*, is argued to reinforce that, because Māori loanwords are no longer being forced to conform to English grammatical rules, they are therefore becoming more accepted in NZE. Davies and Maclagan (2006) go as far as to state that “not imposing the grammatical rules of the dominant language in our culture reveals a growing acceptance of te reo Māori¹⁴, and reflects the social and political status of Māori in general” (p. 89).

Degani (2010) explores the lexical impact of Māori loanwords on NZE in a corpus comprised of articles from three online newspapers: *The New Zealand Herald*, *The Dominion Post* and *The Press*. Three Māori loanwords were selected from the list of most frequent loanwords in Kennedy (2001) and given an in-depth analysis: *aroha* (love), *mana* (prestige, authority¹⁵) and *marae* (Degani, 2010). Proper noun and non-proper noun forms of each loanword were considered, as were hybrid forms.

Semantic extension in *mana* and *aroha* saw these nouns also being used as verbs and adjectives, while semantic bleaching was occurring where they were used as proper nouns (Degani, 2010). In contrast, there was “a marked tendency for [*marae*] to be used as a noun”, with occasional use as an adjective or in metonymic extension (Degani, 2010, p. 185). While semantic changes undergone by these loanwords indicate they are becoming more entrenched in NZE, there were also many instances where they were used in newspapers to exploit Māori cultural connotations for business or exoticisation purposes, especially the word *aroha* (Degani, 2010). Furthermore, Degani (2010) argues that the word *mana* is used positively when associated with European values, i.e., good politicians making positive changes, but that it is negatively associated with Māori culture, i.e., “bad Māori” and child abuse (p. 189). The loanword *mana* also has a strong connection with the All Blacks, which is exploited on the world stage (Degani, 2010). In this manner, the three loanwords often serve to reinforce existing cultural stereotypes about Māori (Degani, 2010).

¹⁴ The Māori language.

¹⁵ See Appendix A and Moorfield (2018) for full definition of the concept.

Her overall evaluation of Māori loanwords in NZE is not positive, concluding that “Maori [*sic*] appears as an appendix to NZE” (Degani, 2010, p. 192), and agreeing with Harlow’s (2005) statement that Māori loanwords in NZE actually “serve to reinforce the monolingually-based status quo” (p. 140). She also asserts that “the dominance of proper noun over common noun usage” in Māori loanwords reduces their contribution to NZE (Degani, 2010, p. 192).

This opinion is revised somewhat in Degani’s (2017) later work on the cultural schema of *marae*, where she continues the semantic exploration of the loanword in NZE. Although not conducted in newspapers, this study is mentioned here because of its strong links to the topics in Degani (2010). Linguistic anthropological techniques are used by using visual stimuli to facilitate the spoken narratives of 140 young Māori women. Their narratives locate the loanword *marae* within a set of highly specific and complex cultural schema (Degani, 2017). These include the *marae* as an expression of tribal identity, and as a connection to the land, rituals, customs and values, and their ancestors (Degani, 2017). The *marae* is partially associated with education and the spiritual, as well as being a social place of belonging and a place to transmit cultural and historical knowledge and values (Degani, 2017). There is also “a complex system of semantic interconnections among other Māori concepts that include” other loanwords such as *iwi* (tribe), *hapū* (sub-tribe), *kawa* (customs, *marae* protocol) and *tangihanga* (funeral) among others¹⁶. In fact, the cultural schema that the *marae* represents to the interviewees are complex and manifold; far more complex than those to be found in the newspapers of Degani’s (2010) earlier work. However, although these “findings demonstrate that the English language can serve as a vehicle for expressing Māori cultural conceptualisations when it is used by Māori people who have knowledge of Māori language and culture” (Degani, 2017, p. 678), it is doubtful that all loanwords mentioned in Degani (2017) will be found in newspapers, as this cultural knowledge is not shared by all New Zealanders. It is predicted that the present study may find some overlap with Degani (2017) in loanwords that co-occur with *marae*, but that any cultural associations will be diluted and certainly not as semantically rich.

¹⁶ For a full list of associated loanwords, see Degani (2017, p. 677).

Hybrid compounds of Māori loanwords in NZE is a matter which has only been considered in great detail to date by Degani and Onysko (2010). They investigated the combination of Māori loanwords with English words, which result in creative new lexical items in NZE such as *waka-jumper* and *kumara vine*¹⁷ (Degani & Onysko, 2010, p 209). *New Zealand Herald* newspaper articles spanning 2003-2008 were collected to construct a corpus of NZE comprising of just over 73 million words, to explore whether the hybrid compound productivity of Māori loanwords is an accurate gauge of their frequency in NZE (Degani & Onysko, 2010). From the lists of most frequent loanwords given in Kennedy (2001), 63 loanwords, including some proper nouns, were selected for analysis.

It was found that over a quarter of all loanwords occurred in a hybrid compound (Degani & Onysko, 2010). This presence was interpreted as both an indication “of the integration of Maori [*sic*] terms and concepts in NZE and... the vitality of the Maori [*sic*] element” in NZE (Degani & Onysko, 2010, p. 218). The most productive loanword in the creation of hybrid compounds was also the second most frequent loanword in the corpus: *Kiwi*¹⁸, in which over 50% of all occurrences belonged to a hybrid compound; while *Pakeha* was the most flexible (Degani & Onysko, 2010). In fact, loanwords which functioned as markers of identity (*Maori*, *Pakeha* and *Kiwi*) were the most likely to occur in a hybrid compound by frequency (rather than percent), as well as having the widest range of hybrid compounds (Degani & Onysko, 2010).

Although the present research will not specifically focus on hybrid compounds, the findings in Degani and Onysko (2010) are relevant because they demonstrate that, although not a fixed rule, higher frequency Māori loanwords are more likely to be creatively compounded in NZE because they are better established. By extension, this provides evidence that grammatical and morphological assimilation of a loanword into a receptor language may be able to measure its entrenchment.

In the next study, Onysko and Calude (2014) are primarily concerned with the effect ethnicity has on Māori loanword use in NZE. They performed sociolinguistic and statistical diachronic analyses on three high frequency proper

¹⁷ For full explanations of these hybrid compound terms and their cultural significance, see Degani and Onysko (2010).

¹⁸ *Kiwi* is “an affectionate term for a New Zealander” (Onysko & Calude, 2014, p. 164).

noun loanwords which also function as identity markers: *Maori*, *Kiwi* and *Pakeha* (Onysko & Calude, 2014). Their study dealt with both a spoken corpus (the *WSC*; see Kennedy, 2001) and a written newspaper corpus (the *New Zealand English Press Corpus*, or the *NZEPC*). The *NZEPC* was comprised of samples taken every 3 years between 1996-2011 from the following newspapers: *The New Zealand Herald* and *Sunday Star Times* (Auckland), *The Dominion Post* (Wellington), *The Press* (Christchurch) and *The Southland Times* (Southland region). Pluralised forms were included in the data, and any tokens not referring to ethnic meaning were excluded (Onysko & Calude, 2014). Counterparts to Māori loanwords in English were also included in the data, to provide an overview of word choice.

Most pertinent to this thesis were findings that different loanwords exhibited different diachronic behaviour, as well as an ethnic bias in use. Over the period 1996-2011, newspaper frequencies of *Maori* were relatively stable, while *Pakeha* decreased, and *Kiwi* increased (Onysko & Calude, 2014, pp. 162-163). *Maori* was also once again the most frequent loanword by a large margin, with its average of 419.91 tokens per million words almost double that of *Kiwi* (282.46 per million words), and dwarfing *Pakeha* (21.71 per million words) (Onysko & Calude, 2014, p. 159). Higher frequencies of all three loanwords were also found in newspapers from regions with higher Māori populations; with the notable difference between the North and South Islands' results regarding *Maori* and *Pakeha* being attributed to this (Onysko & Calude, 2014).

Similar results were seen in the spoken data, where speaker and addressee ethnicity made a significant difference in frequencies of all three loanwords, especially in the case of *Maori*, which was more frequent with a Māori speaker or listener (Onysko & Calude, 2014). *Pakeha* was usually preferred by Māori speakers, often having negative connotations for Pākehā themselves (Onysko & Calude, 2014, p. 163). Additionally, it most usually occurred in Māori contexts “to emphasise bipolar ethnic differences” (Onysko & Calude, 2014, p. 164). In contrast, *Kiwi* was viewed positively by Pākehā and thus used more frequently by that demographic, especially when speaking to non-Māori listeners (Onysko & Calude, 2014).

Finally, pluralisation of the data in Onysko and Calude (2014) agreed with that of previous studies (Davies & Maclagan, 2006; Degani, 2010) in that ‘s’ suffix pluralisation of Māori loanwords is a controversial issue and in decline in written

NZE, as it did not occur in *Maori* or *Pakeha* after 2002 except in a rare proper noun and “an oral quotation” (Onysko & Calude, 2014, p. 165). However, it appears that no such stigma exists in the loanword *Kiwi*, which, when pluralised, usually takes the form *Kiwis* (Onysko & Calude, 2014).

From Onysko and Calude (2014), it is expected that the loanwords *Maori*, *Kiwi* and *Pakeha* will continue to be strong markers of national and ethnic identity in the present study. Frequencies of those mentioned here should also be similar, with *Maori* and *Pakeha* being particularly high. Regional differences in individual loanwords are predicted to vary according to regional demographics, being more prevalent in regions with high Māori populations. Zero-inflections of pluralised Māori loanwords are also predicted to occur in all loanwords except *Kiwi*.

The final and most recent study on Māori loanwords in newspapers to be discussed here is that of Calude et al. (Forthcoming), who constructed a diachronic corpus topically constrained to the theme of *Matariki* (Māori New Year¹⁹), an annual event which “is becoming increasingly salient in New Zealand society” (Calude et al., Forthcoming, p. 3). The *Matariki Corpus* was created by searching the keyword ‘*Matariki*’ in the *New Zealand Herald* and subsidiary newspapers between 2005-2016. To date, this is the only topic-bound corpus on Māori loanwords in NZE.

Most notable among their findings was that, “even when holding topic constant, the ethnicity of the writer of a given newspaper article remains a significant predictor associated with an increase in loanword use for Māori writers” (Calude et al., Forthcoming, p. 14). Pākehā authors generally used less Māori loanwords than Māori authors, even when they were writing about the same thing: *Matariki*. Correlations were also found between author ethnicity, loanword frequency, and the likelihood of a loanword occurring in an article title. Although there was no significant relationship between loanword use in an article and its title if the author was Māori, “if the author was Pākehā, every additional loanword in the article multiplied the odds of a loanword appearing in the title of their article by 1.21 times” (Calude et al., Forthcoming, p. 9).

Network analysis was also utilised in order to create networks of co-occurring loanwords within the same articles. As the loanword *Māori* was ten times

¹⁹ *Matariki* may also signify the Pleiades constellation (Calude et al., Forthcoming).

more likely to occur than the next most frequent loanword (Calude et al., Forthcoming), two networks were created: one with *Māori*, and one without²⁰. As an intra-article attractor of collocates, *Māori* was extremely productive; co-occurring in 22 or more articles with 4 loanwords²¹, and in 12 to 21 articles with 6 loanwords (Calude et al., Forthcoming, p. 11). In contrast, only 12 loanword pairs co-occurred more than 6 times (Calude et al., Forthcoming, pp. 13-14).

It was also found that regional newspapers were 1.75 times more likely to use loanwords than newspapers with a national distribution (Calude et al., Forthcoming, p. 8). Regarding individual loanwords, frequencies of the loanword *Matariki* significantly increased over the corpus timeframe.

These findings have important implications for the present study, as in a corpus bound to a single, but different, topic (Māori Language Week), it should be expected that more loanwords will occur in articles written by Māori authors, as well as in regional newspapers. However, it remains to be seen whether common loanword collocates remain the same when a corpus topic is changed, or in corpora not constrained by a single topic. There is also the possibility that loanword collocates in a specifically non-Māori context are different than those in a Māori context.

2.2.2. Māori loanwords in the television media

Although de Bres (2006) does not deal with written language in mainstream newspaper media, her study is of relevance to the present research because it explores Māori loanword frequencies in the mainstream New Zealand television media. Television media was selected for her corpus as it is representative of current linguistic norms, is a medium for linguistic change (Garrett & Bell, 1997, as cited in de Bres, 2006), and because of its prestige (see Bell, 1984). De Bres (2006) conducted a diachronic comparison of Māori lexical items as they appeared in one week of broadcasts from news programmes in 1984 and 2004 from two different stations: *Television New Zealand (TVNZ)* and *3 News (TV3)*.

Six comparable national or regional news items were selected from each year and station for analysis, along with all “‘Māori-related’ news items from each

²⁰ The loanword *Matariki* was excluded from both of these analyses.

²¹ These were: *marae*, *kapa haka* (a *haka* group, or Māori cultural performing group), *whānau* (family) and *iwi* (tribe) (Calude et al., Forthcoming, p. 13).

source” (de Bres, 2006, p. 20). This yielded 33 news items for analysis from 351 total news items (de Bres, 2006). In coding the data, lexical hybrids were excluded, although loanwords which had semantically shifted from their original Māori meaning²² were retained; as were proper nouns (de Bres, 2006). Tokens were counted in complete lexical units (as in Macalister, 2006b) to avoid conflation.

De Bres (2006) found that there was no clear increase in Māori loanword use between 1984 and 2004; frequency of use was low in both years, although *TVNZ* did use more loanwords overall than *TV3*. Unfortunately, the frequency of spoken NZE words in this study were not calculated, and so this interpretation is an estimate only. Neither can any comparison be made with average loanword frequencies in other literature (de Bres, 2006). De Bres (2006) also concluded that one week’s worth of data for each year surveyed was not sufficient for analysis.

In contrast, morphological assimilation did show a change over time. The occurrence of ‘s’ suffixes on plural loanwords was far lower in 2004 than in 1984 (de Bres, 2006, p. 24). According to de Bres (2006), “these differences suggest a move away from the morphological integration of Māori lexical items and the introduction, and/or greater acceptability, of some Māori lexical items on the news” (p. 31). Proper nouns occurred at a lower rate than in previous studies, perhaps because they were often represented lexically on screen rather than spoken; however, they were still “the greatest proportion of Māori lexical items in all three data sources” (de Bres, 2006, p. 22). In accordance with previous research, *Māori* was still the most frequent Māori loanword (de Bres, 2006).

Topic was shown to be a vital factor in frequency of Māori loanwords in mainstream television broadcasts (de Bres, 2006). In the 1984 data, the highest number of tokens occurred in a news item on the Treaty of Waitangi. Similarly, in 2004, the highest number of tokens in both *TVNZ* and *TV3* items regarded the allocation of fisheries assets by the Treaty of Waitangi Fisheries Commission. De Bres (2006) writes that,

one possible reason for this is that news items about highly specific Māori issues such as these are more likely to include a greater number of specifically Māori lexical items than news items that could be seen as more tangentially related to Māori people and culture. (p. 26)

²² i.e., *kiwi* (a flightless bird native to New Zealand) versus *Kiwi* (a New Zealander).

This topical attraction can be compared to the so-called ‘news-driven effect’ seen in Davies and Maclagan (2006). Further support for the topical attraction of loanwords can be seen in the complete lack of loanwords from the semantic category of flora and fauna in the data, which was attributed to news item subject (de Bres, 2006). Thus, this could be considered an effect of topical non-attraction. Loanwords were also rare in national and regional news items with the exception of proper nouns such as place names (de Bres, 2006). In fact, de Bres (2006) considered that use of non-proper noun loanwords (“general Māori lexical items”) was a matter of lexical choice (p. 23).

Given that there were differences in loanword use between the television stations in de Bres (2006), it might be expected that there may also be differences between newspapers in the present study. Similarly, frequencies should be higher in this thesis than on a non-topically bound corpus as it is centred around a Māori-related topic.

2.2.3. Māori loanwords in picture books

While the majority of research on written Māori loanwords in NZE has taken place in more traditional corpora such as newspapers, a different approach is taken by Daly (2007), whose work focusses on Māori loanword use in children’s picture books. Picture books are not only of inherent value in the construction of national identity in young New Zealanders, but they also have “a much higher rate of Māori loanwords in their English texts than do other domains such as newspapers and school journals” (Daly, 2013, p. 74). This is demonstrated in the *New Zealand Picture Book Collection (NZPBC)*, a corpus comprised of 13 picture books written in NZE between 1995-2005, and published by New Zealand’s Huia Publishers (Daly, 2007). Proper nouns and lexical units were considered loanwords, as in previous studies such as Macalister (2006b). The average frequency of loanwords was found to be 56 per 1,000 words, with a range of 30-90 loanwords per 1,000 words across the 13 books (Daly, 2007). This high frequency is attributed in part to repetition inherent in children’s literature, the high incidence of Māori related themes in New Zealand picture books – itself a result of a 100% rate of Māori authors in the corpus – and the fact that Huia puts an emphasis on publishing books which are a reflection of Māori experiences (Daly, 2007). Another reason for higher

loanword frequencies in this genre may have been that some picture books were originally written in Māori and then translated into English, however there was only a small difference in frequencies between original language of publication (Daly, 2007). Daly (2007) argues that the very high frequencies of Māori loanwords in picture books serve the dual roles of giving “voice to and [reaffirming the] linguistic heritage” of Māori children, as well as increasing the receptive, and possibly productive, vocabulary of non-Māori children (p. 31)²³. It may also reflect increasing use of Māori loanwords in NZE in general.

Another issue Daly (2007) considered was loanword marking, through the textual presentation thereof such as italicisation, glossing and macrons (Daly, 2007). No instances of italicisation were found in any of the picture books (Daly, 2007). While it was noted that it is Huia’s policy not to italicise loanwords, it is also suggested that “the lack of italicisation in some way allows for normalisation and acceptance of the vocabulary item” (Daly, 2007, p. 27). Regarding the glossing of loanwords, these were most likely to be “an embedded gloss, which makes the meaning of the word more transparent without the reader having to consult a glossary” (p. 28). Likewise, the majority of picture books (11) always used macrons on loanwords where possible, and only two did not (Daly, 2007). Daly (2007) noted that macron use was particularly interesting in that it was both “necessary to make a distinction between phonemes in the borrowed item, to ensure their correct pronunciation, and to ensure the correct meaning of the word is applied”, while also functioning in some cases as marking the loanword “as being different to the other New Zealand English items” (p. 29). She predicted that increased use of macrons will result in the macronisation of Māori loanwords becoming standard practice in NZE (Daly, 2007). This has interesting implications for entrenchment research, as macrons could arguably indicate both high and low levels of loanword entrenchment depending on stance taken. However, overall, Daly (2007) agrees with previous literature (e.g., Davies & Maclagan, 2006; Degani, 2010) in that the

²³ This possible increase in vocabulary was proven in Daly (2009), where repeated spoken readings of books from the *NZPBC* by parents to their children increased the receptive knowledge of Māori loanwords and of *tikanga Māori* (Māori culture) in all non-Māori participants, as well as resulting in a small increase in production of loanwords in both parents and children when a picture book was found enjoyable.

lack of marking (here, italicisation and glossing) is indicative of how well a Māori loanword is entrenched in NZE.

A more recent study by Daly (2013) restricted to just two picture books and their Māori language translations taken from the *NZPBC* found a similarly high frequency of Māori loanwords. Although not as high as in the *NZPBC*, the average of 35 and 44 loanwords per 1,000 words respectively was still much higher than the 7.7 per 1,000 word in newspapers found by Macalister (2006b). This was attributed to the prominence of Māori language and culture, and the New Zealand environment, in the two picture books (Daly, 2013). Interestingly, both books were also “written with educative purposes” in mind regarding Māori language and culture (Daly, 2013, p. 13). While these two books represent only a small sample of the *NZPBC*, they suggest that author intent may also play a role in Māori loanword frequencies in other corpus genres, such as newspapers. Moreover, while the *NZPBC* corpus as a whole is restricted to a particular genre rather than a topic, the strong presence of culturally specific topics within it provides evidence that topic-bound corpora are likely to produce higher Māori loanword frequencies.

2.2.4. Māori loanwords in other studies

Metge’s (2009) article on Māori loanwords is much more brief than other literature mentioned here, and so the summary of her work will be necessarily short. However, it is no less important, as she offers a concise insight into semantic change in Māori loanwords in their original forms, and how they are used in NZE today.

Metge (2009) demonstrates that while many Māori loanwords have become well established in NZE, much of their cultural meaning and context has been lost during the process of assimilation. Specific examples given include *utu* (revenge) and *mana* (prestige, authority)²⁴. However, despite the loss of specific cultural context, these loanwords are still markers of Māori culture and thus are seen as a positive gain for NZE (Metge, 2009). Furthermore, in contextualising certain loanwords within a framework of their original meanings and their cultural significance, Metge (2009) provides a number of loanwords with strong semantic and cultural connections. It is possible that these loanwords will have a collocational effect on one another within a corpus on Māori loanwords. Thus,

²⁴ For a discussion on how the given definitions of these loanwords are, in fact, incomplete and have been semantically reduced from the original Māori, see Metge (2009).

Metge (2009) not only suggests some possible high frequency loanwords in current NZE, but also gives important cultural and contextual information which will inform the present study.

2.3. Summary

The topic of loanwords in contact languages is a broad and wide-ranging one. Even a narrower focus such as the one presented here on Māori loanwords in NZE has been well-mined. While this literature review is far from comprehensive in regards to Māori loanwords in NZE, or even Māori loanwords borrowed into other languages elsewhere, it has sought to provide an overview of the most relevant and established research on the topic, with a particular focus on newspaper corpora. It has provided an investigation into changing loanword frequencies over time, ascertaining that there has been a clear diachronic trend for the increasing use of Māori loanwords in NZE since first contact between the Māori and English languages on these shores; as well as on entrenchment issues of listedness, markedness, and morphological assimilation where applicable. Use of Māori loanwords in NZE is complex, and may be affected variously by author or speaker identity, genre and topic. Attitudes to and perceptions of Māori loanwords are also a complex issue, as they are closely associated with concepts of identity. With these issues in mind, this thesis will now seek to address the following research questions:

- 1) How are Māori loanwords behaving in a diachronic corpus spanning 2008-2017, in comparison to previous diachronic studies on loanword frequencies in newspapers? Are they becoming more frequent and established as previous studies suggest?
- 2) Is there any difference in loanword behaviour in a corpus constrained to a specific issue related to Māori culture (Māori Language Week) than in more general non-topic-bound corpora?
- 3) By considering Māori loanword frequencies, listedness, markedness, and morphological assimilation all as variables of the single issue of entrenchment, what conclusions may be drawn about Māori loanword entrenchment in NZE?
- 4) Given that Māori loanword use is often the subject of public comment, can any explicit and implicit perceptions and attitudes to the use of

Māori loanwords be seen in a corpus bound to the topic of Māori Language Week?

- 5) What implications do the answers to these questions have for the use of Māori loanwords in NZE, to the revitalisation of the Māori language, and to the wider use of loanwords in general?

3. Methodology

3.1. Background

The *New Zealand Herald* and its subsidiary newspapers were selected as the basis of this study, as they include both regional newspapers from across New Zealand, and two large national newspapers, the *New Zealand Herald* and the *Herald on Sunday* in its collection (NZME, 2018). They are part of the NZME media company, one of the largest media companies in New Zealand (Calude et al., Forthcoming). The *New Zealand Herald* (hereafter the *NZH*) and/or other NZME newspapers have also been used extensively in previous corpus linguistics studies on Māori loanwords in New Zealand newspapers (Bauer, 1993a; Calude et al., Forthcoming; Degani, 2010; Degani & Onysko, 2010; Kennedy, 2001; Macalister, 2001; Onysko & Calude, 2014).

The first hurdle to meet was the database from which to access these newspaper articles. In the few months between the commencement of the present research and the most recent work on Māori loanwords in the *NZH* (Calude et al., Forthcoming), the *New Zealand Herald* website (<http://www.nzherald.co.nz>) was overhauled. This made searching for articles according to almost any sort of parameters impossible. After some investigation, the online newspaper database *Newztext* (Newztext Newspapers, 2018) was selected for the construction of the corpus, as it was deemed to be the most easily searchable and search-replicable database of New Zealand newspapers available at the time.

Two search terms for the construction of the corpus were used: ‘maori language week’ and ‘te wiki o te reo maori’²⁵. Neither macrons nor capitals were utilised, as capitalisation was found not to affect the search results, and all searches including macrons on the loanword *Māori* yielded zero hits.

The articles resulting from each search made up two sub-corpora. These were named the *Te Wiki sub-corpus* (hereafter the *TW sub-corpus*) and the *Māori Language Week sub-corpus* (hereafter the *MLW sub-corpus*). These sub-corpora were considered separately for the purposes of data collection and coding, and so only concern the methodology section. At the completion of all data collecting and

²⁵ Henceforth where either of these phrases occur in this thesis in all lower case, they refer exclusively to the search terms used in the corpus construction.

coding, the two sub-corpora were merged to form one corpus, the *Māori Language Week Corpus*.

A preliminary investigation of the number and range of articles on the topic of Māori Language Week from the *NZH* newspapers in the *Newztext* database revealed that the earliest entry of either search term occurred in 1999, and that article numbers per year prior to 2008 were less than 10. It was thus decided that the present study would focus on a 10-year timeframe: 2008-2017. A target size of 200 newspaper articles per search term (20 per term per year) was set as a guideline. A more detailed description of the search parameters, corpus construction and methodology of data coding now follows.

3.2. The search parameters

The corpus was constructed from *NZH* articles which were published between 1 January 2008 and 31 December 2017. Using the *Newztext* database of newspapers, each search term ('maori language week' and 'te wiki or te reo maori') was searched for using double quotation marks in all zones, including article title and article body text. Search results were sorted by score, and only *NZH* newspapers were selected. These search parameters can be seen in Figure 1 below.

The screenshot shows the 'Search Newztext Newspapers' interface. The search term is 'te wiki or te reo maori'. The search zone is set to 'All'. The search terms are set to 'Required'. The dates are set to '1 Jan 2008' to '31 Dec 2017'. The results are sorted by 'Score'. The search is performed on 'Newspapers'.

External Sites: Otago Daily Times

International: The Independent (UK)

National Business Review [Contents]: NBR Full Text 1993- NBR Index 1985-1992

Newspapers [Contents]: Fairfax NZ Herald 1998- Stuff

Newswires: 1XX Regional News BusinessDesk Radio NZ Scoop

Figure 1: Initial search parameters in Newztext (TW sub-corpus).
Image reproduced with permission.

It was noted that initial search results yielded some articles which were full-text Māori language or Māori language translations of English articles. There was also a high number of television programme guides netted by the search. To this end, a secondary search was made in the search bar at the bottom of the results page (see Figure 2) which excluded the following phrases from all article titles: ‘version’ ‘translation’ ‘no headline’ ‘couch guide’ ‘on tv today’ ‘here is the news’. This was done for the results of both sub-corpora search terms.

Restrict current search by requiring or excluding:

"version" "translation" "no headline" "couch guide" "on tv today" "here is

Select zone(s): All Author Source Title

Search

Figure 2: Secondary search parameters in Newztext.
Image reproduced with permission.

3.3. The search results

3.3.1. TW sub-corpus search results

Using the search term ‘te wiki o te reo maori’, the search yielded 118 newspaper articles. Figure 3 gives a breakdown of the articles according to year and newspaper publication. Due to the low number of results, all 118 articles were included for consideration in the *TW sub-corpus*.

Year	Source
2017: 23	The Daily Post - 62
2016: 24	Northern Advocate - 11
2015: 14	Bay of Plenty Times - 8
2014: 17	Hawkes Bay Today - 7
2013: 10	New Zealand Herald - 7
2012: 9	Wanganui Chronicle - 3
2011: 1	Midweek Chronicle Wanganui - 2
2010: 18	Taupo Weekender - 2
2008: 2	The Daily Post Sat - 2
	The Northern Advocate - 2
	Weekender Rotorua - 2
	Bay News - 1
	Dannevirke News - 1
	Hamilton News - 1
	Hastings Leader - 1
	Hawkes Bay Today Sat - 1
	Horowhenua Chronicle - 1
	Kapiti News - 1
	Napier Courier - 1
	The Northern Advocate Sat - 1
	The Whangarei Report - 1

Figure 3: Summary of article results by year and source newspaper in Newztext (*TW sub-corpus*).
Image reproduced with permission.

3.3.2. MLW sub-corpus search results

Using the search term ‘maori language week’, a total of 459 articles were netted by the search. A breakdown of these articles according to year and newspaper publication can be seen in Figure 4.

Year	Source
2017: 39	The Daily Post - 157
2016: 71	Bay of Plenty Times - 78
2015: 44	Northern Advocate - 59
2014: 50	New Zealand Herald - 56
2013: 33	Hawkes Bay Today - 49
2012: 37	Midweek Chronicle Wanganui - 8
2011: 16	Wanganui Chronicle - 8
2010: 65	Herald on Sunday - 6
2009: 65	The Northern Advocate - 5
2008: 39	Hawkes Bay Today Sat - 4
	Napier Courier - 3
	The Northern Advocate Sat - 3
	Weekender Rotorua - 3
	Bush Telegraph - 2
	Hastings Leader - 2
	Kapiti News - 2
	Katikati Advertiser - 2
	The Daily Post Sat - 2
	Bay News - 1
	Bay of Plenty Times Sat - 1
	CHB Mail - 1
	Dannevirke News - 1
	Horowhenua Chronicle - 1
	Te Awamutu Courier - 1
	The Business - 1
	The Northland Age - 1
	Wanganui Chronicle Sat - 1
	Whakatane News - 1

Figure 4: Summary of article results by year and source newspaper in Newztext (MLW sub-corpus). Image reproduced with permission.

As the number of articles yielded by the *MLW sub-corpus* search terms was higher than the target size of 200 articles, it was necessary to devise an unbiased means of excluding some of the data. An online random number generator²⁶ was utilised to select 20 random articles from each year, as per the sub-corpus target of 200 articles. In the single year that resulted in less than 20 articles, 2011, all articles were included for consideration. When an article selected by the random number generator had already been included in the *TW sub-corpus*, this article was not included in the *MLW sub-corpus* to avoid duplicates in the data, and the next

²⁶ <https://www.random.org/integer-sets/>

randomly chosen non-duplicate article was included in its place. This resulted in 196 articles for consideration in the *MLW sub-corpus*.

In addition to this screening process, it was noted that some unsuitable articles still remained in both sub-corpora. Articles in the two corpora were thus manually screened a final time for suitability for analysis. Methodology of article exclusions from the combined sub-corpora will now be described.

3.4. Article exclusions

The combined number of articles yielded from the two search terms was 314. Next, all articles with their full text in the Māori language and duplicate articles not already excluded during the initial search screening process were removed.

Duplicate articles were considered to be two or more articles with identical internal text, regardless of a difference in title, date of publication or the newspaper it was published in. Where any such duplicates occurred, the earliest article in the search results was retained (with a preference for *TW sub-corpus* articles) and all others were discarded.

Certain articles were found to have some textual overlap with other articles, while not being 100% identical. This usually occurred when an article was re-edited for use in another publication. Such articles were evaluated on a case-by-case basis, and both were retained only if enough textual difference was judged to exist to preclude them from being exact copies. In instances where a first article was identical to a second, but the second article included additional original text, only the longer article was retained.

Another unexpected issue in the corpus was the presence of articles which were guides to upcoming events or lists of awards. The loanwords in such articles were extremely difficult to categorise. In many cases, it was problematic discerning whether a token was a non-proper or a proper noun (i.e., an event or award title), or whether a longer event title was actually a codeswitch. In cases where categorising the data was too difficult, these articles were also discarded.

3.5. Codeswitching exclusions

It has earlier been established that all articles in which the full text was in the Māori language (that is, not fully in NZE) were excluded. However, there were also some examples of codeswitching in the data. Differentiating individual word

codeswitches and longer codeswitches is a difficult and contentious issue, although Poplack and Dion (2012) found that single word other-language items occurring in a receptor language are almost always borrowings and not codeswitches. For the purposes of this study, individual loanwords were defined in the corpus according to whether they occurred in lexical units rather than unique structural words (following Macalister, 2006b, p. 10). Codeswitches were deemed to be any lexical material which communicated more than one single material item or intangible concept. Section 3.7.2.1 addresses distinguishing between loanword types in more detail.

As codeswitches were judged to conflate the loanword data as well as overall text word counts, any instance of codeswitching text was discarded from the articles. In the few cases where distinguishing between a loanword and codeswitching was unclear, that occurrence was considered a codeswitch and discarded from the article text.

3.6. Irregular article retentions

Another unexpected issue in the data was the small number of articles that were collections of letters to the editor. In such cases, only the relevant letter (in which the search term occurred) was retained, and all other text was discarded from that article. If more than one relevant letter to the editor occurred within the same article, the two letters were separated into different files and all other text from the original article was discarded.

Finally, it was noted that there were five files in the data that did not produce any loanword tokens after codeswitches were removed. These were retained in the data.

After all exclusions and retentions were made, 99 files remained in the *TW sub-corpus*, and 191 files remained in the *MLW sub-corpus*. The two sub-corpora were then combined into the single *Māori Language Week Corpus*, with a total of 290 articles and 108,925 words for analysis.

3.7. Coding the data

For each article included in the corpus, individual text-only (.txt) files were created into which all retained textual data was copied. Articles were then assigned file names which included information about which search they occurred in, their date

of publication, which newspaper they were published in, and given a number pertaining to their entry into the corpus by newspaper and year. This was done according to the following convention:

Table 2: Example of article file name coding

Sub-corpus	Year	Day	Month	Newspaper publication code	Position of article in corpus entry of that year, by publication
TW	2016	04	JUL	HBT	002

Thus in the example given in Table 2, the file TW_2016_04_JUL_HBT_002 is an article resulting from the search term ‘te wiki o te reo maori’ (from the *TW sub-corpus*) that was published on 4 July 2016 in the *Hawke’s Bay Today* (*HBT*), and was the second article from that publication from 2016 recorded in the corpus (for a full explanation of newspaper codes, see Section 4.3).

Once articles were given file names, data about the articles (termed metadata) was coded for, followed by coding of the loanwords (termed loanword data).

3.7.1. Coding the metadata

Article metadata data that were recorded included information on:

- day, year and month of publication
- article author, where possible; including name, age, gender and ethnicity
- newspaper of publication
- whether the newspaper was regional or national
- newspaper section
- page number on which the article was published
- article word count
- word count of any text excluded as codeswitching or irrelevant letters to the editor
- word count of titles
- date of article retrieval from *Newztext*
- which search term netted the article
- loanword frequencies in all articles, and article titles

Regarding title metadata, in files created from letters to the editor articles, a title was considered to be the title of the actual letter, and not the article title. Any subtitle data was considered to be part of the body of the article word count, not the title word count.

An attempt was also made to code each article according to its main subject or topic. In some cases, these were clear (such as events taking place during Māori Language Week, the history of Māori Language Week, or the revitalisation of the Māori language), but in the majority of articles, such distinctions were blurred and difficult to make. As such, these categories were not coded for; however, topics of specific articles will be commented upon where analysis requires it.

3.7.2. Coding the loanword data

All Māori loanwords were manually extracted from the articles and then checked using LancsBox (Brezina, McEnery, & Wattam, 2015) for accuracy of frequency counts, before being coded for which article they occurred in. Each individual loanword was then manually annotated as to whether or not it was included in the loanword frequency data according to the following categories.

With the exception of the types *Maori*, *Pakeha*, *Kiwi* and *Matariki*, which were retained to gain a better understanding of how these high frequency proper noun loanwords are used, all proper noun loanwords were excluded from frequency counts.

Macron use did not occur in any of the data. Thus, none of the loanwords in the corpus have macrons. However, Māori loanwords were used in the writing of this thesis: hence the Māori language, and Pākehā; but the loanwords *Maori* and *Pakeha*.

3.7.2.1. Distinguishing between loanword types

As was mentioned earlier, individual loanwords were distinguished as occurring in lexical units rather than as unique structural words (see Macalister, 2006b, p. 10). For example, *haka*²⁷ was considered one distinct lexical unit except where it occurred in *kapa haka* (a *haka* group, or Māori cultural performing group); then, *kapa haka* was considered one lexical unit.

²⁷ A traditional Māori war dance; see Appendix A for full definition.

Distinctions between orthographically identical loanword types were made where semantic differences existed, for example: *hapu* (pregnant) and *hapu* (tribe); *wananga* (seminar, conference) and *wananga* (a tertiary institution that caters for Māori learning needs)²⁸.

In certain cases, the only thing distinguishing one lexical unit from another was the presence of a second loanword, while in other cases the presence of that second loanword was optional. Here semantics played an important role. One of the most complex examples was that of *kura kaupapa (Maori)*. *Kura*, *kaupapa* and *Maori* were all categorised as distinct types when they occurred in isolation. However, different combinations resulted in different lexical units and so were categorised accordingly. This resulted in the categorisations of Table 3.

A full list of all loanword types and their definitions, along with macrons where applicable, is given in Appendix A. Much care was taken to be as consistent in coding and as accurate to the Māori language as possible, and any error remains the sole fault of this researcher.

Table 3: Distinction of loanword types in *kura*, *kaupapa*, *Maori*

Loanword type	Definition (adapted with permission from Moorfield, 2018)	Notes
<i>kura</i>	school	Any following lexical units not mentioned in this table constituted a variant of <i>kura</i>
<i>kaupapa</i>	purpose, topic	
<i>Maori</i>	(Māori) indigenous person of <i>Aotearoa</i> New Zealand. See Moorfield (2018) for further definitions and usages Also: the Māori language (Deverson & Kennedy, 2005)	
<i>wharekura</i>	Māori language secondary school	
<i>kura kaupapa (Maori)</i>	Māori language primary school	The following presence of <i>Maori</i> was optional in this type; as was preceding <i>te</i>
<i>kaupapa Maori</i>	(<i>kaupapa Māori</i>) Māori customary approach	The following presence of <i>Maori</i> was obligatory for this type

3.7.2.2. Variants of loanword types

Variation in the loanword data included regional or dialectal variants, pluralisation, compounds, lexical innovations, and orthographic variation or error. In each case,

²⁸ With macrons, these loanwords should be spelled *hapū* and *wānanga*.

such variants were considered as belonging to the standard type (that is, the type which occurred most frequently in the corpus). Three examples of variation in loanword types that occurred in the corpus are given in Table 4.

Table 4: Examples of loanword type variants

Loanword type	Variants
<i>Maori</i>	<i>Maori</i>
	<i>Mori</i>
	<i>Maoris</i>
	<i>non-Maori</i>
<i>powhiri</i>	<i>powhiri</i>
	<i>pohiri</i>
<i>kapa haka</i>	<i>kapa haka</i>
	<i>kapahaka</i>

However, it was noted in the data that a small number of certain variants had unexpectedly high frequencies. These were *te reo* (the language, the Māori language) and its variant *reo* (language, dialect); and *kohanga reo* and its variant *kohanga* (both referring to Māori language preschools). In many cases, the variants used appeared to represent a lexical choice made by the author or interviewee of the article. These high frequency variants were therefore re-categorised into separate and distinct loanword types, in order to better investigate the behaviour surrounding their use.

3.7.2.3. Marked loanwords

It was noted that many loanwords in the corpus were marked (for a full explanation of marked loanwords, refer to Section 6.2.1). All loanwords included in the frequency data were coded according to whether they were marked or unmarked; if so, how they were marked; and their corresponding article metadata.

3.7.2.4. Semantic coding

Loanwords were coded according to their semantic field, adapted from categories used in Macalister (2006b). See Section 5.1.3. for more details.

3.7.2.5. Loanwords in article titles

Frequency data on loanwords in article titles were recorded, along with their corresponding article metadata.

3.7.2.6. Loanword exclusions

The following sub-sections explain criteria for loanwords excluded from the overall frequency count in the corpus. Such exclusions were: proper nouns (except for *Maori*, *Kiwi*, *Pakeha* and *Matariki*), search terms, Māori Language Week year themes, and other unclear tokens.

3.7.2.6.1. Proper nouns

All proper noun loanwords not mentioned in the preceding section were excluded from final loanword frequency counts. Such proper noun exclusions were:

- Names of people, including first and last names
- Names of places, including countries (*Aotearoa*), cities, towns, regions and electorates (*Whangarei*, *Te Awamutu*, *Waikato*); and names of marae, pa, event centres, tourist centres and geographical bodies
- Names of organisations, including businesses, libraries, educational institutions, commissions, trusts and government departments, as well as names of Māori tribes or sub-tribes
- Any other proper noun, including: days of the week, names of awards, award categories, job titles; book, song and album names; lyrics; act and fund names; and treaties

3.7.2.6.2. Search terms

The search terms used to construct the corpus were excluded from loanword frequencies to avoid conflating the data. Thus, any occurrence of the loanword *Maori*, *te reo* or *wiki* in the search term phrases ‘te wiki o te re maori’ or ‘maori language week’ was not counted as a loanword. The search terms themselves were considered one lexical unit and recorded, along with their article metadata, for reference purposes. In cases where *te* was omitted in the search term ‘te wiki o te reo maori’, it was counted as a variant of that same type.

3.7.2.6.3. Māori Language Week year themes

It was noted in the data that for each year, a different theme was allocated to Māori Language Week. In order to prevent these from being confused as other proper nouns or codeswitches, or from conflating the loanword frequencies, year themes

were not included in loanword frequencies. However, their occurrences and article metadata were recorded for reference purposes.

Year themes usually, but not always, constituted a phrase. Thus, they followed the same exclusionary rules as described for the search terms above. As an example, in 2011, the theme for Māori Language Week was *manaakitanga* (hospitality). *Manaakitanga* was recorded as a loanword in the frequency data only where it was not used to refer to the theme of that year’s Māori Language Week theme.

3.7.2.6.4. Unclear meanings and other exclusions

Any loanword with unclear contextual meaning was excluded from the loanword frequencies. There was 1 type which was excluded because of this: *mirimiri*.

Loanwords that occurred in email addresses, websites, phone numbers, and hashtags with loanwords in them were not counted in the loanword data.

Regarding codeswitching, a total of 1,202 words occurring in 52 articles were considered to be codeswitches. These codeswitches varied in length from 2 words to 310 words in length (a full article translation), with the average length being 23.13 words. A summary of the codeswitching data is seen in Table 5. As was established earlier, all instances of codeswitches were excluded from the corpus data.

Table 5: Summary of words excluded as codeswitching

	Words excluded	No. of articles with exclusions	No. of articles without exclusions
<i>TW sub-corpus</i>	835	20	79
<i>MLW sub-corpus</i>	367	32	159
Total words excluded	1,202	52	238

4. The data

4.1. Overview of the corpus

The corpus in its final form comprised of 290 NZE articles on the topic of Māori Language Week, all taken from *NZH* newspapers. The size of the corpus was 108,925 words and 10,535 types²⁹. The 3,795 Māori loanword tokens that occurred in the corpus were distributed across 186 loanword types, as shown in Table 6. All loanword types were non-proper nouns apart from the four exceptions previously mentioned. See Appendix A for loanword types, their definitions and frequencies.

Table 6: Types and tokens in the corpus

	Types	Tokens
Māori loanwords	186	3,795
Non-Māori words	10,349	105,130
Entire corpus totals	10,535	108,925

4.2. The articles

The diachronic distribution of the number of articles in the corpus can be seen in Table 7 below. In the end, the number of articles taken from each year varied greatly, from the lowest (17) in 2011 to the greatest (43) in 2016. The average number of articles per year between 2008-2017 was 29 articles.

Table 7: Diachronic article, type and token data

Year	No. of articles	Loanword types	Total corpus types	Loanword tokens	Total corpus tokens
2008	22	62	1,900	286	7,334
2009	20	36	1,851	276	6,900
2010	26	21	2,358	166	8,586
2011	17	28	1,630	202	5,750
2012	30	58	2,329	440	10,679
2013	28	60	2,850	366	11,409
2014	36	56	2,831	602	14,623
2015	33	58	2,797	448	12,498
2016	43	61	3,297	596	16,536
2017	35	52	3,156	413	14,610
Totals	290	186	10,535	3,795	108,925

²⁹ Word and type frequencies were calculated using LancsBox (Brezina, McEnery, & Wattam, 2015).

Variation in number of articles per year is reflected in the frequency of loanword types and tokens per year, as well as the overall loanword token count per year. This will be discussed in greater detail in sections to follow.

4.3. The newspapers

A total of 21 newspapers provided the foundation for the corpus. Regrettably, while North Island and national newspapers were included in the data, the database *Newztext*, from which the newspapers were sourced, excluded any South Island-origin newspapers. Furthermore, while the regional newspapers were distributed across a wide geographical range (see Table 8), no articles from the Auckland, Gisborne or Taranaki regions arose in the search results. Over half of all articles originated from the Bay of Plenty region, while the majority of articles came from either the Bay of Plenty (*Bay of Plenty Times*, *The Daily Post*), Hawke’s Bay (*Hawke’s Bay Today*), Northland (*The Northern Advocate*) or were a nationwide newspaper (*New Zealand Herald*, *Herald on Sunday*). A representation of the percentages of distribution of newspaper articles in the corpus by source region can be seen in Figure 5.

Table 8: Newspaper of publication article data

Code	Newspaper name	Region	No. of articles
AGE	The Northland Age	Northland	1
BN	Bay News	Bay of Plenty	1
BPT	Bay of Plenty Times	Bay of Plenty	40
DVN	Dannevirke News	Manawatu-Wanganui	1
HBT	Hawke’s Bay Today (including HBT Saturday)	Hawke's Bay	27
HCH	Horowhenua Chronicle	Manawatu-Wanganui	1
HNW	Hamilton News	Waikato	1
HOS	Herald on Sunday	National	3
HSL	Hastings Leader	Hawke's Bay	1
KN	Kapiti News	Wellington	2
NPC	Napier Courier	Hawke's Bay	2
NZH	New Zealand Herald (including NZH Saturday)	National	37
TAC	Te Awamutu Courier	Waikato	1
TB	The Business	National	1
TDP	The Daily Post (including TDP Saturday)	Bay of Plenty	117
TNA	The Northern Advocate (including TNA Saturday)	Northland	40
TWK	Taupo Weekender	Waikato	2
TWR	The Whangarei Report	Northland	1
WCH	Wanganui Chronicle (including WCH Midweek)	Manawatu-Wanganui	7
WNW	Whakatane News	Bay of Plenty	1
WRT	Weekender Rotorua	Bay of Plenty	3
			290

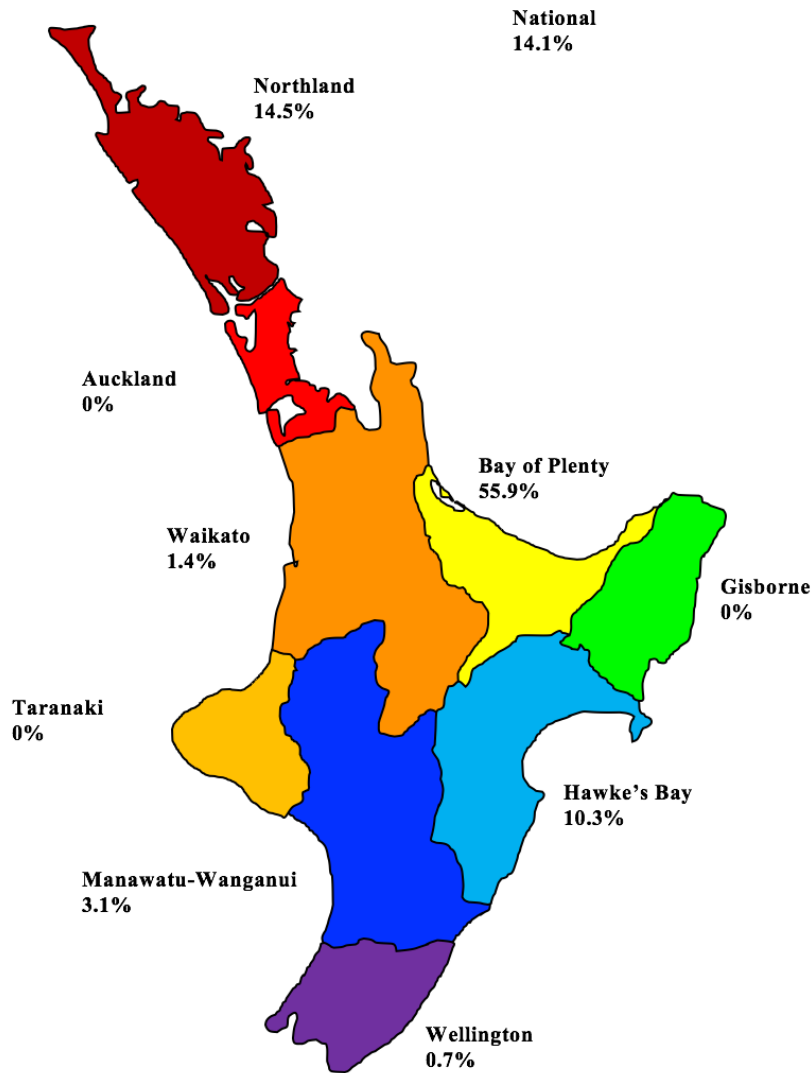


Figure 5: Map of newspaper article distribution (percentage of corpus)

This geographical distribution of the newspaper articles was far narrower than expected. Investigation as to the reason for this revealed that no South Island newspapers were included in the *Newztext* database for *NZH* newspapers. In fact, no definitive list of source newspapers for the entire time frame of the corpus was available as the newspapers *Newztext* receives change from month to month. A query to *The Knowledge Basket* (which oversees *Newztext*) gained the following reply: “Unfortunately the NZME and STUFF, particularly, change things quite often. So its [*sic*] a moving target” (D. Keet, personal communication, February 27, 2018).

Thus, although 14.1% of the corpus is comprised of articles with a national distribution, the present research cannot make any definitive statements regarding Māori loanword use nationwide; nor can it do so in regards to use in the South Island, and whether it is different from that in the North Island. However, a regional comparison of articles by regional source is still possible.

4.3.1. The newspaper authors

Once articles without a given author, or articles with more than one author, were removed from consideration, there were only 161 articles in the corpus with a named author. These 161 articles were written by 110 authors.

Although author information pertaining to gender, age and, in particular, ethnicity was sought after, regrettably these were ultimately unobtainable due to practical constraints. Thus, none of these variables and their influence on Māori loanword usage was able to be explored in the present study. However, ethnicity of newspaper readership demographics will be able to shed some light on this issue in pursuant sections.

4.4. The search term data

The search terms used to source the articles for the corpus occurred in the data a combined total of 520 times; with ‘te wiki o te reo maori’ having a frequency of 124, and ‘maori language week’ a frequency of 396 (see Table 9 for diachronic distribution).

Table 9: Search term frequencies

Year	Frequency		
	‘te wiki o te reo maori’	‘maori language week’	Combined
2008	2	34	36
2009	0	29	29
2010	5	44	49
2011	2	22	24
2012	10	43	53
2013	11	30	41
2014	22	45	67
2015	20	46	66
2016	30	57	87
2017	22	46	68
Total	124	396	520

4.5. The Māori Language Week year theme data

Māori Language Week year themes were recorded despite being excluded from loanword frequency data. These year themes, along with their translations and frequency of occurrence in the corpus are given below (Table 10). While in most instances the year themes occurred in their respective years, this was not always the case. A very small number of year themes referred to a previous year's theme. However, year of occurrence was not recorded for any of these phrases. These data are more for reference purposes, and will be dealt with more in detail only where necessary in following sections.

It is important to mention that these year themes and their translations were retrieved from data in the corpus, and so some translations here may not be strictly accurate to official translations given by Te Taura Whiri i te Reo Māori (the Māori Language Commission). Additionally, while it is suspected that the first Māori Language Week to have a theme was in 2007, with the theme of *tāpoi* (tourism), only one official reference to this was obtainable (a brief mention in Te Taura Whiri i te Reo Māori, 2007) and so this claim requires confirmation.

Table 10: Year themes and frequencies

Year	Year theme	Translation	Corpus frequency
2008	<i>te reo i te kainga</i>	Māori language in the home	5
2009	<i>te reo i te hapori</i>	Māori language in the community	4
2010	<i>te mahi kai</i>	the language of food	11
2011	<i>manaakitanga</i>	hospitality (see Appendix A)	6
2012	<i>arohatia te reo</i>	cherish the language	6
2013	<i>nga ingoa Maori</i>	Māori names	5
2014	<i>te kupu o te wiki</i>	the word of the week	9
2015	<i>whangaihia te reo (Maori) ki nga matua</i>	nurture/encourage the language in parents	4
2016	<i>akina te reo</i>	give te reo Māori a go	14
2017	<i>kia ora te reo (Maori)</i>	let the Māori language live	9
			73

5. Loanword use

5.1. Linguistic characteristics

5.1.1. Loanword frequencies

From a total of 108,925 tokens and 10,535 types in the corpus, 186 Māori loanword types were found³⁰. 3,795 Māori loanword tokens were extracted. Māori loanwords were found to be present in the corpus at a rate of 34.84 tokens per 1,000 NZE words. The type *Maori* alone occurred at a rate of 15.14 tokens per 1,000 words. When that same type was excluded, the average loanword frequency per 1,000 words was 19.71.

Of all the loanword types occurring in the corpus, 36 were considered to have a high frequency of 10 or more tokens, following Poplack et al. (1988, p. 55). These are given in Table 11 (proper nouns are in bold font).

As the data in Table 11 demonstrate, just 19.35% of loanword types account for over 90% (91.38%) of the loanword tokens in the corpus. Further analysis of loanword types can be seen below (Figure 6), where the occurrence of each of the 186 types is broken into bands of frequency.

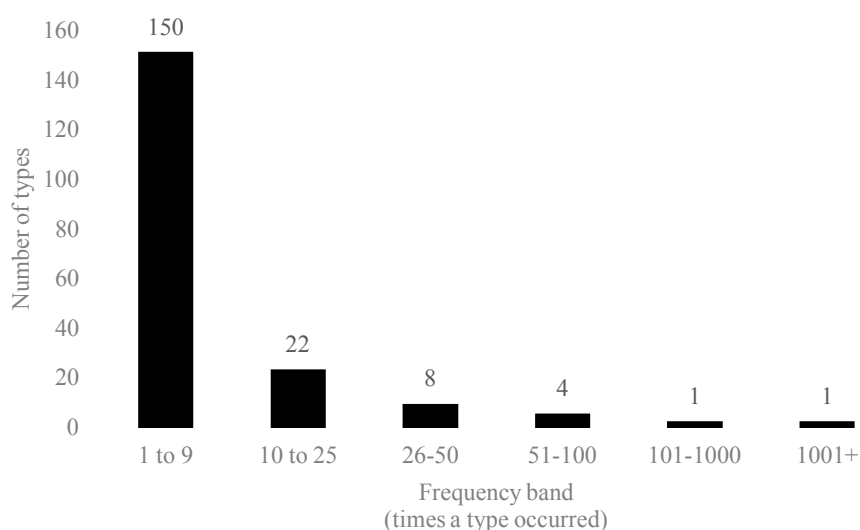


Figure 6: Loanword types by frequency band

³⁰ For a complete list of all loanwords, along with their English translations, see Appendix A.

Table 11: Loanwords with a frequency of 10 or greater

Rank	Loanword	Frequency
1	Maori	1,649
2	<i>te reo</i>	949
3	<i>iwi</i>	82
4	<i>reo</i> (no preceding <i>te</i>)	59
5	<i>whanau</i>	55
6	<i>marae</i>	51
7	<i>kapa haka</i>	47
8	Pakeha	39
9	Kiwi	37
10	<i>kia ora</i>	33
11	<i>kohanga reo</i>	32
12	<i>kura</i>	30
13	<i>kura kaupapa (Maori)</i>	28
14	Matariki	26
15	<i>haka</i>	23
16	<i>taonga</i>	23
17	<i>manaakitanga</i>	21
18	<i>hangi</i>	20
19	<i>kohanga</i>	19
20	<i>waiata</i>	19
21	<i>whakapapa</i>	18
22	<i>kai</i>	17
23	<i>kaumatua</i>	17
24	<i>tamariki</i>	17
25	<i>hapu</i> (tribe)	16
26	<i>kaupapa</i> (not <i>kura kaupapa</i> , <i>kaupapa Maori</i>)	16
27	<i>korero</i>	16
28	<i>kuia</i>	15
29	<i>moko</i> (<i>puna</i>)	15
30	<i>tikanga</i> (not <i>tikanga Maori</i>)	14
31	<i>powhiri</i>	13
32	<i>pepeha</i>	11
33	<i>waka ama</i>	11
34	<i>hui</i>	10
35	<i>moko</i> (tattoo)	10
36	<i>tangi</i> (<i>hanga</i>)	10
	Tokens in the top 36	3,468
	Tokens in the corpus	3,795

As the bulk of this corpus' loanword tokens can be found in the top 36 types, so too can the overwhelming majority of loanword types be found in the frequency band of less than 9 occurrences. In fact, Figure 6 shows a near exponential, semi-Zipfian frequency distribution of the data. Zipf's law, or the rank-size law, dictates that "the frequency (f) of a word is inversely proportional to its statistical rank (r)" (Fagan & Gençay, 2011, p. 139).

While several studies deal with Māori loanwords in a corpus made up of NZE newspapers (Davies & Maclagan, 2006; Degani, 2010; Kennedy, 2001; Macalister, 2006b; Onysko & Calude, 2014), only two of those address overall corpus loanword frequencies. Thus, a comparison of the frequency data in the present research with those two studies will now be made.

The first study to be addressed is that of Kennedy (2001), whose work was with the *WWC* (*Wellington Corpus of Written New Zealand English*) compiled by Bauer (1993b) and the *WSC* (*Wellington Corpus of Spoken English*) compiled by Holmes et al. (1998). Because this thesis is only concerned with written loanword data, frequencies will be compared with those from the *WWC* only.

The first major point of difference the *WWC* has with the present study is that its corpus word count is much larger, at almost ten times the size. The *WWC* was constructed from a variety of mixed-genre texts including both fiction and non-fiction, of which only approximately 18% were taken from the press (Kennedy, 2001, p. 79).

The second study compared here is that of Macalister (2006b). Dealing with a corpus of almost 5.5 million words, his is the largest corpus-based study of written Māori loanwords in NZE to date³¹. The corpus was compiled from three sources: newspaper articles (69.6%), *School Journals* (19.4%), and parliamentary *Hansard* papers (11%).

It should be stated that Macalister (2006b) considered both common and proper nouns in his study of Māori loanwords, as did Kennedy's (2001) analysis of the *WWC*. With the exception of four proper nouns (*Maori*, *Pakeha*, *Kiwi*, *Matariki*), the present study only counts common nouns. Additionally, Kennedy

³¹ Note that Onysko and Calude (2014) dealt with a corpus of 6 million words; however, only 5 million of those were written, while 1 million words were spoken. It is also possible that Davies and Maclagan (2006) had a larger corpus than Macalister (2006b), but no word count was given as they were concerned with article counts and not word counts, and so no comparison can be made.

(2001) dealt with lemmatised loanwords, while both Macalister (2006b) and the present study considered a loanword as a lexical unit. Table 12 provides a summary of comparable data between the three corpora.

Table 12: Comparison of loanword frequency data with Kennedy (2001) and Macalister (2006b)

Corpus	Kennedy (2001)	Macalister (2006b)	Present study
Year range	1986-1990 (see Bauer 1993a, p. 1)	1850-2000*	2008-2017
No. of texts	500 mixed-genre written texts	n/a	290 newspaper articles
Percentage of press texts in corpus	18%	69.6%	100%
Loanword types	1,001	n/a	186
Loanword tokens	5,952	n/a	3,795
Total types	43,755	n/a	10,535
Total tokens	1,015,371	5,422,546 overall (3,773,149 newspapers)	108,925
Loanwords per 1,000 words	5.86	6 whole corpus average; 8.8 in the year 2000; 7.7 in the year 2000 newspapers	34.84

*Note that the year range of 1850-2000 uses indicator years at intervals of 30 years; thus data in Macalister (2006b) are actually from 1850, 1880, 1910, 1940, 1970 and 2000.

The most immediately obvious difference between the three corpora, beyond the aforementioned discrepancies in corpus size and genre composition, is loanword frequency. Despite the differences in timeframes, both earlier studies give loanword frequencies of approximately 6 per 1,000 words (Kennedy, 2001, p. 68; Macalister, 2006b, p. 12). This figure is dwarfed by the present data, where loanwords accounted for on average 35 words per 1,000.

In Macalister (2006), the frequency of loanword per 1,000 words varies according to genre of text. Newspaper loanword frequencies are slightly higher than the average, with 7.7 loanwords per 1,000 words, while *School Journals* fall below, with 5.2 per 1,000. However, it is the *Hansard* parliamentary debate transcripts that are most notable here. From 1910 onwards, frequencies in *Hansard* were more than double their written counterparts. This is due to the inclusion of debates on Māori issues, which had a “higher-than-average proportion of Māori words” (Macalister, 2006b, p. 13).

It is important to qualify that the *Hansard* transcripts were originally spoken, unlike the newspapers and *School Journals*. Nevertheless, this discrepancy in frequency according to topic, what Davies & Maclagan (2006, p. 77) term the

‘news-driven effect’, can also be seen elsewhere. Topical issues are of vital importance when measuring Māori loanword frequency in newspapers (Davies & Maclagan, 2006). Loanword use in newspapers has been shown to often be restricted to Māori topics (Degani, 2010, p. 192). Topic also influences Māori loanword frequencies in mainstream television broadcasts (de Bres, 2006). A higher number of Māori loanword tokens occurred in news items on the Treaty of Waitangi, and on the allocation of fisheries assets by the Treaty of Waitangi Fisheries Commission, in 1984 and 2004 respectively, than any other news item in a study on Māori lexical items in NZE conducted by de Bres (2006, p. 26). A purported reason for this is that specifically Māori issues attract more specific and frequent Māori loanwords (de Bres, 2006, p. 26). Thus, the strikingly high frequency of Māori loanwords in the present corpus is likely attributable, at least in part, to being bound to the topic of Māori Language Week.

It should be acknowledged here that there are two other possible explanations for the high frequency of Māori loanwords in this study: 1) an increase in loanword usage over time since the earlier studies were conducted; and 2) author ethnicity. The first issue will be considered in due course. Regarding the second issue, it has been demonstrated that speaker and listener ethnicity have a strong effect on loanword frequencies in spoken data (de Bres, 2006; Kennedy, 2001; Onysko & Calude, 2014) and written data (Davies & Maclagan, 2006) on Māori loanwords in NZE. Speakers who self-identify as Māori have also been shown to use far more Māori words in their speech than Pākehā speakers or speakers of other ethnicities, as a way of marking ethnic identity and stance towards the use of the Māori language (Kennedy, 2001, p. 77). Regrettably, such sociolinguistic data were not available to the present research (refer back to Section 4). However, given that Māori loanwords are documented markers of Māori identity, it is likely that in a topic-bound corpus on Māori Language Week, author ethnicity is a factor in the high frequency of Māori loanwords, and should be further explored in the future.

5.1.2. Loanword frequencies in article titles

Loanword frequency in article titles was also analysed. From a total of 290 articles in the corpus, 167 had loanwords in their titles. Table 13 gives a breakdown of this data.

Table 13: Title loanword data per article

No. of loanwords per article title	No. of articles
3 loanwords	2
2 loanwords	22
1 loanwords	143
0 loanwords	123
	290

Over half of all articles in the corpus included one or more loanwords in their titles. However, a clear majority (49.3%) of articles with loanwords in their titles had just one loanword. Only two articles in the entire corpus had three loanwords in the title, while 22 articles had two title loanwords.

No title loanwords were coded as marked (see Section 6.2), although three did occur in quotation marks to indicate that they were the focus words of the articles.

In terms of type frequencies of title loanwords, 193 title loanword tokens were distributed across 24 loanword types. Additionally, only five loanword types occurred more than once. These most frequent title loanword types can be seen in Figure 7.

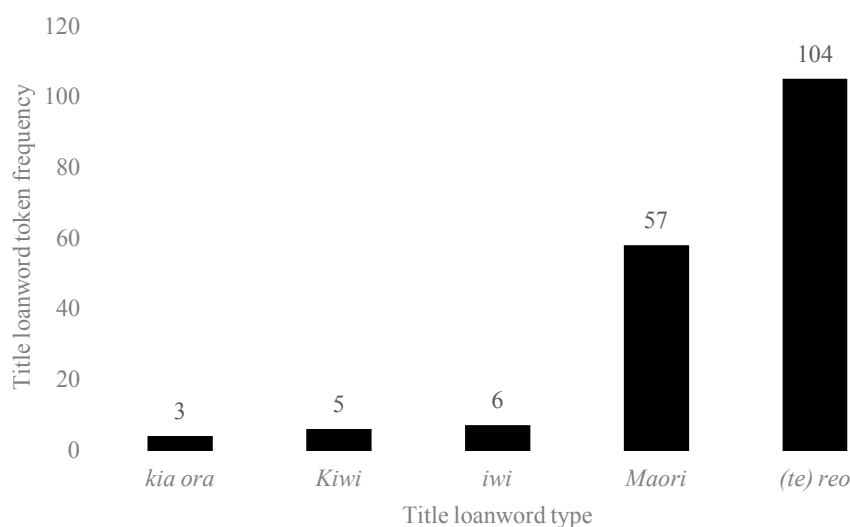


Figure 7: Frequency of most common loanword types in article titles

In line with overall corpus frequencies, *Maori* and *te reo* were the highest frequency loanword types in article titles. These two types alone accounted for 83.42% of all loanword titles in the corpus. Interestingly, although *te reo* had almost half as many

tokens in the corpus as *Maori* (*te reo*: 949; *reo*: 59; *Maori*: 1,649) the number of *te reo* tokens in article titles³² was nearly double that of *Maori*. Moreover, these two types only co-occurred as ‘*te reo Maori*’ in the titles of 14 articles.

The next most frequently occurring title loanwords were *iwi*, *Kiwi* and *kia ora*. The result for *iwi* is unsurprising given that it was ranked the third most frequent loanword in the corpus, with 82 tokens; however, it only had a frequency of 6 loanword tokens in article titles. In contrast, *Kiwi* and *kia ora* were not anticipated to have such a frequent presence. It could be that, although not the most common loanwords in the present study, they are perceived as being more recognisable by readers than other, more frequent loanwords.

5.1.3. Semantic field

An analysis of the semantic domain or field of loanwords was made by adapting the categories used in Macalister (2006b, p. 9). This included considering *Maori* as a separate semantic type from other proper nouns. Two categories (place names and person names) were not included in the loanword data and were thus excluded as semantic categories; while two new categories were added for the purpose of the present research. These were: formulaic expressions, and (*te*) *reo*. In the (*te*) *reo* semantic category, *te reo* and *reo* tokens were combined due to their high frequency. A breakdown of semantic field categories of loanword types and tokens is given in Table 14.

Table 14: Type and token frequencies by semantic field

Semantic field	No. of types	No. of tokens
Flora and fauna	15	17
Formulaic expressions	13	51
Proper nouns - <i>Maori</i>	1	1,649
Proper nouns - Other	3	102
Material culture	46	315
Social culture	106	653
(<i>te</i>) <i>reo</i>	2	1,008
	186	3,795

³² For the purposes of frequency count of loanwords in titles, the type *reo* was combined with *te reo*. This did not affect the interpretation of figures: *reo* only occurred in 4 article titles.

Maori and *te reo* have already been established as being the most frequent loanwords in the corpus. Discounting these, social culture was the most prevalent semantic field of loanwords by both type and token. This was followed by material culture, again in both type and token frequencies. With only three types, ‘other proper nouns’ had a frequency of 102 tokens; a lower figure than anticipated. The categories flora and fauna, and then formulaic expressions, had the least frequent types, at 15 and 13 types respectively. However, formulaic expressions had the higher number of tokens (51), which can be accounted for by the inclusion of *kia ora*, a high frequency type with 33 tokens. A large number of *kia ora* tokens occurred in the *BPT*³³ data as part of a campaign to increase the use of the loan in daily commercial transactions in Tauranga. ‘*Kia ora te reo Māori*’ was also the theme of Māori Language Week 2017. Although year themes were excluded from the data, both of these topics attracted the use of *kia ora* tokens.

Generally, formulaic expressions were utilised as expressions of identity or support of the Māori language. The farewell tag *nga mihi (nui)* was used to sign off two letters to the editor, and one response to a letter, by authors to position themselves as allies of the Māori language in cases where their support was called into question. Formulaic expressions were also given as encouraging examples of simple ways to use more Māori language in daily life, as below:

(1) “It may be as simple as saying ‘**morena**’ every morning or saying ‘**kia ora**’ instead of thank you. Be brave and give something a go” (TW_2017_12_SEPT_TDP_001).

Ka pai was another formulaic expression used to express Māori identity, as is the case in (2), where the Māori language and identity are explicitly linked by the author, who concludes their discourse with *ka pai*; thus positioning themselves as having that same Māori or “dual heritage” identity.

(2) “These days, wanting to speak te reo Maori seems to be driven by a growing unease with the fact that worlds are hidden from view without the language, that being Maori and not being able to speak the language limits the experience of being born in this time and place and being lucky enough to have a dual heritage. **Ka pai?**” (MLW_2012_24_JUL_HBT_003).

³³ See Section 4 for a complete list of newspapers in the corpus and their abbreviations.

Flora and fauna loanwords were the loanwords least likely to occur in the data by token, and tended to cluster together. See (3) below, and (4), in which bracketed translations are given by the author.

(3) "...yesterday Flaxmere College staff and students celebrated with a boil-up of pork bones, potatoes, **kumara** and **puha** for lunch" (MLW_2017_16_SEPT_HBTS_001).

(4) "Mr Royal has researched Maori food and developed his own company, Kinaki Native Herbs, supplying traditional Maori food products to restaurants and food producers worldwide. He uses **pikopiko** (native bush fern), **horopito** (native bush pepper), **kawakawa** (native bush basil), **piripiri** (another bush pepper) and **hakeka** (native ear fungus) in many of his recipes" (TW_2010_26_JUL_TDP_004).

All 17 flora and fauna tokens were distributed across just seven articles, with seven tokens occurring in one article alone, and four in a second. Thus, flora and fauna loanwords in the data can be stated to occur only when the topic dictates. When they do occur, they will generally tend to attract other loanwords of the same semantic category. This is supported by research in loanword use in the mainstream television media, where no flora and fauna words occurred, due to them being a rare subject of news reports (de Bres, 2006, p. 23).

Returning to overall semantic field frequencies of loanwords in the corpus, distribution of these data by semantic field can be seen in Figures 8 (types) and 9 (tokens). Social culture is clearly the most common semantic field in regards to both loanword types and tokens in the data, even when proper nouns and (*te*) *reo* are ignored. This domain accounted for 60.22% of all loanword types, and 89.91% of all loanword tokens (including proper nouns and (*te*) *reo*). This is line with findings that social culture loanwords in newspapers have become more prevalent in NZE in recent years (Macalister, 2006b, p. 16).

Unfortunately, comparison with Kennedy (2001) proved difficult in the realm of semantic field, as his categories were as follows: flora and fauna, place names, names of persons, structural words, and miscellaneous words. However, it was noted that flora and fauna loanword types have risen 8.06% from what they were in the *WWC* in 1993 (5.7%), and that loanword tokens have fallen from 5.2% to 0.45% in the present data.

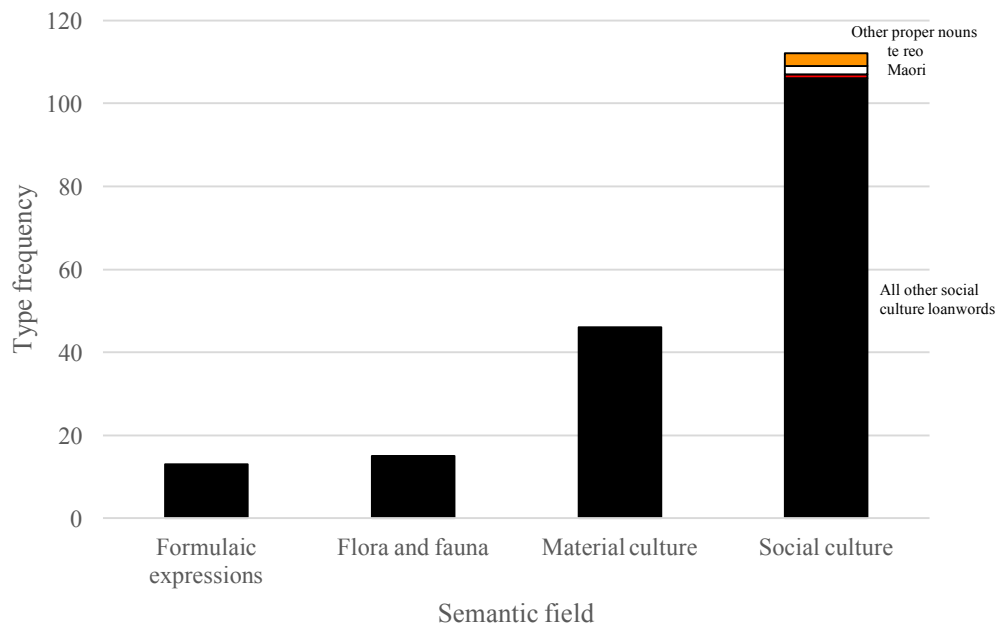


Figure 8: Percentage of loanword types by semantic field

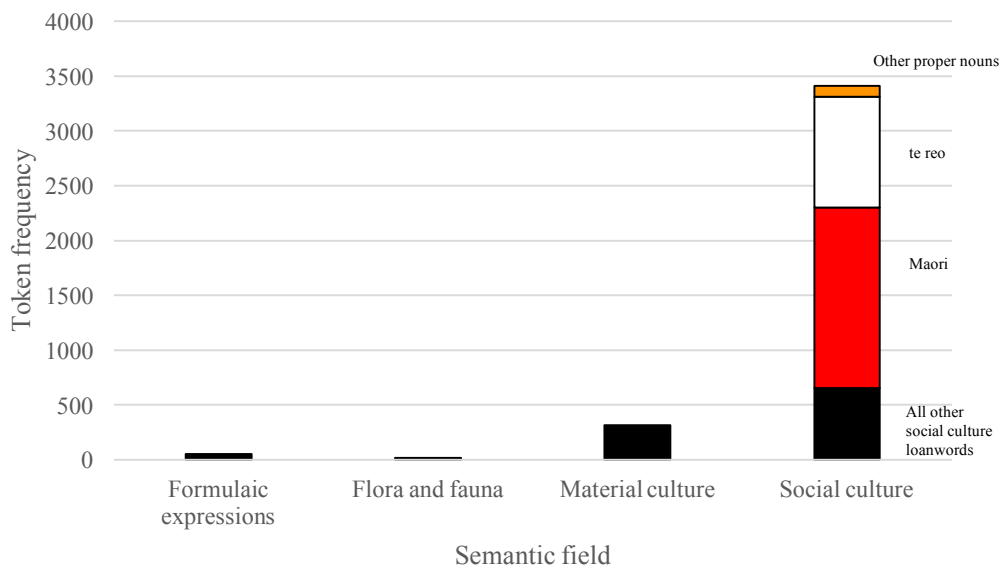


Figure 9: Percentage of loanword tokens by semantic field

Similarly, a direct quantitative comparison with Macalister (2006b) was not possible, as his semantic domain analysis excluded proper nouns, and exact figures were not given. However, some differences were noted when comparing the most like data set: non-proper noun tokens from newspapers in 2000. Although loanwords in the field of social culture were most common in both Macalister (2006b) and this study, they were not as dominant in the present data. Furthermore,

newspapers in 2000 had more flora and fauna loanwords than material culture loanwords, both by type and token (see Figures 7 and 8 in Macalister, 2006b, p. 16). Interestingly, the present data seems more comparable with data on non-proper noun tokens by semantic field from *Hansard* than from newspapers (see Figure 5 in Macalister, 2006b, p. 15).

This comparison reveals two factors at play. The first is that social culture loanwords have been increasing since the 1970s (see Macalister, 2006b), and this has continued into the present data, which spans 2008-2017. Secondly, it is likely that this increase has been boosted by the effect of the topical constraint (the social culture concept of language) under which this corpus was built. This same topical effect may account for the reverse in rank of material culture and flora and fauna loanwords, as there were few articles in the corpus that dealt with this topic.

5.1.4. Pluralisation by suffix

Only 24 instances of pluralisation in the data followed English morphological patterns, in which an ‘s’ suffix was added to a plural Māori loanword. These occurrences were spread across six types: *Maori*, *Kiwi*, *kiwi* (bird), *moko(puna)*, *kia ora* and *ka kite*. Altogether, ‘s’ pluralisation accounted for less than 0.01% of all tokens in the corpus. With only two exceptions, all of the types that were pluralised with an ‘s’ at least once had more than 15 tokens and thus sit in the top 30 most frequent types. These can be seen in Table 15 below.

It is interesting that most cases of loanword pluralisation (excluding *kiwi* ‘bird’, *ka kite* and *kia ora*) were words for people. After some investigation, it would seem that the former is a case of entrenchment, while the latter are unpredictable irregularities.

Both *ka kite* and *kia ora* occurred in the same article, where the increased use of the Māori language during Māori Language Week is discussed:

(5) “It’s one of my favourite awareness ‘weeks’ of the year as it has a tangible effect – the **kia oras** and **ka kites** can be heard from shops, workplaces, schools and communities around the country” (TW_2017_12_SEPT_TDP_001).

Although this is the only instance of *ka kite* occurring in the corpus, *kia ora* was one of the highest frequency loanwords (33 tokens). Nonetheless, it was only given

an ‘s’ suffix once. Contextually speaking, both of these examples can be considered anomalies, as such formulaic expressions would not normally be pluralised, either in Māori or English.

Table 15: English suffix pluralisation in the tokens

Type	Semantic category	Tokens pluralised with an ‘s’
Maori	Culture	0/58
	Language	0/1,114
	Names	0/66
	People group	1/274
	Other	0/137
	All variants	1/1,649
Kiwi	Other	1/3
	People	19/34
	All variants	20/37
Kiwi	Bird	1/1
moko(puna)	Grandchild	0/3
	Grandchildren	2/12
	All variants	2/15
kia ora	kia ora	1/33
ka kite	ka kite	1/1
Totals		26

The loanword *kiwi* was split into two types in the corpus. The first refers to the endangered national bird of New Zealand. The second, and most common, is *Kiwi* as a proper noun commonly used by New Zealanders to describe themselves; a semantic extension from the original meaning which has been in use since at least World War II (Macalister, 2001, p. 21). These types will be considered together as they are both long since well-established in NZE (Macalister, 2001) and therefore demonstrate the morphological entrenchment of Māori loanwords when they are no longer perceived as being foreign in NZE.

The only time the type *kiwi* (bird) occurs in the data, it is given an ‘s’ suffix following English grammatical rules, as in the following excerpt:

- (6) “Many things spring to mind when foreign people consider Aotearoa. Think of Lord of the Rings, our anti-nuclear stance, Buzzy Bee, pavlova, **kiwis**, the silver fern, our

landscape, All Blacks and, in light of recent events, the All Whites”
(MLW_2010_17_AUG_NZH_004).

Here, the kiwi bird is listed amongst several established icons of New Zealand culture, and this status is indicated by its grammatical entrenchment in NZE. It is likely that the loanword in this form is so entrenched in NZE that it is no longer considered a loanword.

The type *Kiwi* as a proper noun was further divided into two semantic sub-categories: one referring specifically to people, and one for all other tokens. Examples of the former can be seen in (7) and (8), and the latter in (9).

(7) “There are some key aspects that make up our identity and they must be recognised to give us a wider view of what drives us as **Kiwis** to be different”
(MLW_2010_17_AUG_NZH_004).

(8) “According to the website, Tauranga is the number one place in the country that **Kiwis** say is mispronounced on Google Maps” (TW_2017_12_SEPT_BPT_002).

(9) “We love the fact that the trends cover everything from the predictable to some delightfully quirky **Kiwi-isms**” (MLW_2010_01_MAY_NZH_001).

In cases where *Kiwi* referred to more than one New Zealander, *Kiwis* was always used. In total, over half of all *Kiwi* tokens were that variant. Moreover, (9) is an ironic case of a Māori loanword innovation, coined to describe lexical innovations in NZE. In contrast to all other Māori loanwords, *Kiwi* is being morphologically integrated into NZE. This is in line with findings that *Kiwi* is one of the most productive and flexible loanwords in hybrid compounds (Degani & Onsyko, 2010). It also reinforces Macalister (2001) who states that *Kiwi* (in both its semantic forms) “is today seldom thought of as a Maori [*sic*] word, and behaves as does any other English word, functioning both nominally and adjectivally, fitting into new coinings and adapting to new uses” (p. 22).

The next loanwords to be addressed are two high frequency types which did have the rare ‘s’ suffix pluralisation. These are the types *moko(puna)* and *Maori*. In the exceptional instances when these loanwords do conform to English grammatical rules, valuable insights are to be had.

The type *moko(puna)* encompassed both the singular and plural translations ‘grandchild’ and ‘grandchildren’, as it can express both meanings in Māori (Moorfield, 2018). It may also be shortened to *moko*, but remains distinguishable from the type *moko* (traditional Māori tattoo). *Moko(puna)* occurred 12 times with plural meaning, and three times with the singular in the corpus. However, only two of the pluralised tokens had an ‘s’ suffix. Both of these tokens occurred in the same article, as below:

(10) “‘It’s now our **mokopunas**’ turn, a lot of my nieces and nephews are all teaching and it’s feeding down to the **mokos**” (TW_2013_02_JUL_TDP_008).

These examples both come from the reported speech of an older female Māori interviewee. What is interesting is that the sole instance of *Maori* with an ‘s’ plural suffix is also spoken by a female Māori in her eighties. In this article, a Māori *kuia* (female elder) recounts her marriage:

(11) “‘The [Catholic] priest wouldn’t marry us, he reckoned **Maoris** were unreliable” (TW_2013_06_JUL_TDP_009).

Given that *Maori* has a token count of 1,649, this sole example is a significant finding. It is also in line with diachronic decreases in ‘s’ suffix pluralisation of Māori loanwords in the press noted as by Davies and Maclagan (2006).

In contrast to *Kiwi*, which is well-established in NZE, both *moko(puna)* and *Maori* are uniformly not grammatically assimilated into NZE. The only exceptions to this are the above cases of reported speech. This is in line with changes in style guide rules reported in Davies and Maclagan (2006), which have shifted from treating Māori words grammatically as any other NZE word, to dropping the ‘s’ suffix and enforcing non-grammatical conformity except in instances of direct quotes (*The Press* style books 2004, as cited in Davies & Maclagan, 2006, p. 88).

Pluralisation of Māori loanwords has been shown to be an indicator of attitude towards the Māori language in the literature (see Davies & Maclagan 2006; Harlow 2005; Onysko & Calude 2014). This stems from the view that to add “English plural suffixes to nouns of Māori origin...is to show disrespect to the Māori language” (Harlow, 2005, p. 141). While Harlow (2005) believes this to be

hypocritical, and a covert expression of opposition to the Māori language, this is a view reflected in stylistic changes in New Zealand newspapers. Davies and Maclagan (2006, p. 89) argue that “not imposing the grammatical rules of the dominant language in our culture reveals a growing acceptance of te reo Māori, and reflects the social and political status of Māori in general”.

However, this is not what I conclude. The inclusion of English pluralised suffixes in Māori loanwords in speech not only indicates a double standard in the press, it also shows that the concept of grammatical non-conformity as an expression of respect for a language is a double standard. Both the Māori speakers in the above articles express great respect and affection for the Māori language, and they have no problem with altering the grammar of loanwords to fit the receptor language. It is here that a question must be posed: who exactly dictates what respect for the Māori language looks like?

While this section does not consider loanwords that never conform to NZE grammatical rules, or alter orthographically, these few examples of conformity and non-conformity offer an insight into what appears to be a growing divergence between perceptions of how Māori loanwords should be correctly used, and their actual use. This theme of perception and use will be returned to in more depth in Section 6.

5.2. Extra-linguistic patterns

5.2.1. Loanword use in newspapers

In order to conduct an analysis of loanwords according to their newspaper of publication, the newspapers were divided into six categories to better represent the spread of newspaper articles themselves. The five newspapers with the highest number of articles were considered separately (*BPT*, *HBT*, *NZH*³⁴, *TDP* and *TNA*), while all other newspaper articles were grouped together due to their low independent frequency in the corpus. A summary of newspaper article, type and token data (Table 16) shows that the newspaper *TDP* (*The Daily Post*) was the highest ranked newspaper in all of these areas by a significant margin; in fact, *TDP* accounts for almost half of the data in the corpus alone. Interestingly, although *HBT*

³⁴ For the purposes of this analysis, the *HOS* (*Herald on Sunday*) was considered part of the *NZH* newspaper.

(*Hawke's Bay Today*) had the lowest number of articles, and second lowest total token count, it also had the second highest frequency of both loanword tokens and types. The *NZH* (*New Zealand Herald*) newspaper ranked second for total number of tokens, having approximately 3,000 more tokens than *TNA* (*The Northern Advocate*), despite them having an equal number of articles (40), indicating that article length is higher in the *NZH*. Moreover, the *NZH* actually had less loanword tokens and types than *TNA*. This is likely due to the nationwide audience of *NZH*, as opposed to the smaller regional audience of *TNA*, and the higher Māori population in Northland. The newspaper category 'other', comprised of 16 newspapers (*WCH*: seven articles; *WRT*: three; *KN*, *NPC*, *TWK*: two; 11 remaining newspapers: one article each)³⁵, was the least prominent in the corpus, with lowest or tied-lowest figures in all areas of Table 16.

Table 16: Summary of newspaper data by article, type and token

Newspaper	Articles	LW tokens	Total words	LW types	Total types
<i>BPT</i>	39	437	14,698	39	7,584
<i>HBT</i>	27	561	11,838	75	5,890
<i>NZH</i>	40 (1)	459	16,552 (591)	45	8,743
<i>TDP</i>	117 (1)	1,599	42,244 (450)	111	22,159
<i>TNA</i>	40 (2)	478	13,524 (69)	74	7,276
Other	27 (1)	261	10,069 (202)	39	5,317
	290	3,795	108,925		

Note: figures in brackets denote data regarding the five articles included in the corpus with a zero-loanword count.

Finally, it should be stated that five articles in the corpus did not include any loanword tokens at all. These zero-loanword count articles were still included in the corpus (see Section 3.6) and thus are included in the data considered here. Their distribution according to newspaper article and total token count is given in brackets in Table 16 for interest only; they were considered noteworthy as, although no loanwords occurred in them, at least one search term did occur in each.

³⁵ See Section 4.3 for a full list of newspaper abbreviations.

5.2.1.1. Most frequent loanwords in newspapers

A comparison of the five most frequent loanword types, with their frequencies according to newspaper, is given in Table 17. Average frequencies per 1,000 words for those same loanwords are given in Table 18. The top two most frequent loanwords across all newspapers are unsurprisingly *Maori* and *te reo*, given their generally high frequency in the corpus. However, beyond this, most frequent loanword types diverged according to newspaper. *Iwi*, the third most frequent loanword in the corpus, is third or fourth most frequent in all papers but the *TNA* and ‘other’ newspapers; while *reo*, the fourth most frequent loanword in the corpus, only occurs in the top five of two newspapers: *HBT* (ranked third) and *TDP* (ranked fifth). Moreover, 48 of the 59 total occurrences of *reo* are split between these two newspapers, indicating the possibility that *reo* is a regional-specific loanword variant of *te reo*. *Whanau* (fifth overall in corpus frequency) was ranked third in both *TDP* and *TNA*. The remaining 10 types spread out across the top five most frequent loanwords in newspapers all had frequencies of 11 or more, putting them in the top ranks of loanwords overall.

The newspaper *TDP* is the only publication to share the same five most frequent loanwords with those of the whole corpus. There was only a slight difference in ordering (*whanau* preceded *iwi* and *reo* in *TDP* rankings, as opposed to being fifth in the whole corpus), making it the newspaper most representative of the whole corpus in this respect.

While it is unsurprising that the majority of the loanwords in Table 17 are among the most frequent loanwords in the corpus, the variation between newspapers is noteworthy. When the five most frequent loanwords in the whole corpus are removed, thematic links between the remaining loanwords are quite different. In the *BPT*, the loanwords *Kiwi* and *kia ora* could be considered generic markers of New Zealand identity. The *HBT* differs in giving prominence to Māori cultural activities (*waka ama*³⁶, *kapa haka*). The *NZH* has a marker of New Zealand identity (*Kiwi*), but Māori language education loanwords such as *kura kaupapa* (*Maori*) and *kohanga reo* are also highly ranked. In *TNA*, there is an interesting juxtaposition of *marae*, a locus of the Māori community, and *Pakeha*, a marker of non-Māori identity. This is a general commentary only, and a more detailed analysis

³⁶ A traditional Māori sport involving outrigger canoes (*waka*).

Table 17: Token frequencies of the five most frequent loanwords, by newspaper

Rank	BPT		HBT		NZH		TDP		TNA		Other		Whole corpus	
	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.
1	<i>Maori</i>	217	<i>Maori</i>	212	<i>Maori</i>	218	<i>Maori</i>	710	<i>Maori</i>	196	<i>Maori</i>	96	<i>Maori</i>	1,649
2	<i>te reo</i>	138	<i>te reo</i>	139	<i>te reo</i>	111	<i>te reo</i>	385	<i>te reo</i>	108	<i>te reo</i>	68	<i>te reo</i>	949
3	<i>iwi</i>	17	<i>reo</i>	26	<i>iwi</i>	23	<i>whanau</i>	30	<i>whanau</i>	12	<i>kapa haka</i>	11	<i>iwi</i>	82
4	<i>kia ora</i>	12	<i>iwi</i>	12	<i>kura kaupapa (Maori)</i>	9	<i>iwi</i>	24	<i>marae</i>	11	<i>kai</i>	9	<i>reo</i>	59
5	<i>Kiwi</i>	5	<i>kapa haka = waka ama</i>	10	<i>kohanga reo = Kiwi</i>	8	<i>reo</i>	22	<i>Pakeha</i>	10	<i>Matariki</i>	8	<i>whanau</i>	55

Table 18: Average loanword frequencies per 1,000 words, by newspaper

Rank	BPT		HBT		NZH		TDP		TNA		Other		Whole corpus	
	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.	Loanword	Freq.
1	<i>Maori</i>	14.76	<i>Maori</i>	17.91	<i>Maori</i>	13.17	<i>Maori</i>	16.81	<i>Maori</i>	14.49	<i>Maori</i>	9.53	<i>Maori</i>	15.14
2	<i>te reo</i>	9.39	<i>te reo</i>	11.74	<i>te reo</i>	6.71	<i>te reo</i>	9.11	<i>te reo</i>	7.99	<i>te reo</i>	6.75	<i>te reo</i>	8.71
3	<i>iwi</i>	1.16	<i>reo</i>	2.20	<i>iwi</i>	1.39	<i>whanau</i>	0.71	<i>whanau</i>	0.89	<i>kapa haka</i>	1.09	<i>iwi</i>	0.75
4	<i>kia ora</i>	0.82	<i>iwi</i>	1.01	<i>kura kaupapa (Maori)</i>	0.54	<i>iwi</i>	0.59	<i>marae</i>	0.81	<i>kai</i>	0.89	<i>reo</i>	0.54
5	<i>Kiwi</i>	0.34	<i>kapa haka = waka ama</i>	0.84	<i>kohanga reo = Kiwi</i>	0.48	<i>reo</i>	0.52	<i>Pakeha</i>	0.74	<i>Matariki</i>	0.79	<i>whanau</i>	0.50

on individual loanwords per newspaper (not just the top five) is beyond the scope of this study. However, it does suggest that use of Māori loanwords in different newspapers may be oriented differently in different communities or readerships, even within a topic-bound corpus such as this one. This possibility will now be further explored by investigating average loanword frequencies per 1,000 words in newspapers as a whole, as well as potential effects of newspaper readership demographics on loanword use.

5.2.1.2. Average loanword frequencies in newspapers

The next analysis to be made is that of loanword token and type frequencies per article according to their newspaper of publication, as well as overall average frequencies within newspapers per 1,000 words. This was done so as to better compare differences in loanword usage between newspapers across the corpus.

As Table 19 demonstrates, there is a clear difference between newspapers regarding average token and type counts per article. The corpus average number of loanwords tokens per article is 13.08, which is unsurprisingly best represented in *TDP*'s loanword token count. However, it is the *HBT* which has the highest average number of loanwords per article, at almost 21 tokens. On the other end of the spectrum, the 'other' newspapers had the lowest average loanword token count, at 9.67 per article.

Regarding average numbers of types per article, the range was spread from less than 1 in *TDP* to 2.78 in the *HBT*; with the corpus average frequency of loanword types being 1.56 per article.

Table 19: Loanword token and type article averages by newspaper

Newspaper	Loanword averages per article	
	Tokens	Types
<i>BPT</i>	11.21	1.00
<i>HBT</i>	20.78	2.78
<i>NZH</i>	11.48	1.13
<i>TDP</i>	13.67	0.95
<i>TNA</i>	11.95	1.85
Other	9.67	1.44
Entire corpus	13.08	1.56

The average number of loanwords per 1,000 words according to their newspaper of publication was also of particular interest. As stated in Section 5.1.1, the loanword frequency per 1,000 words in this study is much higher than that of previous studies, even when *Maori* tokens are removed from frequency counts. Figures 10 and 11 (following) show newspaper loanword frequencies per 1,000 words with and without proper nouns respectively.

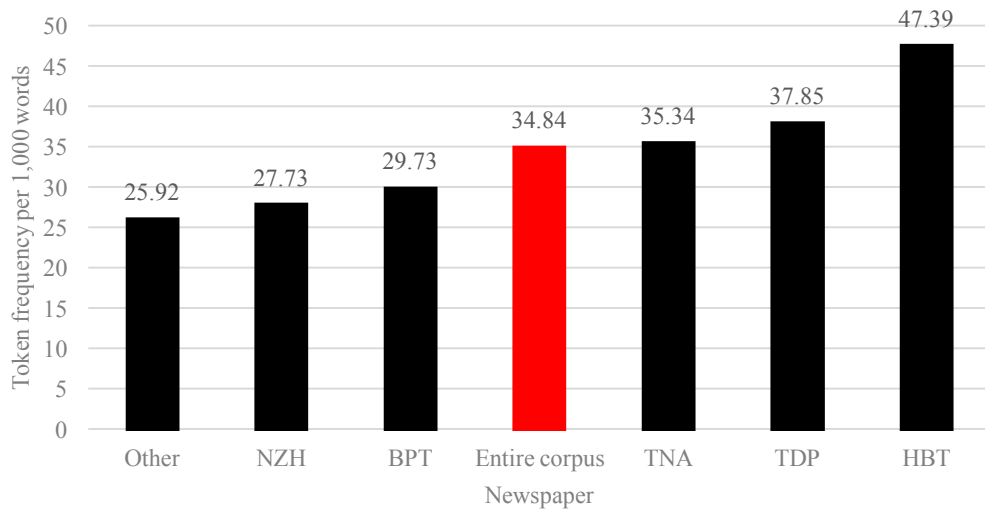


Figure 10: Loanword token frequencies per 1,000 words by newspaper (including proper nouns)

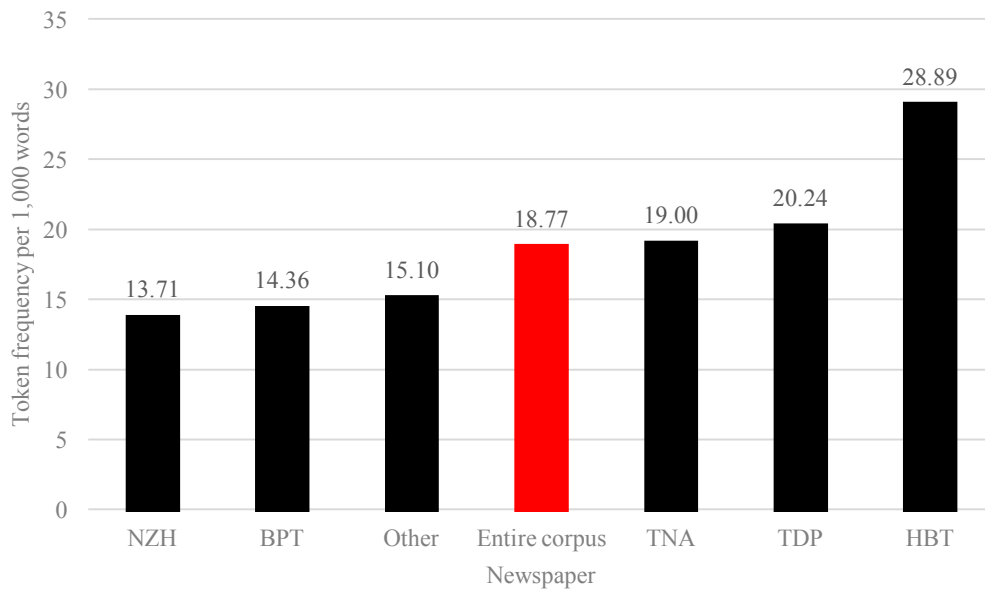


Figure 11: Loanword token frequencies per 1,000 words by newspaper (excluding proper nouns)

When compared to the corpus average frequency of loanwords per 1,000 words (including proper nouns, Figure 10), *TNA* is the most closely representative individual newspaper, with only half a word difference. *HBT* has the highest average of all the newspapers by a large margin – 10 more loanwords per 1,000 than the next highest, *TDP*. *BPT*, *NZH* and all other newspapers sit below the corpus average in the mid- to late-twenties; however even the lowest figure is four times greater than comparable figures from Kennedy (2001) and Macalister (2006b). These data also indicate that there is much inter-regional or inter-newspaper variation in frequency of loanword use.

When proper noun data are removed (*Maori*, *Kiwi*, *Pakeha*, *Matariki*), average frequency counts per 1,000 words are naturally lower in all newspapers and the corpus as a whole (see Figure 11); however, their ordering is almost identical in comparison to Figure 10. *HBT* still has the highest loanword count per 1,000 words out of all the newspapers, and the three newspapers above the whole corpus average (*HBT*, *TDP*, *TNA*) are still the same. The only notable difference made by excluding the proper noun data is that the ‘other’ newspapers now outstrip *BPT* and *NZH* for frequency. This indicates that the ‘other’ newspapers use a higher rate of proper nouns. But again, even when proper noun data are removed, all newspapers in the corpus have a much greater average loanword frequency per 1,000 words than in previous years.

5.2.1.3. Diachronic effect of loanword use in newspapers

Given that this corpus is a diachronic one, it was important that the frequency of loanword tokens in newspapers be analysed by year, in order to gain an accurate picture of how newspaper loanword frequencies are behaving between 2008-2017. To this end, yearly average loanword frequencies were calculated per 1,000 words for each newspaper. These data are in Table 20, with a visual representation in Figure 12 (data includes proper nouns).

When viewing the individual loanword frequencies per 1,000 words, only two newspapers (*HBT* and ‘other’) appear to have an easily interpretable diachronic trend; in this case, a decrease (Figure 12). It has already been established that the *HBT* has the highest average frequency of loanwords per 1,000 words, and this is reflected in its high averages in every year for which data were available. However, given the ‘other’ grouping of newspapers was an amalgamation of disparate

Table 20: Diachronic loanword frequencies per 1,000 words, by newspaper

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Overall
BPT	20.41	34.24	8.84	58.25	20.50	32.58	35.96	21.86	39.92	38.09	29.73
HBT	57.63				53.06	38.23	54.09	44.78	40.76	37.60	47.39
NZH	36.08	21.24	12.44	16.52	48.13	17.92	34.89	36.10	39.90	22.71	27.73
TDP	32.44	48.28	27.06	45.28	35.67	37.01	48.53	30.25	34.97	36.61	37.85
TNA	50.36	61.58	18.93	14.85	37.29	19.69	19.53	55.81	37.29	22.47	35.34
Other								46.01	32.52	17.64	25.92

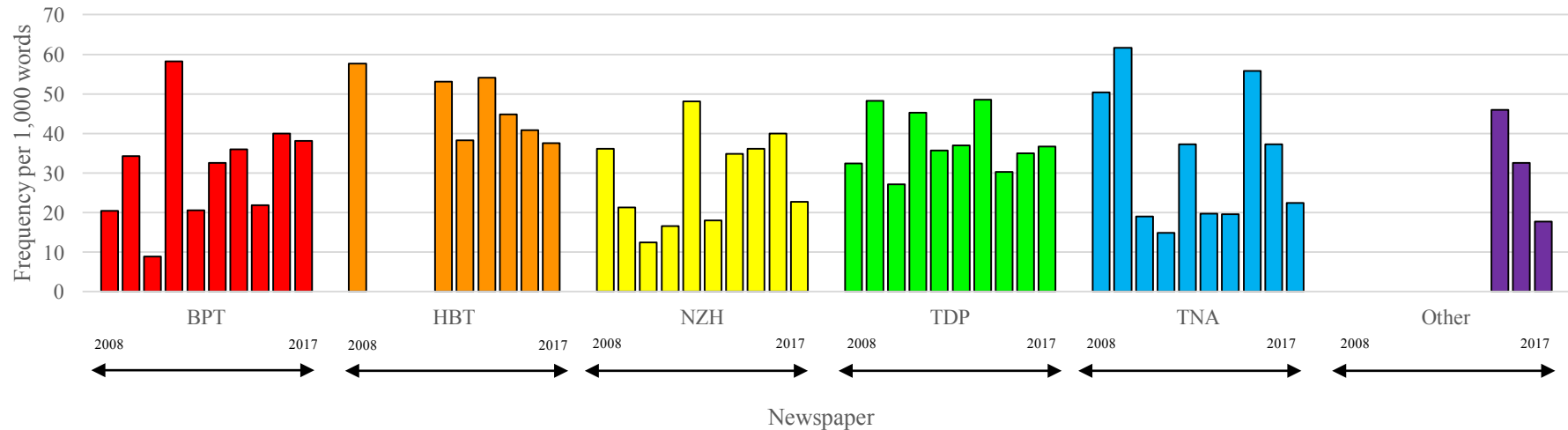


Figure 12: Diachronic loanword frequencies per 1,000 words per year, by newspaper

newspapers, little conclusion can be drawn from this. Of the remaining four newspapers, *TDP* was the most consistent in terms of diachronic frequency; however, even within this there was irregular yearly fluctuation. The *BPT* and *TNA* newspapers even less consistent, with yearly frequencies peaking and dipping unpredictably; although the *BPT* and *NZH* showed possible increases beginning in the middle of the decade.

It is important to clarify that there were two variables which had considerable influence on the diachronic interpretation of newspaper loanword data. Data were highly reliant on the number of articles per newspaper per year, as well as individual article word counts. Thus, a more standardised approach was necessary.

A Kendall Tau analysis was therefore conducted on the five individual newspapers ('other' newspapers were excluded), using the software R (R Core Team, 2018). It was found that all five newspapers had slight increases in loanword use over the 10-year period. However, none of these increases were found to be statistically significant. Similarly, Kendall Tau analysis of loanword types by newspaper found no significant diachronic change in any of the newspapers.

For this reason, it was decided not to conduct a diachronic loanword analysis of individual types as they feature across newspapers. Such an analysis will be done ignoring newspaper effect in Section 5.2.3. However, before this may be done, there is one more factor to consider in the variation of loanword use in newspapers: ethnicity.

5.2.1.4. Effect of ethnicity on loanword use in newspapers

It was not possible to measure the correlation between articles and author ethnicity as this data were not available in the corpus; however, ethnicity demographics for readerships by region were obtainable. A comparison of Māori and non-Māori population percentages was compiled from data from the 2013 census (Statistics New Zealand, 2018). This year was chosen as it was situated within one year of the mid-point of the year range covered by the corpus.

Ethnic demographics are given below for the five major newspapers considered. Northland data corresponds to *TNA*, Hawke's Bay to the *HBT*, and nationwide data to the *NZH*. Bay of Plenty data were split into Tauranga and Rotorua, correlating to the *BPT* and *TDP* respectively. This was because, although

both of these cities are located in the same region, they have very different ethnic breakdowns when it comes to Māori population, as Figure 13 makes clear.

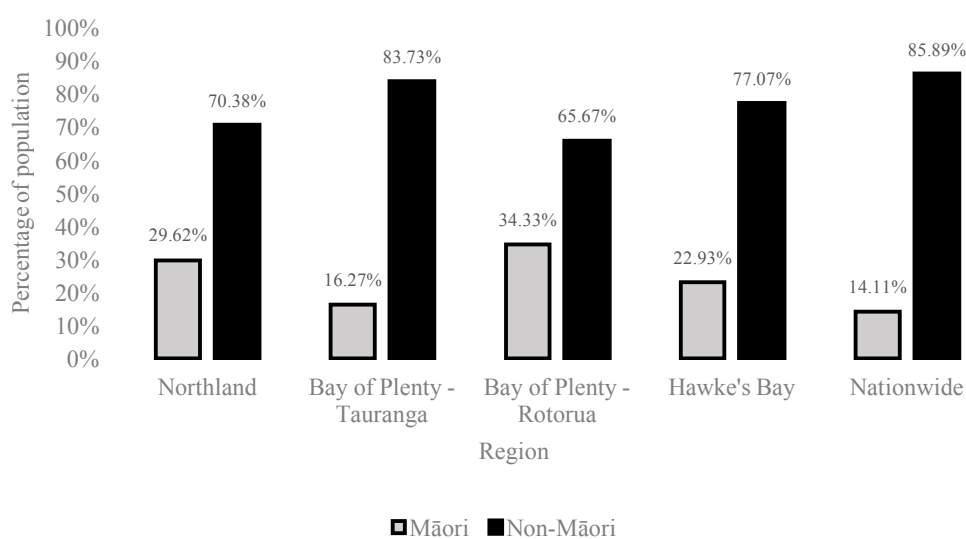


Figure 13: Percentage of Māori and non-Māori populations by region

Rotorua, represented by *TDP* newspaper, has the highest proportional Māori population out of all the regions surveyed above. Northland (*TNA*) is the next highest, followed by the Hawke's Bay (*HBT*); with Tauranga (*BPT*) the lowest of the regional newspapers. The correlation between newspaper average loanword frequencies and percentage of Māori population per newspaper region can be seen in Figure 14, where proper nouns are excluded.

In most cases, average token frequency per 1,000 words appears to correlate with the percentage of population identifying as Māori in each region. Lower Māori populations correlate with lower loanword frequencies, and higher Māori populations correlate with higher loanword frequencies. The exception to this is the *HBT*, from the Hawke's Bay region, which has the highest loanword frequencies, but only the third highest percentage of Māori-identifying residents, rather than the highest as might be expected. Similarly, Rotorua has the highest population of residents identifying as Māori, while the second highest loanword frequency is in its newspaper, *TDP*. It is likely that this is linked to the fact that Rotorua became New Zealand's first bilingual city in 2017 ("An 'inspirational' day", 2017).

Predictably, the newspaper with the lowest overall Māori population is the *NZH*, which has a national readership.

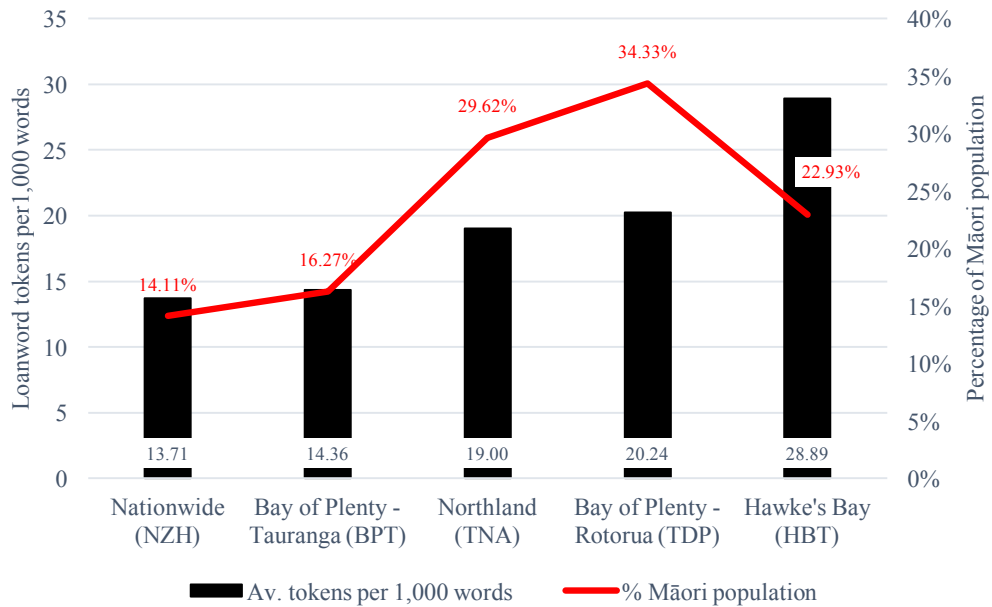


Figure 14: Comparison of regional Māori population data and newspaper loanword frequencies per 1,000

It can thus be stated that, generally, regional loanword use in newspapers will reflect Māori demographics in that same region; however, the Hawke’s Bay does not follow this trend. Further analysis will investigate whether these trends in loanword use and ethnic demographics are reflected in other aspects of the present data.

5.2.2. Diachronic use of loanwords in the whole corpus

A thorough analysis of diachronic loanword frequency data was next conducted from a whole corpus perspective, putting aside the factor of newspaper. The average number of loanword tokens per 1,000 words for each year was calculated from the summary of data (see Table 7, Section 4.2.) and compared with the whole corpus average.

What is immediately obvious from the below figure, is that the average frequencies of loanwords per 1,000 words is higher in almost every year spanned by the corpus data than the whole corpus average. However, even the three years which fall below the whole corpus average (2010, 2013 and 2017) are still much greater than comparable results from earlier studies (see Kennedy, 2001; Macalister, 2006b). In particular, the average frequency for 2010, predictably the

lowest in the corpus, is still 2.5 times greater than what it was in 2000³⁷ (Macalister, 2006b, p. 12). There was also a remarkably large range between the highest and lowest yearly average frequencies (21.87 tokens) from 2010 and 2012 respectively.

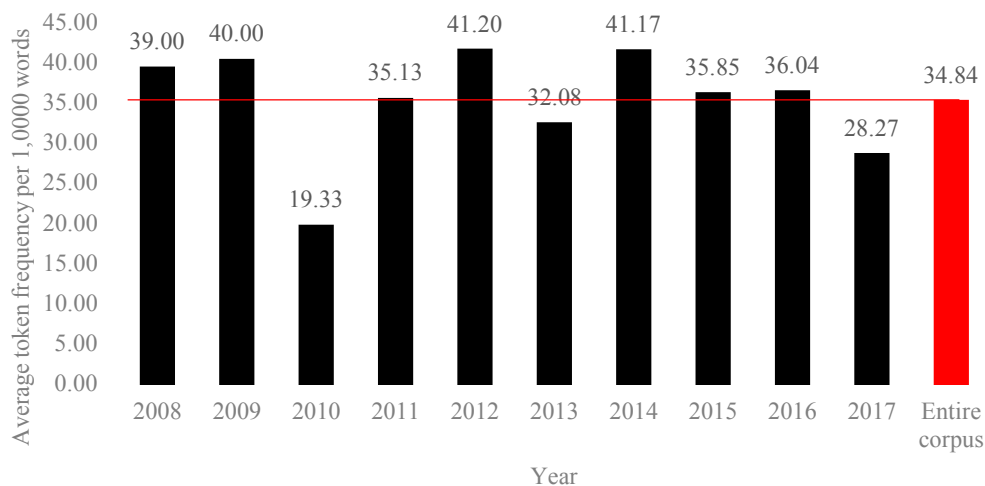


Figure 15: Diachronic loanword token frequencies per 1,000 words

It is difficult to state definitely if there is an increase or decrease in diachronic loanword use per 1,000 words from Figure 15 alone. Thus, a Kendall Tau statistical analysis was run on the loanword token frequencies and total token frequencies per year in order to ascertain whether there was any diachronic significance to this data.

It was found that there was a strong and statistically significant trend of increasing Māori loanword usage between 2008-2017 in the data ($\tau=0.6$, $p=0.020$). This is in line with previous findings that Māori loanword use is on the increase, both in general (Macalister, 2006b) and in newspapers (Davies & Maclagan, 2006; Macalister, 2006b).

5.2.3. Selection of individual loanwords for diachronic analysis

In order to best assess which individual loanwords to focus upon for a more in-depth diachronic analysis, two criteria were applied. The first was to investigate frequencies of loanword types as they occurred across the 10-year period of 2008-2017. Table 21 details the number of types which occurred in all 10 years of the corpus, or in just nine years, eight years and so on, regardless of specific year.

³⁷ When compared to the average frequency of 7.7 Māori loanwords per 1,000 words in newspapers in the year 2000 (Macalister, 2006b, p. 12).

Table 21: Type frequency distribution across 2008-2017

No. of years of occurrence	No. of types	Example of types
10	5	<i>Maori, te reo, marae, kapa haka, Pakeha</i>
9	3	<i>iwi, whanau, haka</i>
8	6	<i>kia ora, kura kaupapa (Maori), taonga, kai, tamariki, hapu (sub-tribe)</i>
7	7	<i>hangi, Kiwi, kohanga reo, moko(puna), powhiri, tikanga, whakapapa</i>
6	7	<i>kaumatua, korero, kuia, kura, reo, tangi(hanga), waiata</i>
5	7	<i>hui, kaupapa, kupu, mahi, mana, Matariki, wananga (university)</i>
4	11	Remaining loanwords were too numerous to include in the present table
3	17	
2	23	
1	100	
Total types	186	

The vast majority of loanword types (100) only occurred in a single year of the corpus, while 84.9% (158 types) occurred in five years or less across the corpus. Given that it is theoretically possible (although unlikely) for any loanword type with a frequency of 10 or greater to occur in every single year, there was the possibility that 36 types might have an even distribution across the corpus, i.e., occur at least once every year between 2008-2017. Predictably, this was not the case.

There were only five loanword types which occurred in every year of the corpus. These were: *Maori, te reo, marae, kapa haka* and *Pakeha*. Three types occurred in every year except 2010: *hapu, iwi* and *whanau*. It is likely that this singular year miss is due to the low number of data available from 2010. Finally, there were six types which occurred in all but two years: *hapu* (sub-tribe), *kai* (food), *kia ora, kura kaupapa (Maori), tamariki* (children) and *taonga* (treasure).

To these results, the second criterion for selecting individual loanwords for diachronic analysis was applied. This was to look at overall loanword frequencies in comparison to diachronic spread. All types which occurred in eight or more years of the corpus, and their diachronic frequency spread, can be seen in Table 22. It is interesting to note that that *iwi*, the type with the third highest token frequency, did

Table 22: Diachronic distribution of most frequent loanword tokens

Loanword	Corp. freq.	Rank	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
<i>Maori</i>	1,649	1	119	134	92	100	165	149	265	191	251	183
<i>te reo</i>	949	2	45	72	34	39	102	79	166	107	173	132
<i>marae</i>	51	6	2	2	2	9	5	7	5	8	10	1
<i>kapa haka</i>	47	7	1	1	1	1	9	9	1	9	14	1
<i>Pakeha</i>	39	8	7	6	2	1	3	4	3	2	3	8
<i>iwi</i>	82	3	13	11		3	6	10	16	8	8	7
<i>whanau</i>	55	5	4	3		3	8	7	8	7	12	3
<i>haka</i>	23	15	3	3		2	7	1	1	3	1	2
<i>kia ora</i>	33	10	1	1		1	2	5	10	5		8
<i>kura kaupapa (Maori)</i>	28	13	5	1			4	2	8	1	6	1
<i>taonga</i>	23	16	7		2	1	5	1	4		2	1
<i>kai</i>	17	22	2		2	2		1	1	2	5	2
<i>tamariki</i>	17	24		3		1	1	1	4	4	2	1
<i>hapu (sub-tribe)</i>	16	25	2			1	1	3	2	4	2	1

Note: Blacked out cells indicate the non-occurrence of a loanword type in that year.

not occur in every year, as might be expected. Nor did *reo* (fourth) or *whanau* (fifth). It is suggested that this would not be the case were there more articles in the corpus from 2010.

After some consideration, a decision was made to conduct an in-depth diachronic analysis on just three loanword types, rather than a more superficial analysis of many loanword types. The types selected were those with the highest corpus frequencies that also occurred in every year of the corpus: *Maori*, *te reo* and *marae*. As *Maori* and *te reo* can both semantically refer to the Māori language, and this corpus is topic-bound to Māori Language Week, they will be considered together in a case study on ‘*te reo Maori*’ in the following section. Section 5.4 will then be concerned with a comparative diachronic analysis of *marae* in the present data to that of previous studies.

While it is regrettable that an analysis of every individual loanword that occurred every year between 2008-2017 is beyond the scope of this study, let alone the loanwords that almost did, it is hoped that this data will provide a foundation for future diachronic analysis of individual Māori loanwords.

5.3. Case study: *te reo Māori*

It has already been established that *Maori* is the single most frequently occurring Māori loanword in the corpus, and that this is followed by *te reo*. These results are not unusual given that the corpus is based on the topic of Māori Language Week, and that phrase, as well as ‘*te wiki o te reo maori*’ were both used in its construction. A significant part of the present research, therefore, is the use of Māori loanwords which pertain specifically to the Māori language itself.

Previous studies have often focussed on the loanword *Maori* because of its prominence in any collection of Māori loanwords. It is the most frequent loanword in almost all of the previous research that considers the issue (Calude et al., Forthcoming, p. 11; de Bres, 2006, p. 23; Kennedy, 2001, p. 71; Kennedy & Yamazaki, 2000, p. 41; Macalister, 2006b, p. 15; Onysko & Calude, 2014, p. 151, 159). *Te reo*, too, is an extremely high frequency loan in previous studies (see *reo* in Kennedy, 2001, p. 71); as well as being noted as a moderately frequent collocate with *Maori* (Calude et al., Forthcoming).

Because the loanword *Maori* has several different semantic connotations, these will first be distinguished, before an analysis of *Maori*, *te reo* and other loanwords as they pertain semantically to the Māori language will be made.

5.3.1. Frequencies of language-referent loanwords

Three loanword types were recorded in the corpus that referred semantically to language, either the Māori language or language in general. These were: *Maori* (1,649 tokens), *te reo* (949 tokens) and *reo* (59 tokens). Together, they accounted for 2,657 tokens, or 70.01%, of the corpus.

However, not all occurrences of *Maori* referred to the Māori language. This loanword was also prevalent in its reference to the Māori as an ethnic group of people. An analysis of the different semantic meanings assigned to this loanword was therefore made (see Figure 16).

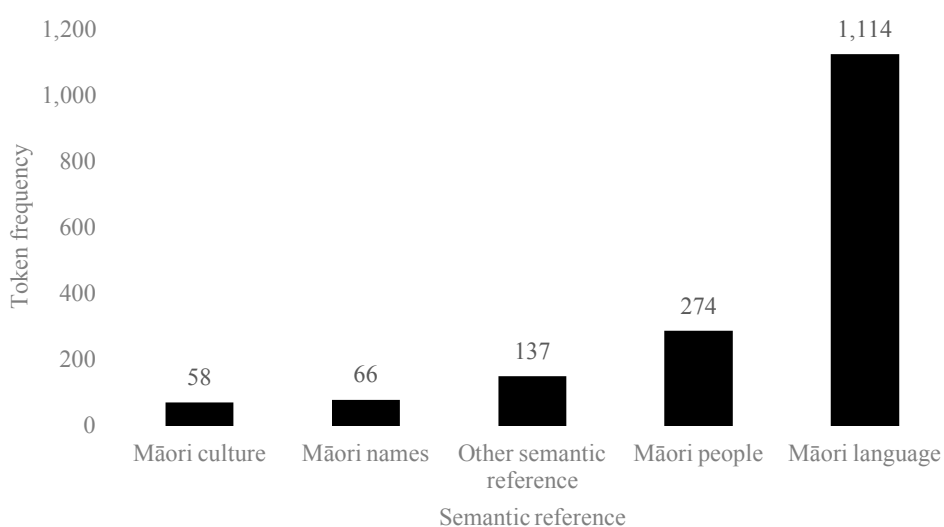


Figure 16: *Maori* token frequencies by semantic reference (excluding ‘te reo Maori’)

A total of 1,114 tokens of *Maori* were used to refer semantically to the Māori language, a figure which dwarfs that of the next most frequent; *Maori* as a semantic referent to people of Māori ethnicity (274 tokens). While it was expected that Māori language-referent tokens would have a high frequency, Māori ethnicity-referent tokens were surprisingly low.

The remaining *Maori* tokens were categorised as referring to Māori culture, Māori names, or other. It was noted that the majority of *Maori*-name referent tokens

names in particular occurred during 2013, when the theme for Māori Language Week was ‘*ngā ingoa Maori*’ (Māori names; refer to Section 4.5), or in 2014 referring to the top Māori names of the previous year. Additionally, most Māori culture or name-referent tokens functioned as modifiers, specifically adjectival modifiers, for example:

(12) “[She] said she had a great introduction to **Maori culture** when she moved to New Zealand from Fiji some 26 years ago” (MLW_2013_05_JUL_TDP_002).

(13) “Others give their loved ones **Maori names** to keep their whanau name alive so 24 ancestors are remembered, while some use **Maori names** to remember an event or even a special place that had a **Maori name**” (MLW_2017_12_SEPT_TDP_001).

Although these modifiers could have been considered as one group, there were enough of them that they were split into a separate category for interest.

Any instance of *Maori* which did not fall into any of the aforementioned categories was considered ‘other’; as was any token which was not easily classifiable. For the purpose of the following analysis, it is the 1,114 *Maori* tokens referring to language that are of main interest.

5.3.2. Sub-types of Māori language-referent loanwords

Maori was not the only loanword type in the corpus which referred to the Māori language; *te reo* and *reo* (literally ‘the language’ and ‘language’ respectively) also served this function. The first thing that was noticed in attempting to semantically categorise these loanwords was that distinguishing between the possible meanings of *te reo* and *reo* was contextually very difficult. A complete semantic and grammatical analysis of these loanwords is beyond the scope of this study. As such, for the purposes of the present analysis, they were both assumed to denote the Māori language except in cases where they specifically referred to another language or dialect. This only occurred twice: the first, in ‘*te reo Pakeha*’ (the English language), and the second, in ‘*te reo Waikato*’ (the Waikato dialect).

It was also observed that the phrase ‘*te reo Maori*’ (the Māori language) had a strong presence in the corpus, even once all search terms were removed. While this phrase was considered as being two separate lexical items and therefore tokens for the measurement of loanword frequencies, they were sufficiently common

collocates to warrant further investigation. To this end, a new categorisation methodology for measuring Māori language-referent sub-types was created. All *te reo* and *Maori* tokens which co-occurred in the bundle ‘*te reo Maori*’ were recategorised as one, new lexical unit, and 414 tokens of these were found to occur. This left a total of 697 remaining Māori language-referent³⁸ *Maori* tokens, 533 Māori language-referent *te reo* tokens, and 56 Māori language-referent *reo* tokens. Additionally, *reo* without the *te* prefix occurred 3 times in the phrase ‘*reo Maori*’.

Table 23: Frequencies of Māori language-referent loanword sub-types

Language loanword sub-types	Language sub-type frequency	Total token frequency in whole corpus
<i>Maori</i> (tokens referring to the Māori language only)	697	697
‘ <i>te reo</i> ’	533	533
<i>reo</i> (no preceding <i>te</i>)	56	56
‘ <i>te reo Maori</i> ’	414	828
‘ <i>reo Maori</i> ’ (no preceding ‘ <i>te</i> ’)	3	6
Totals	1,703	2,120

Note: Figures in the right-most column are double that of sub-type frequencies in two cases (‘*te reo Maori*’ and ‘*reo Maori*’) because these phrases are comprised of two loanword types combined. Therefore, one phrase equates to two tokens according to the overall corpus frequency counts established in Section 5.1.1.

Table 23 shows the five sub-types which all refer exclusively to the Māori language. Between them, they have a distribution of 1,703 tokens. When counted according to overall corpus frequencies set out in earlier sections, they account for 2,120 tokens, or 55.9% of all loanword tokens in the corpus.

Next, percentages of token frequencies for each of these Māori language-referent sub-types were calculated from overall sub-type frequencies (see Figure 17). When occurring in its meaning of the Māori language, the sub-type *Maori* was unsurprisingly the most prevalent, accounting for almost half of all such sub-types. However, there was only a percentage difference of 9.63% (164 tokens) between *Maori* and *te reo* when they had the same semantic meaning. Although there appears to be a preference to use the loanword *Maori* to indicate the Māori language over other options, neither is *te reo* an uncommon choice. Moreover, ‘*te reo Maori*’, which has the same semantic meaning as either of its loanword components in

³⁸ Māori-language referent tokens here is defined as any token which semantically refers to the Māori language only, and not the culture, people etc.

isolation, is also reasonably frequent in the data, accounting for nearly one quarter of all tokens.

In comparison, *reo* is far less frequent, at 3.29%; while the phrase ‘*reo Maori*’ without the preceding *te* has only 3 tokens and a <1% frequency.

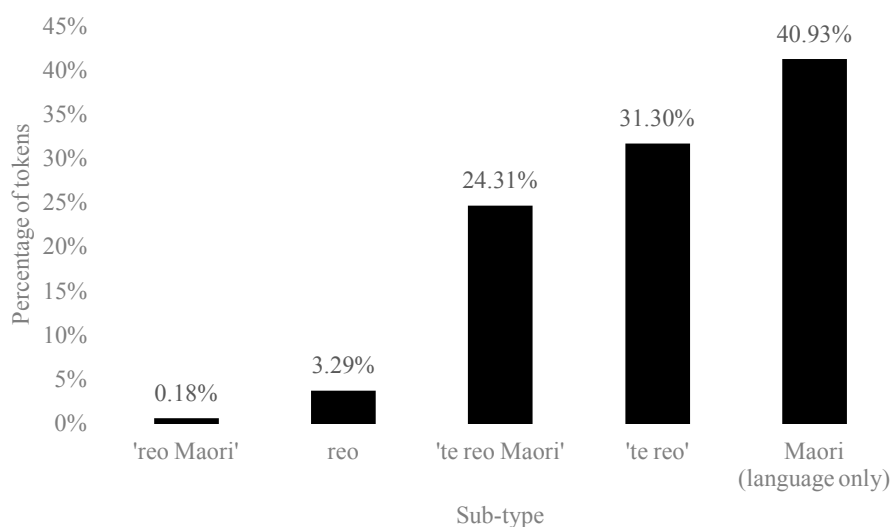


Figure 17: Percentage of tokens of Māori language-referent sub-types

These data are interesting because they represent the lexical choices available to a journalist, or interviewee, when referring to the Māori language. Five sub-type options were found to exist, used to greater or lesser degrees. The questions of exactly which newspapers are using these different sub-types; if there is a visible preference between newspapers; and if there is any diachronic effect on their use, will now be considered.

5.3.3. Sub-types of Māori language-referent loanwords in newspapers

This section will explore any differences in distribution of Māori language-referent sub-types between different newspapers.

As Figure 18 demonstrates, a similar preference trend can be seen in the use of Māori language-referent loanwords across all newspapers in the corpus. All newspapers most frequently employed the sub-type *Maori* to refer to the Māori language. This was followed by *te reo* and then ‘*te reo Maori*’ in all cases. However, there was some difference in the use of *reo*. While every newspaper did use *reo* at least once, it was only the *HBT* and *TDP* newspapers which had *reo*

frequencies of any note. *Reo* occurred in 25 tokens in the *HBT*, and 20 in *TDP*; the two newspapers with highest Māori populations in the data. In all other newspapers, *reo* only occurred in four tokens or less. Moreover, the sub-type ‘*reo Maori*’ only occurred in the *HBT* and *TDP*. This suggests that the sub-type *reo*, as a loanword variant of *te reo*, may be a variant specific to those regions (the Hawke’s Bay and Rotorua); or that it is an incoming variant from Māori into NZE being led in these two regions. Whatever the case may be, it is suggested that the phrase ‘*reo Maori*’ may become more frequent in the future.

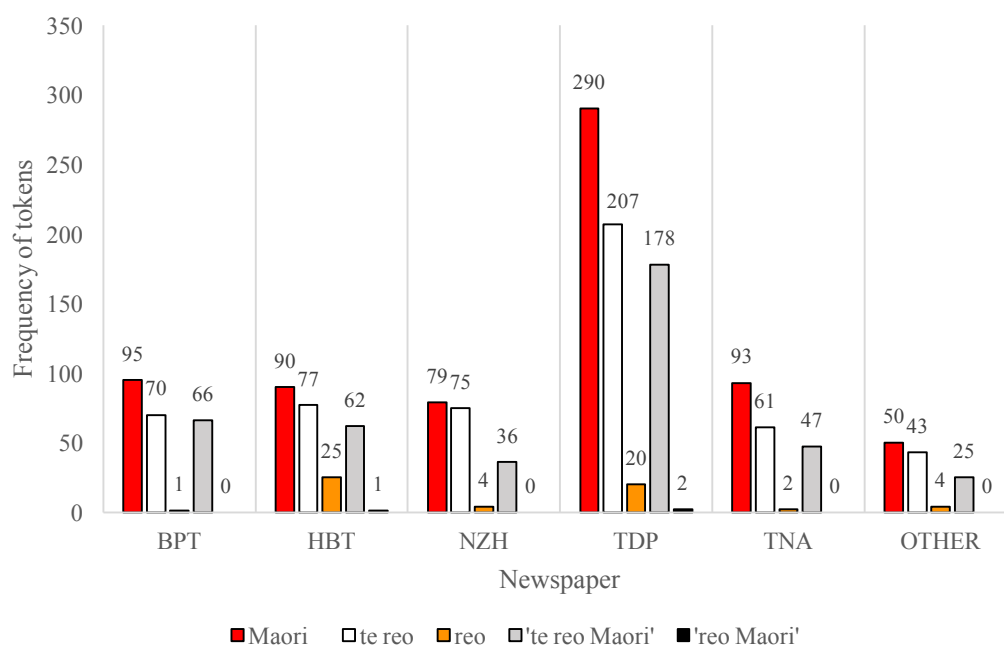


Figure 18: Frequency of Māori language-referent sub-types by newspaper

The frequencies of sub-type tokens that referred exclusively to the Māori language were also compared with all other non-Māori language-referent loanword tokens, according to their newspaper of origin (see Figure 19). Interestingly, there was only one newspaper in which Māori language-referent token numbers surpassed that of all other tokens. This was the *BPT*, which notably has the lowest demographic Māori population of all the newspapers. It was also the only newspaper of which more than half of all tokens referred to the Māori language, at 53.09%; all other newspapers had more non-Māori language referent loanword tokens than otherwise, although this was usually by a slim margin (see Table 24). The ‘other’ newspapers had the next highest proportion of Māori language-referent tokens, at just under 47%.

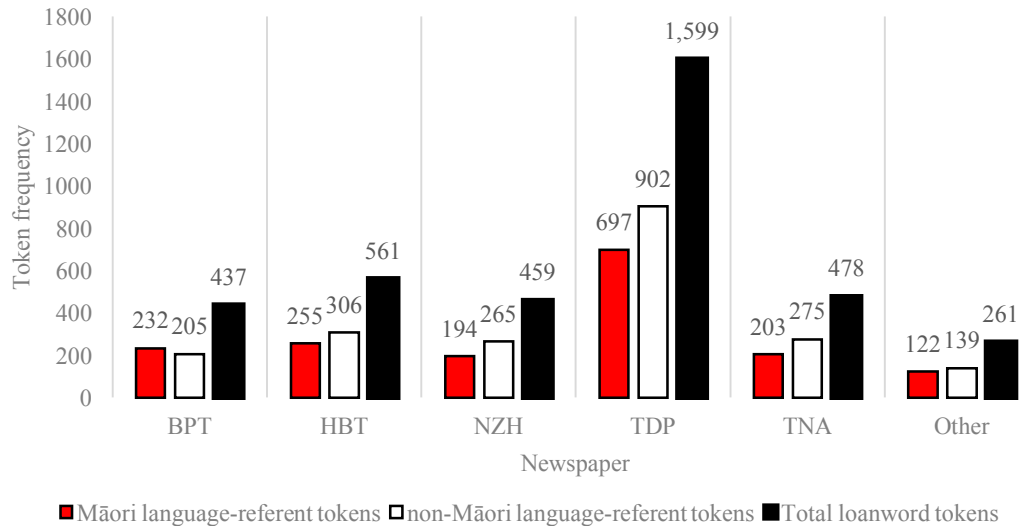


Figure 19: Comparison of Māori language-referent token frequencies to all other token frequencies by newspaper

Table 24: Percentage of Māori language-referent loanword tokens by newspaper

Newspaper	Percentage of Māori language-referent tokens
BPT	53.09%
HBT	45.45%
NZH	42.27%
TDP	43.59%
TNA	42.47%
Other	46.74%

An unexpected finding was that the *NZH* and *TNA* had almost identical rates of Māori language-referent tokens, with 42.27% and 42.47% respectively. Given findings mentioned earlier in this study on *NZH* loanword frequencies, it was a surprise that this newspaper had the lowest proportion of Māori language to non-Māori language-referent loanword tokens. Overall, there was a difference of 11% in variation between these percentages between different newspapers. This, combined with sub-type preference trend data, suggests that in general, newspapers are behaving similarly in their division of use between general Māori loanwords, and loanwords referring specifically to the Māori language, but that ethnicity of readership may have a slight influence.

5.3.4. Diachronic use of sub-types of Māori language-referent loanwords

Finally, diachronic effect on the Māori language-referent sub-type token frequencies was explored. This was made easier as all of these sub-types occurred in every year of the 10-year period of data, except for the least common sub-types:

reo, which did not occur for three years between 2009-2011, and ‘*reo Maori*’ which only occurred in 2014.

A clear diachronic increase in the three main sub-types *Maori*, *te reo* and ‘*te reo Maori*’ can be seen in Figure 20. *Maori* had the highest frequency of occurrences in all years, except, interestingly, in 2017, when it was surpassed by *te reo*. This could indicate that *te reo* is becoming the preferred loanword for referring to the Māori language, although more data from coming years is required before such a hypothesis can be corroborated.

In a similar vein, *te reo* was the second most frequent Māori language-referring subtype in all years but two: 2010 and 2014, when it was surpassed by the phrase ‘*te reo Maori*’.

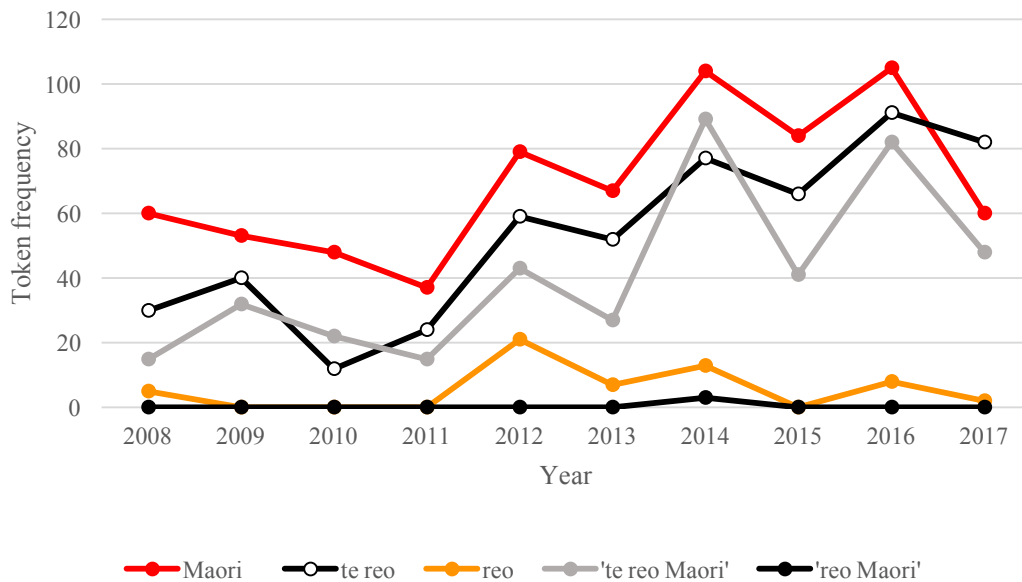


Figure 20: Diachronic token frequencies for Māori language-referent sub-types

Reo has less of a clear trend, with a frequency spike in 2012 and then what appears to be a decline in use; however, it has already been established that variation in use of *reo* appears to be regionally dependant. Similarly, ‘*reo Maori*’ does not have enough data to merit any generalisations, although its three tokens offer the possibility that it is a recent acquisition into NZE given that it does not occur prior to 2014.

To confirm whether or not any of the aforementioned and apparent diachronic increases in frequencies were significant, the Kendall Tau test was made on the following types: *Maori*, *te reo*, *reo*, ‘*te reo Maori*’, and ‘*reo Maori*’. This

Table 25: Diachronic frequencies for all Māori language-referent sub-types

Sub-type	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	
<i>Maori</i>	60	53	48	37	79	67	104	84	105	60	697
<i>te reo</i>	30	40	12	24	59	52	77	66	91	82	533
<i>reo</i>	5				21	7	13		8	2	56
<i>‘te reo Maori’</i>	15	32	22	15	43	27	89	41	82	48	414
<i>‘reo Maori’</i>							3				3
Totals	110	125	82	76	202	153	286	191	286	192	1,703

Table 26: Diachronic token frequencies of all Māori language and all non-Māori language-referent loanwords

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	
Māori language-referent tokens	110	125	82	76	202	153	286	191	286	192	1,703
non-Māori language-referent tokens	176	151	84	126	238	213	316	257	310	221	2,092
Total loanword tokens	286	276	166	202	440	366	602	448	596	413	3,795

was done using the individual sub-type frequencies per year as given in Table 25 (above), compared to overall sub-type frequencies per year. These results are in Table 27.

Table 27: Kendall Tau results for diachronic Māori language-referent loanword types

Māori language-referent type	Tau figure	2-sided <i>p</i> -value	Increase	Significant
<i>Maori</i>	0.405	0.12685		NO
<i>te reo</i>	0.689	0.00729	Large	YES
<i>‘te reo Maori’</i>	0.539	0.03888	Large	YES
<i>reo</i>	0.167	0.57801		NO
<i>‘reo Maori’</i>	0.149	0.72772		NO

No increase of significance was found in the types *reo* or *‘reo Maori’*. The type *Maori* had a slight increase in use across the 10-year period, but this was not found to be significant. In contrast, *te reo* increased significantly in frequency, as did *‘te reo Maori’*. This suggests that these types are a viable and frequent lexical replacement for the type *Maori*, and that they, especially *te reo*, can be expected to continue to increase in frequency in coming years.

A comparison was also made between the diachronic data of Māori language and non-Māori language-referent token frequencies. The data used for this analysis can be seen in Table 26 (above). Both Māori language and non-Māori language referent tokens were found to have slight increases, but neither of these were significant.

Given that this corpus was constrained to the topic of Māori Language Week, it is unsurprising that so many loanwords occurred in reference to the Māori language. It does, however, prove that the topic is a strong attractor of language-associated Māori loanwords. Significant diachronic increases in the sub-types *te reo* and *‘te reo Maori’* (but not *Maori*) as a reference of the Māori language suggest that loanwords used to denote the Māori language are undergoing a transition. Diachronic data on these two sub-types will be needed in coming years, in order to ascertain whether this increase is the result of topical attraction within the corpus, or if these sub-types becoming more common in NZE across the board.

There was also some indication that less frequent loanword types denoting the Māori language such as *reo* are increasing, and more likely to occur in regions

with higher populations of people who self-identify as Māori. However, more information is needed before these claims can be corroborated.

Finally, there was a large discrepancy between the frequencies of *Maori* tokens referring specifically to the Māori language (44.87% of all loanwords), and all *Maori* tokens when semantics are ignored (70.01%). This proves that *Maori* must be considered as polysemous when it is analysed, and coded accordingly, to avoid misrepresenting loanword frequencies. Future research on Māori loanwords must therefore take notice.

5.4. Case study: *Marae*

Before an analysis can be made of the loanword *marae*, a definition of the word and the conceptualisation behind it is in order. Although formerly in NZE *marae* referred to the open space or courtyard before the *wharenuī* (meeting house), a semantic shift took place in the 1960s (Mead, 2003, p. 101). Degani (2017) gives a good description of the contemporary, physical aspect of the *marae*:

In its current usage, the term *marae* refers to a fenced-in complex consisting of a few buildings. The main building is a carved meetinghouse painted in red (*wharenuī* or *whare tipuna*) and with a courtyard in front of it (*marae ātea*). A dining hall and cooking area (*wharekai*), as well as toilet and shower facilities (*whare paku*) and a shelter for visitors are also normally part of it. (p. 669)

For many New Zealanders, the above depiction is sufficient to understand what a *marae* is. However, Degani (2017) goes on to add that there is also a complex set of cultural schema and concepts associated with the word for many Māori³⁹. These diverging representations of *marae* will be addressed in due course.

5.4.1. *Marae* diachronic results

Marae is one of the most salient and commonly used Māori loanwords in NZE (Davies & Maclagan, 2006; Degani & Onysko, 2010; Kennedy, 2001). Davies and Maclagan (2006) found that it is no longer being glossed in newspapers, indicating its establishment as part of the NZE lexicon (p. 86). It is also among the most frequent loanwords when it comes to the construction of hybrid compounding

³⁹ Degani (2017) does not address cultural schema of *marae* in speakers self-identifying as Pākehā or non-Māori in her paper.

(Degani & Onysko, 2010, p. 216). This high frequency is attested in the present corpus.

Marae was the loanword with the third highest diachronic frequency of distribution in this corpus, and the sixth most frequent loanword overall. It accounted for 0.01% (51 out of 3,795) of all loanwords. This is a marked increase from 0.007% (41 *marae* tokens of 5,952 loanwords) in the *WWC* studied by Kennedy (2001)⁴⁰. This also equated to an average of 13.44 *marae* tokens per 1,000 words in this corpus; a much higher figure than the average of <0.03 *marae* tokens per 1,000 words (or 28.24 per one million words) in the comparable *NZH* data taken between 2003-2008 in Degani and Onysko (2010, p. 215). Moreover, in Davies and Maclagan (2006), *marae* was shown to increase in frequency in 2004 as a result of the topic-driven effect of the Foreshore and Seabed issue which was prominent that year.

In the present data, *marae* tokens occurred in 34 articles, and in every year of the 10-year period between 2008-2017. This diachronic distribution can be seen in Figure 21. Note that the frequency count of these tokens excludes any use of *marae* as a proper noun (of which there were 20); for example, ‘Ruamata Marae’ was considered a proper noun (from file MLW_2008_29_JUL_TDP_001).

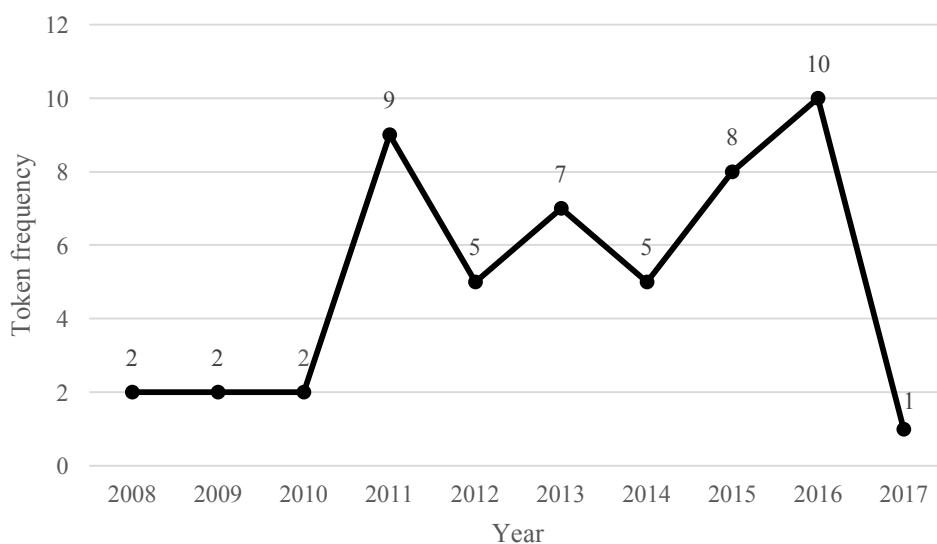


Figure 21: Diachronic frequency distribution of *marae* non-proper noun tokens

⁴⁰ See Kennedy (2001) for total number of loanwords in the *WWC* (p. 68), and for *marae* tokens in the *WWC* (p. 71).

Figure 21 shows a general trend for an increase in *marae* tokens diachronically. Within this, there are two spikes in the frequency of tokens, in 2011 and 2016. These spikes are of interest as their years are respectively the year with the lowest number of articles and total word counts in the corpus (2011), and the year with the highest number of articles and total word counts (2016). This suggests that 2016 is less a spike than actual growth; while 2011 seems even more anomalous.

Average frequencies of *marae* per 1,000 words per year in the corpus were next calculated in order to measure this spike more objectively. As Figure 22 demonstrates, average frequencies of *marae* still show a diachronic increase, while the frequency spike in 2011 is even more pronounced: at 1.57 *marae* tokens per 1,000 words, it is over double the next highest frequency, in 2015. There is also a notable decrease in use of *marae* in 2017. In fact, the average frequency per 1,000 words is lowest in the most recent year of this study, at 0.07 *marae* tokens per 1,000 words. Due to these unclear results, a Kendall Tau analysis was conducted, and no increase of significance was found. It is thus likely that *marae* only occurs in increased frequencies as an effect of topical attraction, rather than as a part of a continuous diachronic increase in use.

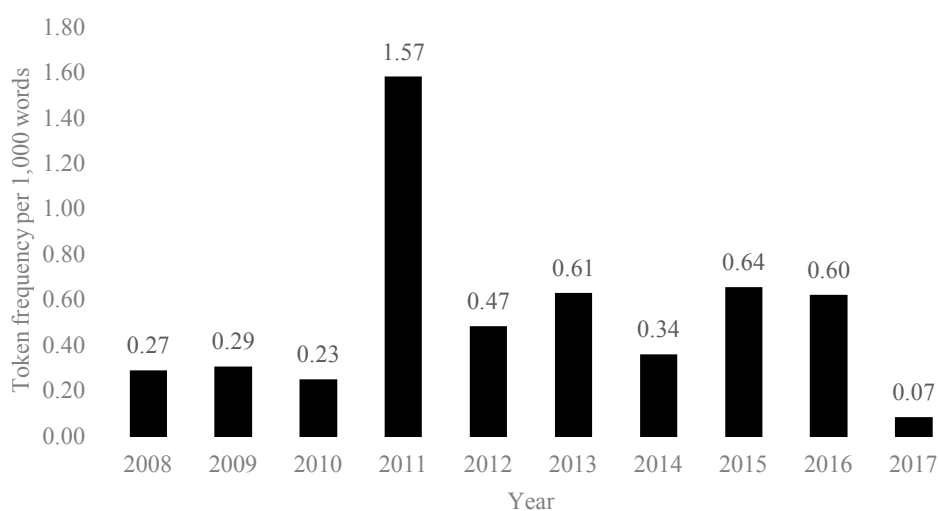


Figure 22: Diachronic frequencies of *marae* tokens per 1,000 words

5.4.2. *Marae* in newspapers

Marae tokens were more likely to occur in *TDP* than any other newspaper, a finding which is unsurprising because of the high proportion of *TDP* articles in the corpus.

Marae tokens did not occur in every year in any newspaper, but were close to it in the *TDP*, occurring in seven of the 10 years in the data. Surprisingly, the newspaper with the least number of *marae* tokens was not the ‘other’ newspapers, but the *BPT*. The *BPT* had only one token, and this occurred in the year with the highest *marae* count, 2011. *TNA* was the newspaper with the second highest count at 11 tokens, as well as the second widest year spread of *marae*, occurring in 4 of 10 years; while *HBT* was the third in both respects, with nine tokens occurring in three years. The complete figures for newspaper frequencies of *marae* by year are given in Table 28. Only six *marae* tokens (11.8%) occur in the corpus prior to 2011, indicating that it may be becoming more frequently used in the newspapers surveyed here. The data also suggest that *marae* was already well-established in the newspaper *TDP* prior to 2011, while it is becoming more established in other papers.

Table 28: Diachronic *marae* tokens by newspaper

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Totals
<i>BPT</i>				1							1
<i>HBT</i>					3	2			4		9
<i>NZH</i>				2		1		3			6
<i>TDP</i>	2	2		3	2	4	5		2		20
<i>TNA</i>			2	3				5		1	11
<i>Other</i>									4		4
Totals	2	2	2	9	5	7	5	8	10	1	51

5.4.3. Semantic analysis of *marae*

A semantic analysis was conducted on the type *marae* in order to investigate whether there was any variation in how it is being used. Following Degani’s (2010) observations that *marae* may indicate either a physical grounds or place, or a community of people, these two semantic distinctions were included in the analysis. Table 29 provides a summary of all *marae* tokens according to their semantic classification⁴¹. As the data show, the most frequent use of *marae* was in its traditional meaning referring to a physical place. There was also a high number of proper nouns, which are included here for interest. Six tokens occurred in which *marae* was used to refer to a community group associated to a *marae* rather than

⁴¹ Note that Table 29 includes proper nouns that were excluded from the total corpus frequency count.

the physical grounds or buildings, and three which could not be considered as being any of the above categories. These three tokens featured two uses of the phrase ‘global *marae*’ (both from the same article MLW_2011_05_JUL_NZH_002), in which the concept of a *marae* is expanded by semantic extension; and the other was as a modifier in the phrase ‘*marae* styled food’ (MLW_2011_05_JUL_BPT_001). Additionally, eight tokens were semantically unclear: in these instances, it was impossible to distinguish between the *marae* as referring to either a physical place or community; or between a physical place or other. These results were in line with previous findings that there has been an increase in the semantic expansion of *marae* in NZE from exclusively meaning a meeting or gathering place, to including a community of people (Davies & Maclagan, 2006, p. 96; Degani, 2010, p. 185); as well as Degani’s (2010) claim that there was a marked tendency to borrow *marae* as a common noun rather than a proper noun or an adjective.

Table 29: Semantic analysis of *marae* (terms adapted from Degani, 2010)

Semantic meaning	Tokens
Grounds/Physical place	34
Proper noun	20
Unclear	8
Community/Group of people	6
Other	3
	71

Next, semantic distinction of the type *marae* was assessed according to newspaper distribution. Once again, proper nouns were included in the data. Figure 23 details the results.

Marae was more likely to be used to refer to a physical place than anything else in all newspapers except the *BPT* (in which the sole occurring token was a modifier categorised as ‘other’), and the collective of ‘other’ newspapers, where there were slightly more proper noun tokens than physical place tokens. The use of *marae* to indicate a community of people only occurred in *TDP* and *TNA*. However, unclear tokens occurred in three newspapers: *TDP*, *TNA* and *HBT*. Given that these three newspapers have not only the highest frequencies of *marae*, but also the highest loanword frequencies overall in the corpus, and the highest Māori-

identifying populations, it is unsurprising that these newspapers should start to exhibit semantic divergence in the use of *marae* ahead of the remaining newspapers.

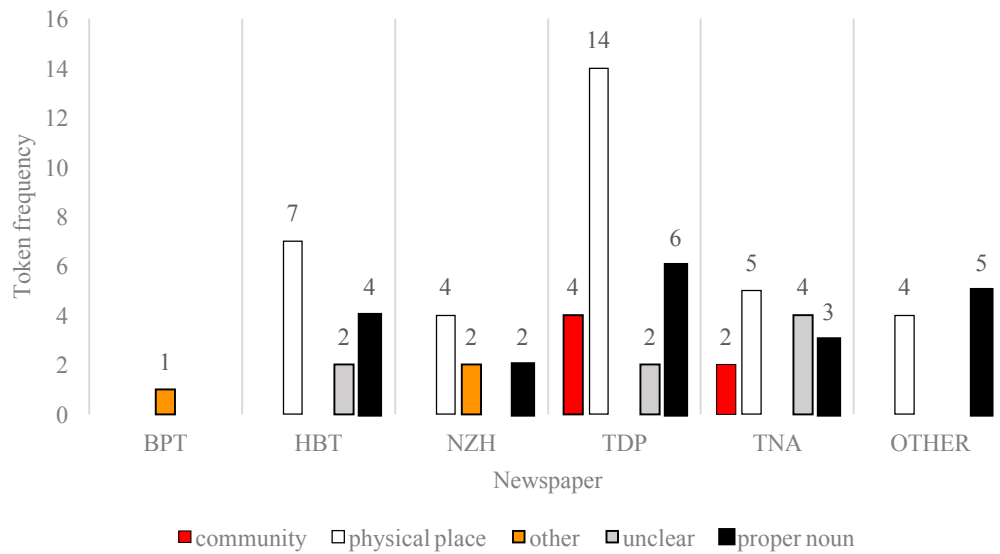


Figure 23: Semantic categorisation of *marae* by newspaper

Finally, semantic categorisation of *marae* was conducted diachronically, in order to see if there was any change across the 10-year period of the corpus data. This is shown in Figure 24.

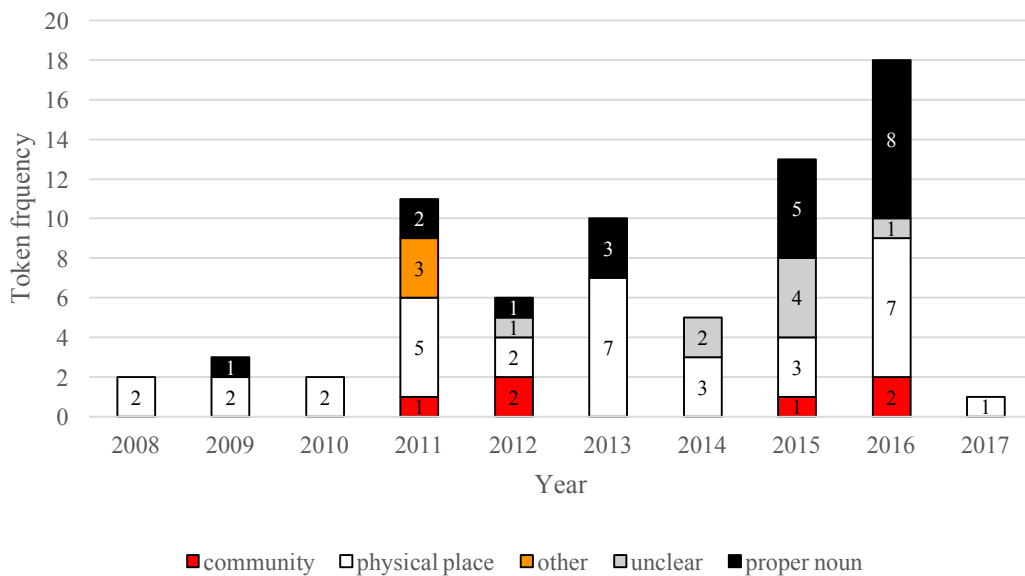


Figure 24: Semantic categorisation of *marae* by year

Only one semantic category of *marae* occurred in every single year of the corpus. Unsurprisingly, this was *marae* as a physical place. Proper nouns were the next most frequent category, occurring in just six years of 10.

Interestingly, no semantic variation beyond the traditional meaning of *marae* and the use *marae* as a proper noun appears to have existed in the corpus prior to 2011. This suggests that the semantic divergence and (thereafter) accelerated increase may have been caused by the 2011 spike in *marae* tokens. It was also in 2011 that the first occurrence of a *marae* token indicating a non-traditional semantic meaning was noted. It is possible that the increased use of *marae* in that year triggered an increase not just in frequency, but also provided the impetus for a widening in semantic variation. Possible causes for the jump in *marae* frequency between 2010 and 2011 will now be explored through loanwords that frequently co-occurred within the same articles.

5.4.4. Collocates of *marae*

Due to the nature of the data, even in articles with high frequencies of loanwords, the majority of these did not occur as near collocates. Thus, it was decided to investigate ‘collotextualisation’. This term was created to encompass whether any loanwords co-occur with another specific loanword within an entire text, “regardless of distance from each other” (A. Calude, personal communication, October 6, 2018). For the purposes of analysing *marae*, an entire text was one complete newspaper article. Co-occurring loanwords were found by starting with all articles in which *marae* occurred, and then locating all high frequency loanword types (with a frequency of 10 tokens or higher; refer back to Table 11) that occurred in those same articles. The following Table 30 includes all loanwords resulting from these criteria that co-occurred with *marae* five times or more.

It should be noted that *Maori* and *(te) reo* were excluded from this analysis, as they have already been explored in Section 5.3. Given that this corpus is constrained to a topic related to language, it is highly likely that *marae* also has a strong attraction to language-related loanwords. However, this issue is not considered here.

Table 30: Most frequent loanwords to co-occur with *marae* in the corpus

Co-occurring loanword ⁴²	Frequency in <i>marae</i> articles
<i>Maori</i>	240
<i>te reo</i>	129
<i>reo</i> (no preceding <i>te</i>)	20
<i>manaakitanga</i> (hospitality)	20
<i>iwi</i>	14
<i>Pakeha</i>	13
<i>kohanga reo</i>	13
<i>whanau</i>	11
<i>kapa haka</i>	10
<i>kura</i>	9
<i>hapu</i> (sub-tribe)	9
<i>waiata</i> (song)	7
<i>moko</i> (tattoo)	7
<i>kohanga</i>	6
<i>kura kaupapa</i> (<i>Maori</i>)	5
<i>kuia</i>	5
<i>pepeha</i> (tribal saying or motto)	5
<i>tangi(hanga)</i>	5

Interestingly only four of the loanwords in Table 30 that co-occur in articles with *marae* (*iwi*, *hapu*, *kuia*, *tangihanga*) appear on the much larger list of semantically related concept loanwords given by Degani (2017) that included: *iwi*, *hapū*, *tikanga*, *kawa*, *tangihanga*, *karanga*, *kuia*, *pōwhiri*, *tangata whenua*, *whānau pani*, *kaikorero*, *manuwhiri*, *tipuna*, *ope*, *manaaki*, *wānanga*, and *kīngitanga* (p. 677)⁴³.

5.4.4.1. *Marae and manaakitanga*

When the types *Maori* and (*te*) *reo* are excluded, *manaakitanga* is the most frequent loanword type to co-occur with *marae*. In fact, of the 21 times *manaakitanga* occurs in the corpus, 20 of those times it is in the same article as *marae*. Additionally, all 20 of those tokens occur in 2011, the year with the spike in *marae* token frequency.

It is here that year theme data collected from the corpus (Section 4.5) will be drawn upon. In 2011, the theme chosen by Te Taura Whiri i te Reo Māori (the Māori Language Commission) for Māori Language Week was *manaakitanga*

⁴² Previously unmentioned loanwords and semantically polysemous loanwords are defined in Table 30. For all other loanword definitions, see Appendix A.

⁴³ For definitions of loanwords mentioned by Degani (2017), see Moorfield (2018).

(hospitality). While *marae* occurs in many articles in which *manaakitanga* does not, it also appears that *manaakitanga* is attracted by the loanword *marae*. This correlation can be seen in Table 31, where all six articles which feature one or more non-proper noun *marae* token from the year 2011 are compared with all four articles with *manaakitanga* tokens.

Table 31: Comparison of *marae* and *manaakitanga* co-occurrences in 2011

File name	<i>marae</i>	<i>manaakitanga</i>	Other information
MLW_2011_04_JUL_TDP_007	1	12	
MLW_2011_05_JUL_BPT_001	1	0	1x uncollected <i>manaakitanga</i> token as year theme
MLW_2011_05_JUL_NZH_002	2	0	Topic: 2011 Rugby World Cup
MLW_2011_05_JUL_TDP_001	1	3	
MLW_2011_06_JUL_TDP_008	1	5	
MLW_2011_12_JUL_TNA_001	3	0	Topic: restoration of a <i>marae</i> orchard
MLW_2011_05_JUL_TDP_003	0	1	Topic: Māori Language Week activities
Totals	9	21	

While the correlation between the two loanwords is not exclusive – *marae* occurs in two articles that *manaakitanga* does not, while *manaakitanga* occurs in one article that *marae* does not – this data would suggest that there is a strong effect of topical attraction between the two types. This is unsurprising when one considers that the *marae* is indeed where hospitality is often offered in Māori culture.

This link between *marae* and *manaakitanga* is only strengthened when considering the article MLW_2011_05_JUL_NZH_002. While the loanword *marae* may ordinarily not be expected to be used in conjunction with the Rugby World Cup, this article in fact was responsible for two examples of semantic extension of *marae*, as seen in (14) and (15). Upon investigation, this was due to the expressed wish of an interviewee to extend New Zealand hospitality to the world while hosting the Rugby World Cup; thus, the meaning of *marae* itself was extended to encompass the whole world in ‘global *marae*’.

(14) “Maori TV aiming to be **global marae** of the Cup”

(MLW_2011_05_JUL_NZH_002).

(15) ““One of our aims was to be the **global marae** of the Rugby World Cup.””

(MLW_2011_05_JUL_NZH_002).

Consequently, it seems that the increase in *marae* tokens in 2011 is the result of a double topical effect: that of association with the Māori Language Week theme *manaakitanga*, and with the Rugby World Cup.

5.4.4.2. Other loanword collocates with *marae*

While no other loanword has nearly as strong a correlation with *marae* as *manaakitanga* in the corpus, a brief analysis of the semantic fields of the remaining most frequent collocates indicates that *marae* is most likely to co-occur with loanwords semantically associated with community or education, followed by activities that may traditionally be conducted on the *marae*. This can be seen in Table 32. Once again, co-occurrences of *marae* with *Maori* and *(te) reo* were not considered.

Table 32: Semantic field frequencies of loanword collocates with *marae*

Semantic field and loanword type	Type frequency	Token frequency
Activities		
- <i>kapa haka</i>		
- <i>waiata</i>	4	27
- <i>pepeha</i>		
- <i>tangi(hanga)</i>		
Community		
- <i>iwi</i>		
- <i>whanau</i>	4	39
- <i>hapu</i> (sub-tribe)		
- <i>kuia</i>		
Culture		
- <i>moko</i> (tattoo)	1	7
Education		
- <i>kohanga reo</i>		
- <i>kohanga</i>	4	33
- <i>kura</i>		
- <i>kura kaupapa</i> (Maori)		
Hospitality		
- <i>manaakitanga</i>	1	20
People		
- <i>Pakeha</i>	1	13
	15	139

The analysis of semantic field indicates that *marae* co-occurs in articles most often alongside loanwords associated with community, as the *marae* is a traditional locus of the Māori community. It follows that activities performed on a *marae* would be attracted by this loanword, especially *pepeha* (a tribal saying, motto or proverb, traditionally spoken on the *marae*) and *tangi(hanga)*. It is predicted that further

analysis into lower frequency loanwords that co-occur with *marae* in the corpus would reveal many more social culture loanwords of activities which may be conducted on a *marae*, for example *mihi(mihi)* and *hui*; as well as associated material culture loanwords including *whare*, *wharenui* and *paepae*⁴⁴. Unfortunately, however, such a detailed analysis is beyond the scope of this study.

In contrast to the above findings, what was not anticipated was the presence of so many *marae* collocates in the semantic field of education (see Table 33 for a full list). However, on closer examination, it appeared that this was likely due to the researcher viewing the data through a Pākehā lens, as an explanation for this correlation was clearly visible in the articles themselves. This will now be demonstrated.

One article mentions the *marae* as being key in the Māori renaissance and in the establishment of *kōhanga reo*⁴⁵:

(16) “In the early 1980s, the native-speaking elders created the **kohanga reo** movement, in which mostly untrained parents and grandparents volunteered to raise children in Maori-speaking settings mostly on **marae** and in private homes”
(MLW_2015_27_JUL_NZH_006).

Further investigation when including both proper and non-proper noun *marae* tokens shows that there is a strong link between *marae* and te reo Māori education:

(17) “Last week Te Wharekura o Ngati Rongomai celebrated the **kura**’s first anniversary, situated on the beautiful shores of Lake Rotoiti at Tapuaekura Marae. The **kura** teaches the tamariki, New Zealand's other official language, Maori”
(MLW_2009_28_JUL_TDP_005).

(18) “Matt O’Dawd, since taking over as principal of Kimi Ora, assured us from the outset that Kimi Ora [primary school] and Te Aranga [marae] would eventually become as one”
(MLW_2016_06_JUL_HBT_001).

(19) “Separate from the Budget the Government was also considering an additional \$15.2 million for communities’ engagement in learning te reo Maori in homes and on **marae**”
(TW_2013_01_JUL_NZH_001).

⁴⁴ See Appendix A for definitions.

⁴⁵ In the following examples, relevant loanwords are given in bold, while relevant proper noun loanwords are underlined.

Table 33: Comparison of marae tokens and education collocates

No. of file	File code	<i>marae</i>	<i>kohanga reo</i>	<i>kura</i>	<i>kura kaupapa (Maori)</i>	<i>kohanga</i>	Total loanwords per file
1	MLW_2015_27_JUL_NZH_006	2	1	0	0	6	9
2	TW_2014_26_JUL_TDP_010	2	3	1	2	0	8
3	MLW_2015_27_JUL_TNA_001	5	0	0	0	0	5
4	MLW_2009_28_JUL_TDP_005	1	0	4	0	0	5
5	MLW_2016_06_JUL_HBT_001	2	0	1	1	0	4
6	MLW_2012_23_JUL_HBT_002	2	1	0	0	0	3
7	MLW_2013_01_JUL_HBT_002	2	1	0	0	0	3
8	MLW_2015_30_JUL_NZH_002	1	1	1	0	0	3
9	MLW_2016_10_MAY_TDP_004	1	1	1	0	0	3
10	MLW_2016_30_JUN_TDP_002	1	1	1	0	0	3
11	TW_2012_18_JUL_TDP_003	1	2	0	0	0	3
12	TW_2013_01_JUL_NZH_001	1	1	0	1	0	3
13	TW_2014_24_JUL_TDP_006	1	1	0	1	0	3
	Total tokens per type	22	13	9	5	6	

Finally, the semantic extension of *marae* to include a community of people is often mentioned in conjunction with other language education organisations:

(20) “The commission is calling on all groups from the public and private sectors as well as universities, **wananga**, **kura**, play centres, kindergartens, **kohanga reo**, sports teams, **kapa haka** groups, **marae** and **hapu** to join the parade”
(MLW_2016_30_JUN_TDP_002).

(21) “Te Puni Kokiri Te Arawa regional director Wally Tangohau said it gave out free resources to **kohanga reo**, schools, organisations and **marae** wanting to learn te reo”
(TW_2012_18_JUL_TDP_003).

In light of this evidence, it makes sense that it should be *marae*, the loanword with the sixth highest frequency in the corpus, and not any of the loanwords that occur with greater overall frequencies (*iwi*, *whanau*), that is the most prevalent loanword outside of *Maori* and (*te*) *reo* to occur in every single year of the corpus. This is because *marae* is not just a centre of the Māori community, coming to represent by semantic extension the community itself, but it is also the centre of the Māori-language learning community. In a topic-bound corpus on Māori Language Week, which celebrates and discusses te reo Māori in its many facets, it is only natural that the *marae* would be a prominent and central part of the corpus. These findings make sense when interpreted in light of Degani (2017), who found “the partial identification of *marae* with a specific type of educational setting that is intended to foster cultural knowledge” (p. 676).

However, this was not the only cultural conceptualisation demonstrated by Degani (2017) to be associated with *marae*; in fact, it is shown to have a much wider and more complex interrelation with other cultural schema than occurs in the present data.

In her study exploring the use of the loanword *marae* in the narratives of young Māori women, Degani (2017) uses methods from the emerging field of cultural linguistics to construct a more expansive cultural schema. The loanword *marae* is shown to be conceptualised in a variety of sophisticated schema, including being an expression of tribal identity; a place of belonging and of reinforcing social connections; a means of expressing connection to the land and ancestors; a place of spiritual significance which reinforces cultural values, rituals and customs; a place

of education as mentioned above, where Māori cultural knowledge can be preserved and continued; and a site for historical events (Degani, 2017, p. 677). Accordingly, “*marae* does not represent an easy terrain for a semantic examination, especially for a Westerner who remains an outsider to Māori culture” (Degani, 2017, p. 668). Although it should not be surprising that such complex schemas do not appear in newspapers targeted to a wider, multicultural New Zealand audience, this is markedly different to previous findings that the use of the loanword *marae* in mass media serves to reinforce stereotypical views of Māori as historically aggressive and ignorant (Degani, 2010).

Fortunately, such overt stereotypes in regards to *marae* are not present in the current findings; however, neither is the intricate tapestry of cultural schema and concepts which the young women in Degani (2017) describe. *Marae* as it occurs in the corpus most usually represents a physical place, and occasionally widens metonymically to encompass a community of people.

5.4.5. *Marae* summary

The *marae* has strong semantic associations with many aspects of Māori life, embedded in and at the centre of Māori culture. It thus is no surprise that the occurrence of the loanword *marae* in the corpus is so frequent. The use of *marae* has increased diachronically since 2008, with a sharp increase occurring after 2011. This is attributable to the topic-driven effect of the Māori Language Week theme *manaakitanga* and the Rugby World Cup as attractor loans. *Marae* also has strong links with loanwords in the semantic fields of community and Māori language education.

However, semantic analysis of the loanword in the corpus when compared to the cultural schema expressed in Degani (2017) has demonstrated that the understanding of what it means to be at the centre of Māori life and culture is not the same to everyone. There is a divergence in the cultural schema behind *marae* according to who uses it. The meanings found in the corpus, representative of a general audience of NZE speakers, indicate that centrality to Māori culture is found in a physical place, though understanding of this is expanding over time to include social elements of a community of people with ties to that physical place, the *marae*. While this semantic extension, or ‘opening up’ likely reflects changing attitudes

and perceptions in NZE towards what it means to be Māori that were not previously evident, they are still very much indicative of a non-Māori worldview.

With the increase in diversity of the semantic meaning of *marae* in newspapers, it is possible that the meaning will continue to widen until it becomes more closely analogous to those in Degani (2017). However, given that the young women Degani (2017) interviewed all described deep held values and formative memories associated with *marae*, it is unlikely that the same semantic expression and diversity will be attained in NZE without a seismic cultural shift. If *marae* is to be widely understood by the average New Zealander as having the same cultural schema as what it does to these young women, life and cultural experience is required. And for many, who have not had or do not have access to such experiences of growing up linked to a *marae*, this entails that the word will not acquire the same cultural schema in NZE. There is in fact a semantic divergence in the loanword *marae* which parallels cultural and ethnic boundaries.

6. Loanword entrenchment and perception

This section is concerned with loanword entrenchment, and author perceptions thereof. Before continuing, a note should be made in regards to terminology. Although the majority of the literature uses the term ‘integration’ to address the extent of loanwords establishment in a receptor language (Poplack & Dion, 2012; Sharp, 2007), it was felt that such a term was not appropriate in the discussion of Māori loanwords in NZE. As New Zealand has a history of colonisation, ‘integration’ is a word loaded with negative connotations, making it culturally inappropriate. Thus, for the purposes of the present research, the term ‘integration’ will only be used when citing other literature directly, and the term ‘entrenchment’ will be preferred by this study.

6.1. Listedness in loanwords

Listedness was measured in the loanwords in the corpus according to Muysken’s (2000) definition, using the *New Zealand Oxford Dictionary* (Deveson & Kennedy, 2005). While there was a small difference in the number of loanword types which were listed and unlisted (see Table 34), this was relatively even, with only 28 more listed than unlisted types in the corpus⁴⁶. However, in terms of tokens, the overwhelming majority were listed: 94.55% (see Figure 25). Only 207 tokens, or 5.45%, of loanword tokens were unlisted. This is conformant to Davies and Maclagan (2006), who state that language used in newspapers is more conservative than spoken language, as it is targeted at a wide audience, and because newspapers “[regard] themselves as guardians of correct English” (p. 74).

Table 34: Types and token frequencies by listedness

	Types	Tokens
Listed	107	3,588
Unlisted	79	207
Totals	186	3,795

⁴⁶ For a complete list of listed and unlisted loanwords in the corpus, see Appendix A.

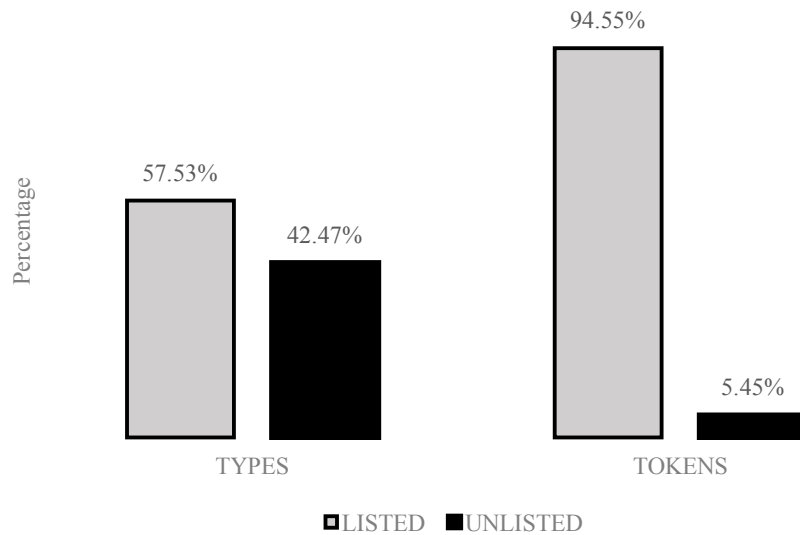


Figure 25: Percentages of types and tokens by listedness

Listedness in loanwords was further analysed according to their frequencies of occurrence. As Table 35 demonstrates, most unlisted loanword types had a low frequency in the corpus, occurring in four tokens or less, while higher frequency unlisted loanwords were quite rare (four types each in the mid- and high frequency bands), as might be expected. Interestingly, loanword types were also more likely to yield a low frequency of tokens, with high frequency-band types the next more common. Loanword token distribution behaved more predictably, with the majority of unlisted tokens (111) having a low frequency of occurrence, and the overwhelming majority of listed tokens having a high frequency (3,588). This is in line with the extremely high frequency of a small number types in the corpus; specifically, *Maori* and *te reo*.

Table 35: Listed and unlisted types and tokens by frequency band

	Listed types	Unlisted types	Listed tokens	Unlisted tokens
Low frequency (1-4 tokens)	61	71	105	111
Mid-frequency (5-9 tokens)	14	4	84	27
High frequency (10+ tokens)	32	4	3,399	69
	107	79	3,588	207

Within the low frequency band of tokens in Table 35, a considerable proportion were single-occurrence tokens only: 79 tokens in total.

Nonce loanwords are defined by Poplack and Dion (2012, pp. 285-286) as unattested loanword types which occur once and once only, used by only one person, in a corpus. The majority of nonce loanwords in a corpus are ephemeral: that is, they will not recur after their initial manifestation (Poplack & Dion, 2012). Nor do they “generally become widespread”, and neither are they predictable (Poplack & Dion, 2012, p. 287). This supports findings in the present data, in which 47 single-occurrence loanword tokens (59.5%) were unlisted and can therefore be classified as nonces. However, there were also 32 single-occurrence loanword types which were listed. These were not considered by Poplack and Dion (2012) in their research. However, it is interesting that a large proportion of these single-occurrence listed loanwords are for the most part flora and fauna (11 tokens) such as *kiwi* and *kereru*; formulaic expression (four tokens) such as *ka kite* and *aue*; or material culture loanwords specific to Māori art (four tokens), including *koauau*, *kowhaiwhai*, *tukutuku*, and *taiaha*⁴⁷. These low frequency, listed loanwords are all extremely culturally specific, and thus unlikely to occur outside of related topics.

There were also four notable unlisted loanword types with high frequencies (10 or more tokens): *Matariki* (Māori New Year, 26), *manaakitanga* (21), *waka ama* (11) and *pepeha* (11).

It has already been shown that the high frequency of *manaakitanga* tokens can all be accounted for by a topical effect of attraction, as the explosion in use in 2011 (and only 2011) was primarily due to *manaakitanga* being the theme for Māori Language Week that year. Use of *Matariki*, too, has been increasing since 2007 (Calude et al., Forthcoming, p. 5). *Waka ama* and *pepeha* were less expected as high frequency unlisted loanwords, and it is likely that they are recent borrowings, given that neither of them occurred in the corpus prior to 2012. Both types also had the highest token count in their first year of occurrence, suggesting their borrowing was actuated by an initial, high frequency of use, which triggered subsequent use and may result in entrenchment in the years to come if use is maintained or increases. This is reinforced by Poplack and Dion (2012), who found that in almost all cases, even nonce items were immediately established in the receptor language. They challenged the notion that nonce items gradually become more frequent and

⁴⁷ See Appendix A for definitions.

diffused over time, and that, with very few exceptions, entrenchment is not gradual, but sudden and definite.

Following this, the four unlisted loanword types with mid-range frequencies of occurrence – *mahi* (work, 9 tokens), *kupu* (word, 7), *wananga* (university, 6), and *tikanga Maori* (Māori custom, 5) – can be claimed to be already established in NZE and becoming more frequently used, rather than on the increase towards becoming established. While *mahi* did not occur before 2012, the remaining three types had irregular use from the first year of the data, 2008. It is suggested that their high frequency use may be triggered by an appropriate topic, as with *Matariki* becoming more salient and frequent parallel to its yearly celebratory events gaining popularity. However, this must be ongoing, as the one-off year use of *manaakitanga* proves.

More data are needed to corroborate these findings, however, they provide strong evidence for several recent borrowings into NZE, and demonstrate that loanword frequencies compared with listedness is a sound method of positioning loanword entrenchment within a receptor language. This concept will be referred to in coming sections.

6.1.1. Listedness in newspapers

Listedness in newspapers will also be briefly dealt with, as it reinforces previous sections' claims that different newspapers are possibly influenced by ethnic demographics of their readerships. The highest number of unlisted tokens was used by *TDP* (see Table 36), which is unsurprising given the majority of loanwords occurred in that newspaper, and that they appear to be functioning as a teacher of loanwords to their readers. Likewise, the *BPT* and *NZH* had the lowest rates of unlisted loanwords, in line with their having the lowest percentage of Māori population in their readership, and accordingly preferring to use only those Māori loanwords seen as most standard. The *HBT* and *TNA* accounted for approximately 20% each of unlisted loanwords. This was not as high as *TDP* in its loanword teaching function, but still higher than all other newspapers in parallel with their higher Māori populations.

Distribution of listed and unlisted loanword tokens was also investigated within each newspaper. A breakdown of these figures is given in Table 37.

Frequencies of unlisted tokens are in line with general loanword frequencies established in Section 5.

Table 36: Distribution of unlisted tokens by newspaper

Newspaper	No. of unlisted tokens	Percentage of unlisted tokens
<i>TDP</i>	75	36.23%
<i>HBT</i>	44	21.26%
<i>TNA</i>	38	18.36%
Other	26	12.56%
<i>NZH</i>	13	6.28%
<i>BPT</i>	11	5.31%
	207	100%

Table 37: Listed and unlisted loanword token frequencies by newspaper

Newspaper	Unlisted tokens	Listed tokens	Total loanword tokens
<i>TDP</i>	75	1,524	1,599
<i>HBT</i>	44	517	561
<i>TNA</i>	38	440	478
Other	26	235	261
<i>NZH</i>	13	446	459
<i>BPT</i>	11	426	437
	207	3,588	3,795

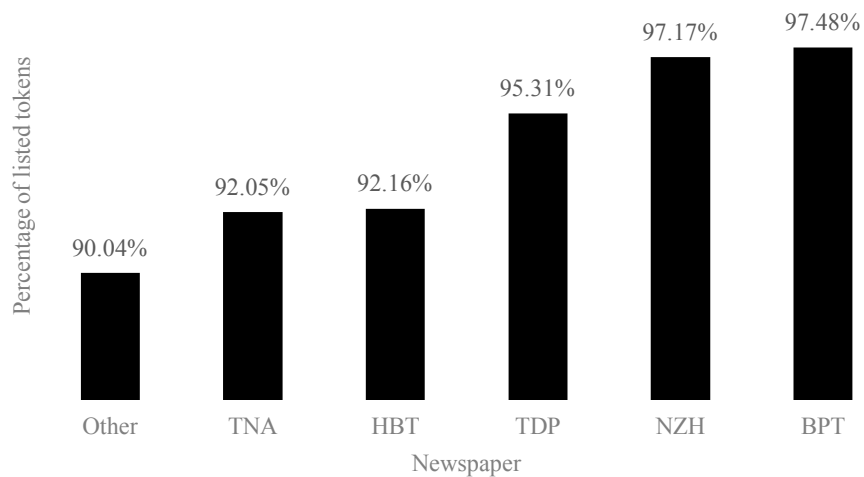


Figure 26: Percentage of listed tokens by newspaper

Additionally, at least 90% of all loanword tokens were listed in every newspaper (Figure 26). *TNA* had the lowest rate of listed tokens, followed by *HBT*; while *BPT*

and *NZH* had the highest. Once again, this is in line with previous findings of loanword use in newspapers, and also serves to reinforce that newspapers prefer to use listed, standardised loanwords, but that demographics of target audiences have a significant impact on to what extent this is done.

6.1.2. Diachronic effect of listedness

The final aspect of listedness to be touched on here before addressing it in conjunction with other variables is that of diachronic effect. A comparison of listed and unlisted loanword token frequencies according to year can be seen in Figure 27.

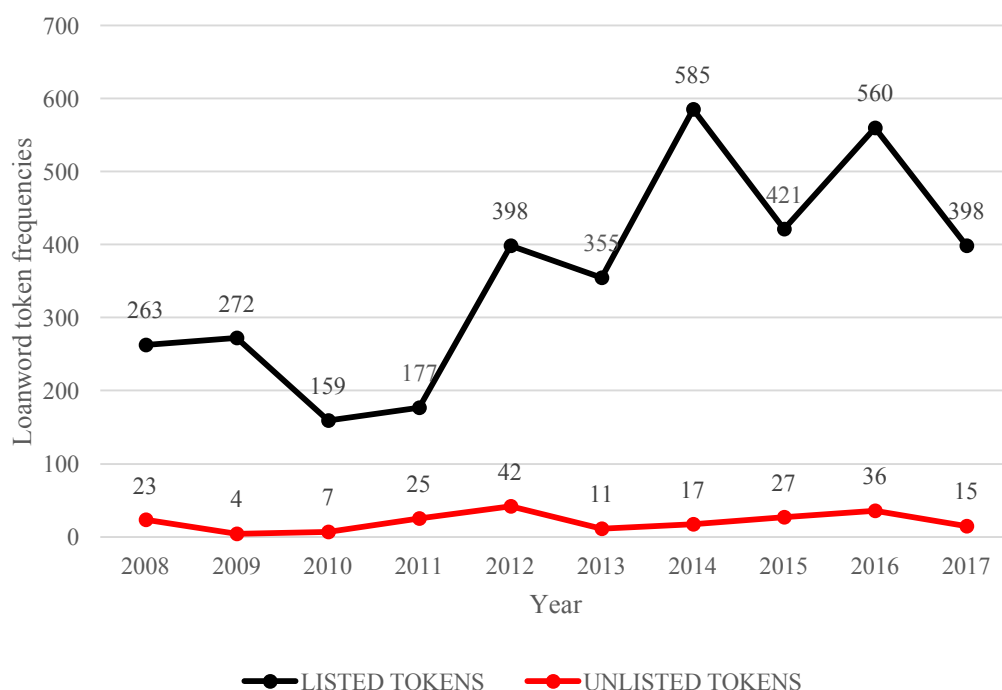


Figure 27: Listed and unlisted loanword token frequencies per year

Unlisted tokens appear to be somewhat unpredictable, with spikes in use in 2012 and 2016. Interpreting an increase or decrease is difficult due to this variation. Following a Kendall Tau analysis, it was found that there has been a slight increase in unlisted tokens, but that this is not significant ($\tau=0.289$, $p=0.28313$). However, there is a clear increase in use of listed loanwords, and this is confirmed as being just significant ($\tau=0.539$, $p=0.038879$).

When compared to the overall significant increase in loanword tokens in the corpus, these results indicate that it is dictionary listed, attested loanwords which

are being used more frequently in newspapers over recent years, and not unlisted loanwords. When unlisted loanwords are used, they are unpredictable in their occurrence, and much lower than their listed counterparts. This lines up with existing evidence that mainstream media in NZE prefers to use Māori loanwords which are already well established in the language. It also demonstrates that the recent diachronic increase in loanword use does not appear to be introducing more loanwords to NZE in any great and consistent capacity.

6.2. Marked loanwords

6.2.1. Summary of terms

During the course of the data collection, it became obvious that many Māori loanwords were translated, defined, or otherwise stylistically presented in a way that indicated the word does not belong to the English language. Although much of the literature uses the term ‘flagging’ for this phenomenon (Poplack et al., 1989; Sharp, 2007; Turpin, 1998), the same literature also in most cases discusses loanwords and codeswitching in a spoken context, often bilingual. Additionally, flagging is also used in the context of written loanwords in Grant-Russell and Beudet (1999); although Kruger (2012) terms the same thing ‘markedness’. Some studies on Māori loanwords mention the glossing of loanwords in written texts (see Davies & Maclagan, 2006; Degani, 2010; Macalister, 2009), however in the present corpus, glossing was not the only mode of flagging in the corpus. To differentiate the use of written flagged loanwords from spoken flagged loanwords, therefore, a new flagging typology was created, and a different methodology of classification was devised⁴⁸.

This new typology can be seen in Table 38, which gives a summary of the various modes by which a loanword may be marked as not belonging to or originating in the receptor language by a speaker or author. As the grey highlighted boxes indicate, loanwords in written language are only concerned with discourse markers such as synonyms or definitions, and typographic markers – all physical marks made within the text. It is proposed that to better distinguish these modes from those in spoken discourse, any loanword in written text which is marked as

⁴⁸ Acknowledgement and thanks must be given to Andreea Calude for her help in the creation of this typology.

belonging to or originating from a language than the receptor one would fall exclusively under the term ‘markedness’; while ‘flagging’ should refer only to those loanwords (or, by extension, codeswitches) marked in spoken language.

Table 38: Typology of loanword marking in spoken and written language

	Phonological	Discourse	Typographic	Physical	Term in the literature	Term in this study
Spoken language	✓	✓	✗	✓	Flagging	Flagging
Written language	✗	✓	✓	✗	Flagging, markedness, glossing	Markedness
Examples	<i>Phonological system change, accent, tone, pauses, hesitations</i>	<i>Explanations, translations, definitions, synonyms</i>	<i>Brackets, dashes, italics, bold font, quotation marks</i>	<i>Quotation marks created in air with fingers</i>		

Next, an annotation system was adapted from Kruger’s (2012) study on translated children’s literature in South Africa, in which loanwords were described as being ‘marked’ if they were glossed, italicised, or translated in texts, and ‘unmarked’ if “the loan word is transferred directly to the target text, without explication or glossing, and usually without marking by italics or quotation marks” (Kruger, 2012, p. 197).

Thus, Māori loanwords in this study were categorised as either ‘marked’: having any textual or stylistic indication that the loanword was of foreign origin; or ‘unmarked’: the loanword was treated as any other NZE word without textual or stylistic remark. Two sub-types were further devised for the purposes of this study: ‘graphically marked loanwords’ and ‘textually marked loanwords’.

Graphically marked loanwords (GML) were any loanwords marked in the text by quotation marks, brackets, dashes, bolding or italicisation. As it was difficult to differentiate with any certainty between use of commas for loanword marking or for standard punctuation, they were excluded from GML categorisation.

Textually marked loanwords (TML) were loanwords that had been translated, defined or explained within the text itself. Translations and definitions were included as TMLs regardless of the accuracy of the translation or definition given by the author. At least two loanwords were included as TMLs where redundancies occurred, i.e., ‘*kura* school’.

As coding of the marked loanwords progressed, it became apparent that there were no exclusive GMLs as such. Any loanword which was only marked graphically was usually not being marked as a foreign word; rather, it was being marked as the focus of a sentence as any NZE word might be. Nor was italicisation a feature in any of the articles. Furthermore, where graphical marking of a loanword did occur, this was always in conjunction with textual marking. A third category was therefore created: mixed-mode marked loanwords (MML).

An illustrative example of each of these categories of markedness is given in Table 39.

Table 39: Examples of markedness by mode

Mode	Abbr.	Example
Graphically marked loanword	GML	a. Most New Zealanders would acknowledge the Māori language is a <i>taonga</i> that should be protected. b. Most New Zealanders would acknowledge the Māori language is a “taonga” that should be protected.
Textually marked loanword	TML	Most New Zealanders would acknowledge the Māori language is a taonga, a treasure, that should be protected.
Mixed-mode marked loanword	MML	a. Most New Zealanders would acknowledge the Māori language is a taonga (a treasure) that should be protected. b. Most New Zealanders would acknowledge the Māori language is a treasure – a taonga – that should be protected.
Unmarked loanword	n/a	a. Most New Zealanders would acknowledge the Māori language is a taonga that should be protected. b. People are saying nothing but “taonga”

Note: commas were not considered as representations of graphical marking, as it was found to be impossible to differentiate their use in marking from punctuation.

A final note regarding marked loanwords is that the proper noun phrase ‘Maori New Year’, a term which was excluded from the loanword data, is included in the marked data where it occurred as a translation of *Matariki*.

With these terms established, the marked loanword data will now be considered.

6.2.2. Mode of markedness

There were 111 marked loanword types in the corpus, with 177 marked tokens. It is clear from Figure 28 that the majority of marked loanword types are always MML (mixed-mode, or both textually and graphically marked). 75 types or 67.6% of marked types were always MMLs. Only 21.6% (24 types) of types were exclusively TML (textually marked loanwords), while 12 types (10.8%) occurred as either in the corpus.

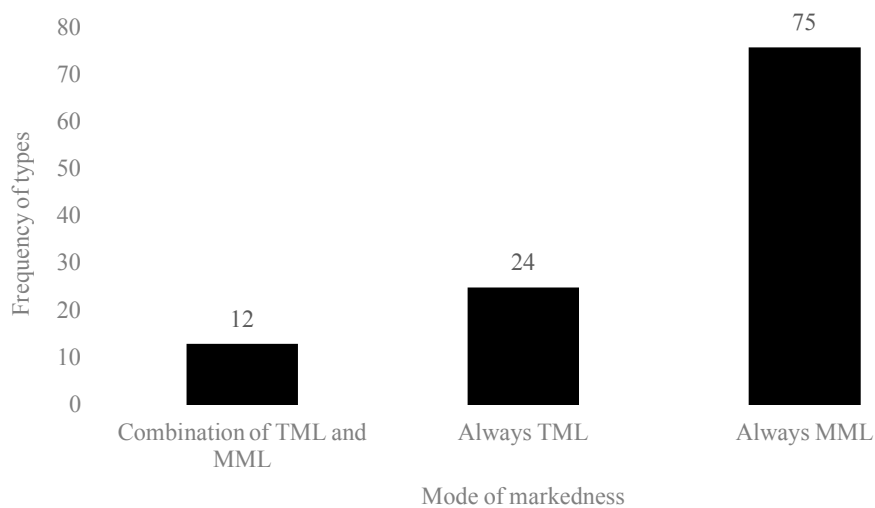


Figure 28: Marked type frequencies by mode of markedness

This preference for using MML when marking loanword types is sustained when individual marked tokens are viewed diachronically in Figure 29. There are only two years in which MMLs have a lower frequency than TMLs: 2011, in which they were equal at six tokens each, and in 2015. Additionally, no significant diachronic trends according to mode of markedness were found when Kendall Tau analysis was made.

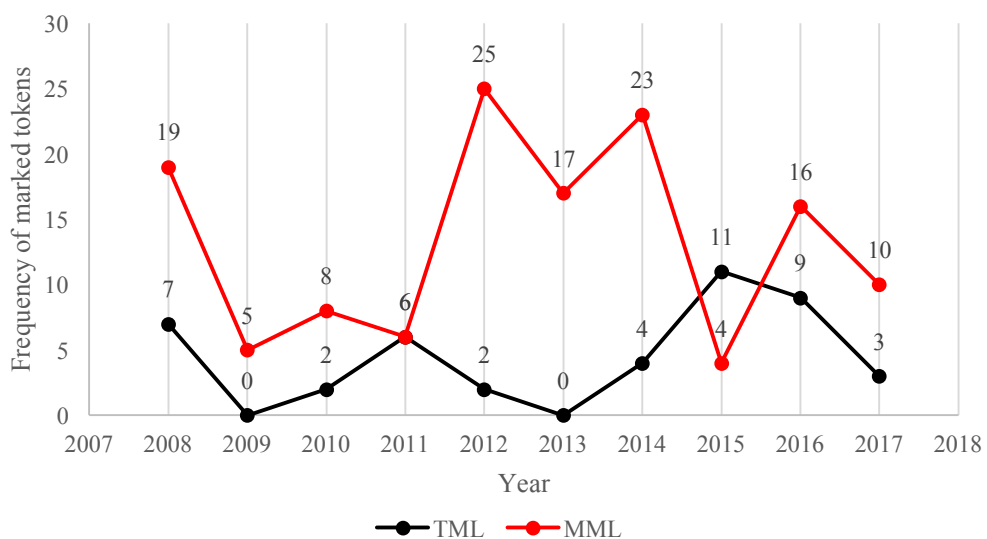


Figure 29: Diachronic frequencies of marked tokens by mode of markedness

Therefore, it can be stated that when marking loanwords in newspapers, there is a definite preference to do so by translating or defining a loanword in a combination

of textual and graphical modes (MML). Additionally, the mode of marking loanwords in newspapers is not subject to any diachronic influence in the current data.

6.2.3. Language order in marked loanwords

The language order of translations and definitions made in the marked loanword data will now be briefly evaluated using two different methods. In the first, a straightforward analysis will be made of the MML and GML data in the corpus. In the second, a slightly different approach will be taken. Data that were not included in the corpus frequency counts will be considered, in the form of the search terms used to construct the corpus itself.

6.2.3.1. Translations of marked loanwords

Marked loanwords that were translated or defined (both GMLs and MMLs) were categorised according to the order of language this was done in. Essentially, this necessitated coding whether the loanword occurred before an English translation (Māori to English in direction), or whether it was given as the translation of an English word (English to Māori in direction). For instance, in the example sentence ‘the Māori language is a *taonga* (treasure)’, the loanword *taonga* occurs before its English translation, and so would be categorised as being ordered Māori to English.

A representation of the number of marked tokens in each of these directional categories can be seen in Figure 30. The majority of marked loanwords in the data occurred in their Māori form first, and were then defined or translated into English. In only 15 tokens was the reverse true, where an English word was translated into its Māori loanword equivalent. When these data are given in percentages, the difference is striking: only 0.08% of all marked loanwords occurred as a translation of an English word. This figure is even lower when it is noted that four of these tokens occurred in an article in which the loanwords were part of an English-to-Māori comprehension test (labelled ‘street poll’ in Table 40).

A list of all marked loanwords translated in the direction of English to Māori can be seen below (Table 40). With the exception of *Maori*, *iwi* and *Matariki*, all

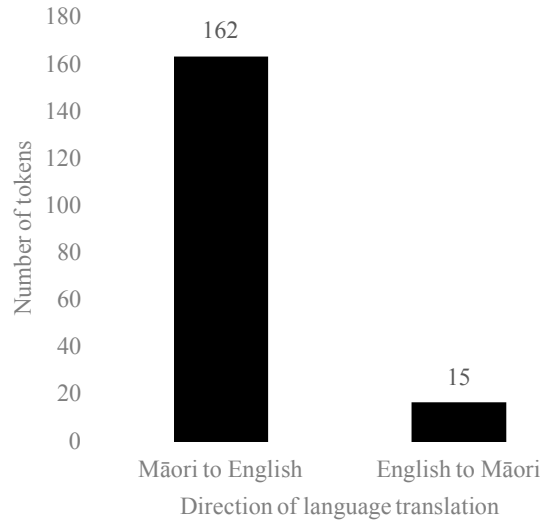


Figure 30: Token frequency of language order of translations/definitions in marked loanwords

loanwords marked in this way were low frequency or nonces. *Matariki* was the only loanword to be translated from English to Māori twice, suggesting that there may be a push for education regarding this event in line with an increase in its use in recent years (Calude et al., Forthcoming). From this, it may be inferred that marked translations in this direction are an educative device. Similarly, marked translations in the direction of Māori to English may either serve the same educative function, or simply be used in order to clarify the meaning of a loanword the author perceives will not be widely understood by their audience.

Although this analysis essentially relies on which word occurs first in the text, the Māori or the English form, it demonstrates that, despite the push for increased use of Māori loanwords in NZE, and the strong theme of the revitalisation of the Māori language within the corpus itself, it is extremely rare for an English word to be translated into Māori. It is likely that this is deemed unnecessary. When a Māori loanword is translated into English, it is usually because the author does not deem it to be comprehensible by enough of their audience to be left unmarked (whether or not this is true) and regardless of listedness; although a loanword that is both listed and marked may give an indication that an author is taking the opportunity to use it as a teaching point. Thus, markedness in either language direction may be used to perform two very similar functions: either an educative function or to fill a perceived knowledge gap. However, both cases convey a positive, supportive stance towards the Māori language.

Table 40: Summary of marked loanword data (English to Māori directional translations)

Year	Paper	Article file number	Marked token	Token freq.	Context in article
2008	<i>NZH</i>	MLW_2008_16_JUL_NZH_002	<i>hiko</i>	1	...and lightning, hiko
2008	<i>TNA</i>	MLW_2008_22_JUL_TNA_002	<i>hapu</i> (pregnant)	1	pregnant (hapu) – street poll
			<i>iwi</i>	82	tribe (iwi) – street poll
			<i>kai moana</i>	2	seafood (kai moana) – street poll
			<i>whare</i>	4	house (whare) – street poll
2010	<i>TNA</i>	TW_2010_02_AUG_TNA_001	<i>Maori</i>	1,649	Maori language (te reo)
2010	<i>TNA</i>	MLW_2010_27_JUL_TNA_004	<i>mau raakau</i>	1	Maori martial art disciplines, collectively called mau raakau ...
2012	<i>TNA</i>	MLW_2012_26_JUL_TNA_004	<i>turangawaewae</i>	2	I consider my standing place _ my turangawaewae _ to be Kamo.
2014	<i>TDP</i>	TW_2014_21_JUL_TDP_005	<i>kaiako</i>	2	teacher (kaiako)
2015	<i>BPT</i>	MLW_2015_28_JUL_BPT_002	<i>whare pukapuka</i>	2	public libraries (whare pukapuka)
2015	<i>TNA</i>	TW_2015_24_JUL_TNA_003	<i>Matariki</i>	26	Maori New Year, Matariki ,
2015	<i>TDP</i>	TW_2015_27_JUL_TDP_006	<i>puku</i>	3	...even now they are as likely to refer to their stomach as their puku ,
			<i>taringa</i>	1	their ears as their taringa
2015	<i>TDP</i>	TW_2015_29_JUL_TDP_004	<i>Matariki</i>	26	Maori New Year, Matariki .
2017	<i>WCHM</i>	TW_2017_13_SEPT_WCHM_001	<i>mita</i>	4	Maori dialects (mita)

6.2.3.2. Translation of corpus search terms

This section will deal with the translation of the two search terms used to build the corpus: ‘maori language week’ and ‘te wiki o te reo maori’⁴⁹. For the purposes of loanword frequencies, these search terms were excluded from all counts to avoid conflating the data. However, their occurrences can provide an insight into a recurrent issue found in the corpus: lexical choice. This was considered of interest as ‘te wiki o te reo maori’ appears to be a relatively recent innovation in New Zealand newspapers.

The search term ‘maori language week’ occurred in the data a total of 396 times, while ‘te wiki o te reo maori’ occurred 124 times.

As Figure 31 demonstrates below, the general frequency trends of both search terms are reflected in the number of articles per year (refer to Table 9 for diachronic frequencies). However, the search term ‘te wiki o te reo maori’ did not occur at all in the year 2009, and prior to 2011, it never occurred more than five times in a single year. This suggests that it is a relatively recent addition to NZE. In comparison, ‘maori language week’ was consistently used over the 10-year period of this study, and frequencies indicate that it is often used more than once in the same article.

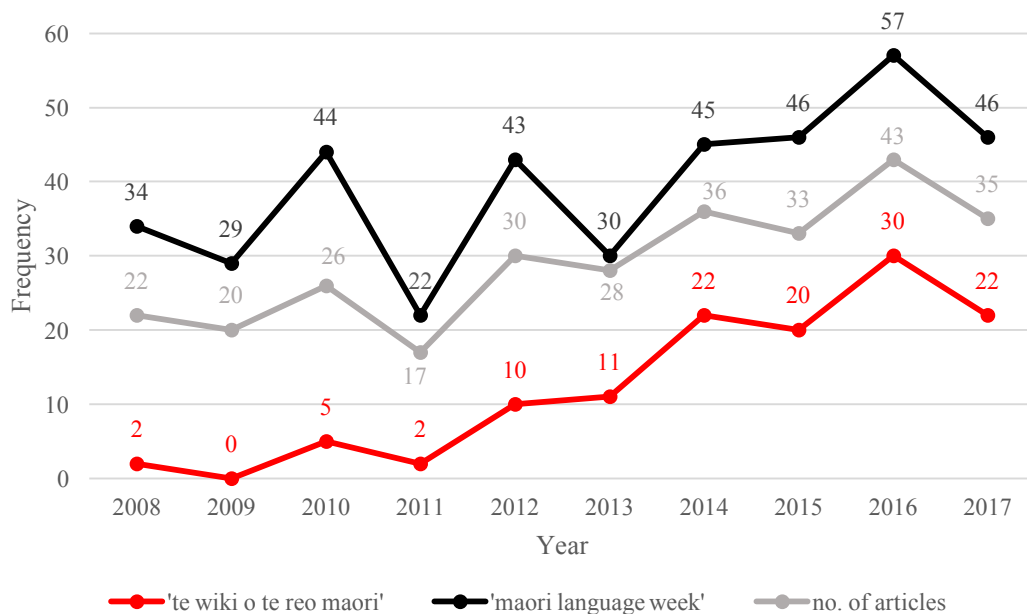


Figure 31: Comparison of article frequencies and search term frequencies

⁴⁹ As established in earlier sections, search terms did not include capitalisation or macrons.

Both search terms show a clear upward trend over time. The search term ‘maori language week’ increased slightly in frequency over the 10-year period, although this was only just significant ($\tau=0.584$, $=0.024764$). Even more notably, ‘te wiki o te reo maori’ had a clear and significant increase in use ($\tau=0.796$, $p<0.01$). From these findings, it is clear that the former phrase is well established in NZE, while the occurrence of ‘te wiki o te reo maori’ has gained acceptance in NZE in recent years, and will likely become more common in future newspaper articles.

A final point of interest regarding these search terms is the fact that two terms were even necessary, as one is simply the English translation of the Māori term⁵⁰. Thus, in any given use of these terms by an author, two choices are made: firstly, whether to use the Māori or the English term; and secondly, whether or not to translate it. Figure 32 compares the percentages of how often each search term was translated according to year.

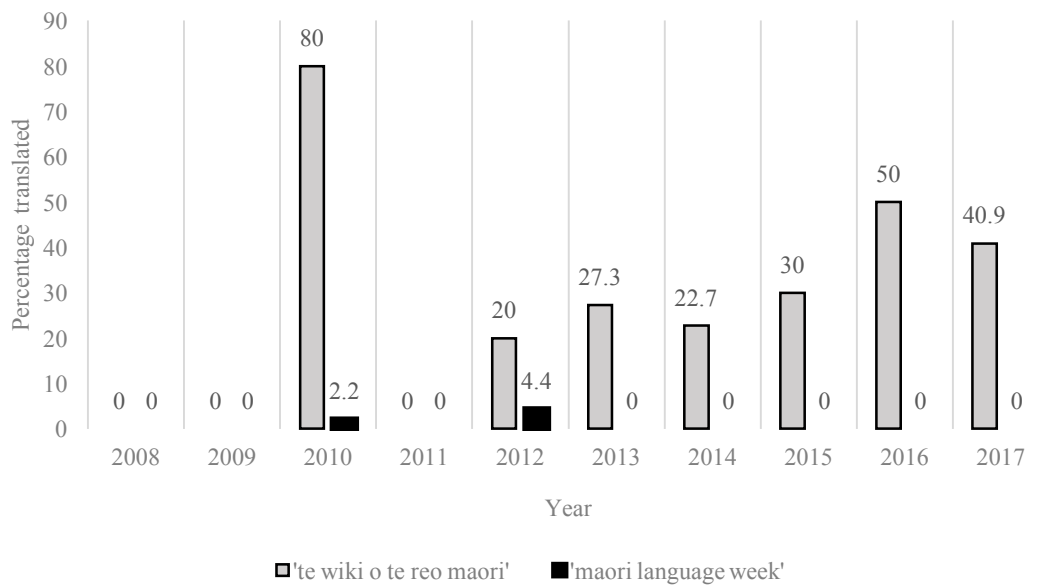


Figure 32: Diachronic percentages of translated (marked) search terms

What is startling about these percentages is that the English form ‘maori language week’ is almost never translated into its Māori equivalent. Despite occurring 398 times in the corpus, there were only three instances of a Māori to English translation

⁵⁰ Although ‘maori language week’ is technically a blend of Māori and English, and *wiki* itself is an English loan into Māori; see Moorfield (2018).

of this: once in 2010 and twice in 2012. Two were published in *TNA* by different female journalists of unknown ethnicity. The third occurred in the *HBT*, written by a male non-journalist who had attended a Māori-English bilingual primary school (additionally, a full translation of the article was provided by him in Māori). The article topics were respectively: the events of Māori Language Week at a non-Māori language pre-school; a traditional Māori sport being played in the UK; and a personal story of Māori language learning through *waka ama*. Thus, there is little overlapping information to provide a hypothesis for these anomalies based on author. However, these instances all came from the two newspapers with the highest regional Māori population. Based on this small slice of information, it might be expected that Māori to English translations of ‘maori language week’ are most likely to occur in regions with a high Māori population, and unlikely to occur in national newspapers.

It is predictable that ‘te wiki o te reo maori’ is translated into English much more often, being translated on average in 35% of all occurrences. While this entails that just under two-thirds of these search phrases are not being translated into English, what is less encouraging is that, as the usage of the Māori term becomes more common, so too it seems, does the likelihood that it will be accompanied by an English translation. Although it is becoming more prevalent in NZE, the onus seems to remain on the Māori form to make itself understood, while translations from English to Māori are perceived as unnecessary. A positive supposition for this could be that ‘maori language week’ is more textually economical than ‘te wiki o te reo maori’, and that some effort is beginning to be made to grow awareness of the Māori form; however, it is more likely that little effort is being made to teach or promote the Māori variant when the English will suffice.

6.2.4. Marked loanwords in newspapers

An analysis was next conducted of the frequency of marked loanwords according to the newspaper of publication they occurred in. A visual representation of this can be seen in Figure 33, where all individual newspapers with marked loanwords are considered, not just the main groupings as dealt with in earlier sections. As expected, the highest number of both marked types and tokens occurred in the newspaper *TDP*. This was congruent with *TDP* having the highest number of both articles and general loanword types and tokens in the entire corpus.

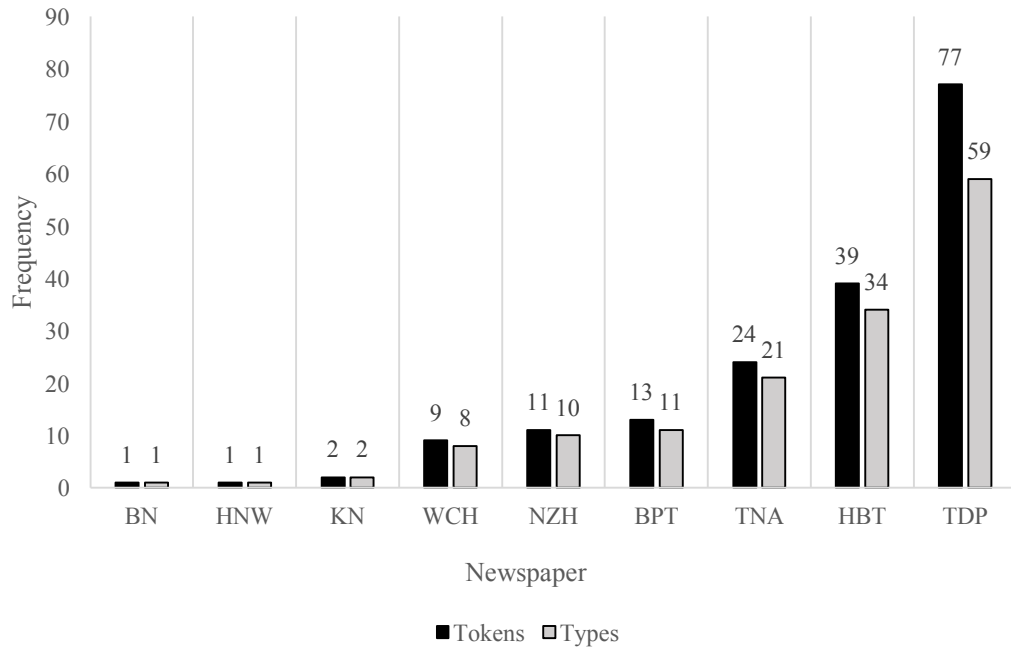


Figure 33: Marked token and type frequencies, by newspaper

Following this, a comparison of marked tokens to overall number of tokens per newspaper, along with their percentages, was made in Table 41.

The average percentage of marked loanwords (out of all loanwords) across the corpus was 4.66%. However, this figure varied according to newspaper. It was the *HBT*, the newspaper with the second highest number of loanwords, that was the most likely to mark loanwords, with a rate of 6.95%. *TNA* was the next most likely to mark loanwords, at 5.02%, but only won out by a slim margin. The newspaper with the greatest number of loanwords (*TDP*) marked them at a rate of 4.82%, which was still above the corpus average. The two newspapers with the lowest percentages of marked tokens were the *BPT* and the *NZH*; both newspapers having the lowest frequencies of overall tokens as well.

Table 41: Marked and unmarked token frequencies, by newspaper

Newspaper	Marked tokens	Loanword tokens	% of marked tokens
<i>TDP</i>	77	1,599	4.82%
<i>HBT</i>	39	561	6.95%
<i>TNA</i>	24	478	5.02%
<i>BPT</i>	13	437	2.97%
Other	13	261	4.98%
<i>NZH</i>	11	459	2.40%
	177	3,795	4.66%

These findings demonstrate that number of loanwords are potentially correlated with how many will be marked. The greater the number of overall loanwords, the more likely they are to be marked, while newspapers with lower frequencies of loanwords are also less likely to be marked.

Interestingly, in regions where there is a higher Māori population, newspapers actually marked more loanwords than in regions with a higher proportion of non-Māori. This would suggest that regional newspapers are not only actively using more Māori loanwords in their texts, but that they are also using them as a vehicle to teach. This was unanticipated, given that higher general comprehension of Māori loanwords might be expected in regions with higher Māori populations than elsewhere, and therefore need less marking.

The newspapers with the lowest frequencies of loanwords also had the lowest rates of marking. In this respect, the *NZH* is particularly noteworthy as it is the only newspaper in the corpus with a national rather than regional readership, and the only one with a remit to cater to a wider reading audience and the lowest common denominator in Māori loanword knowledge. This suggests that newspapers which do not use many loanwords (such as the *BPT* and *NZH*) are more concerned with reader comprehension than education; they use few, more well-known loanwords, only using unfamiliar loanwords when they are absolutely necessary and marking them as a general courtesy.

When analysing marked and unmarked loanword types according to newspaper (see Figure 34 for a comparison of totals), it was found that, predictably, most newspapers had more unmarked than marked loanword types. However, there was one exception to this. Beyond having the highest number of different loanword types in the corpus, the newspaper *TDP* also had seven more marked types than unmarked types. It was the only newspaper with such an unexpected result, and reinforces the supposition that *TDP* is actively seeking to teach Māori loanwords to its readership. The *HBT* had the next highest number of overall types, but only one more than *TNA*. This suggests that, although the *HBT* uses more tokens overall, and less of them are generally marked, their range of loanword types is not as broad as it might be; while *TNA* uses more unmarked types than the *TDP*.

Previous findings on the *NZH* were also strengthened here. While the *NZH* had the third lowest number of loanword types (45), it also had the lowest number

of marked types. This reinforces that the *NZH* has a lower rate of using different types in articles, but seeks to use only those it considers already known to its readership.

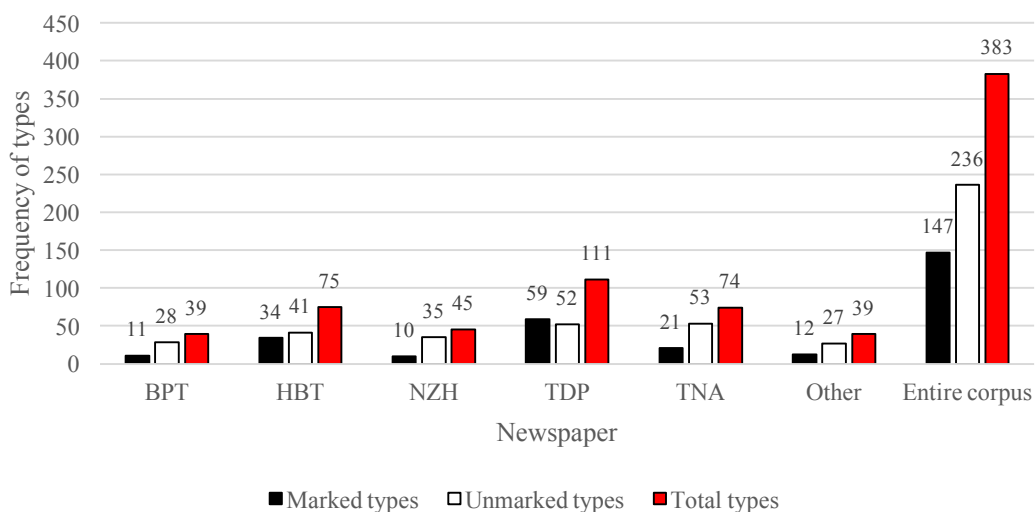


Figure 34: Comparison of marked and unmarked type frequencies, by newspaper

6.2.4.1. Diachronic marked loanword use in newspapers

Marked diachronic data according to newspaper was not considered in great detail, as no significant diachronic effect was found in overall loanword use by any newspaper. This subject will therefore be touched upon only briefly, to give a general impression of marked loanword behaviour across years in newspapers, before turning to a more in-depth study of marked loanword frequencies in general in the corpus, including diachronic data, in pursuant sections.

It was interesting that no one newspaper marked loanwords in every single year of the corpus (see Figure 35). However, *TDP* almost always marked some loanwords; only in 2009 is this not done. These data consolidate earlier findings that *TDP* is making a concerted effort over time to teach Māori words, as part of New Zealand's first bilingual city, and that markedness is one of its teaching tools. This push can certainly be seen not only in loanword behaviour, but also in the opinions given in the articles themselves, in which residents are generally making a concerted effort to revitalise the Māori language in their own city.

In comparison, diachronic markedness in other newspapers is less consistent, suggesting their marking of loanwords is only done as necessary.

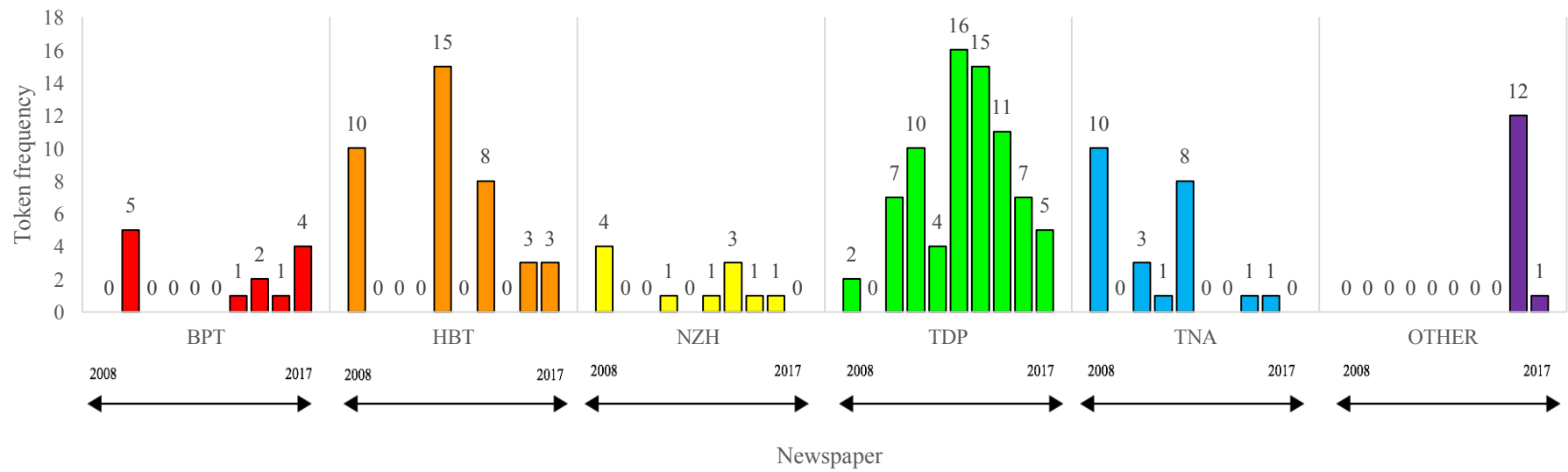


Figure 35: Diachronic marked token frequencies by newspaper

Two newspapers marked some loanwords in six years (*TNA, NZH*) and two marked some loanwords in only five years (*HBT, BPT*). This entails that the newspaper with the highest percentage of marked to unmarked loanwords (*HBT, 6.95%*), had a high concentration markedness in just a few, apparently random years.

A possible explanation for the erratic nature of diachronic marked loanword use in newspapers is that markedness is influenced not only by author perception and intent, but by article topic; and that topic, in turn, is influenced by the locale of publication. However, such an investigation is beyond the scope of this study.

6.3. Marked loanword frequencies in the whole corpus

Having dealt with modes of markedness in loanwords, language order of translations in marked loanwords, and marked loanwords in newspapers, overall marked loanword frequencies in the corpus will now be turned to. Marked loanword tokens accounted for less than 5% of all loanword tokens in the corpus. Additionally, marked loanwords occurred in less than a third of the total number of articles: 78 articles. Thus, loanwords were marked in just 26.9% of articles.

During the course of coding for markedness, 16 tokens (over 11 types) were removed due to difficulty in discerning whether they were marked or not. This left 3,779 loanword tokens and 175 types for analysis of markedness (see Table 42). From these remaining data, 177 loanword tokens were marked in some form, distributed across 111 types.

Table 42: Summary of frequencies of markedness in all loanword types

Rate of markedness	Type frequency	Percentage
100% - Always marked	46	26.3%
Sometimes marked	65	37.1%
0% - Never marked	64	36.6%
Total	175	100%

There was a relatively even distribution across types for markedness: 46 types (26.3%) were marked every time they occurred in the corpus, while just over 35% of all types were either sometimes marked or never marked.

Following Sharp (2007), it can be supposed that loanword types which are never marked in the corpus are well-established in NZE. A further inference can be made that, when a loanword is not marked, it is because the author assumes it is already well entrenched; that is, that it will be recognised and comprehended by

most speakers of NZE. Some high frequency types which were never marked in the corpus were: *kapa haka* (47 tokens), *kia ora* (33), *haka* (23), *hangi* (food cooked in an earth oven, 20), *kohanga* (19), *kaumatua* (an elder person of status within the *whānau*, 17) and *hapu* (sub-tribe, 16). Two proper nouns were also never marked: *Pakeha* (39) and *Kiwi* (37).

Conversely, it can be inferred that loanword types which are always marked constitute recent borrowings Māori into NZE, or borrowings that are not yet established (see Sharp, 2007). Comprehension is therefore not expected from a general readership by authors regarding these loanwords. The most frequently marked loanword in the corpus was *whakatauki* (a proverb), which had a 100% rate of markedness in all four tokens that occurred. Five types with token frequencies of two were also always marked: *kawa*, *pepe/pepi* (baby), *raranga* (weaving), *taha Maori* (Māori side), and *turangawaewae* (standing place⁵¹). The remaining 40 loanword types with a rate of 100% markedness only occurred once each in the corpus their low frequency, indicating that they may be nonce loanwords (Poplack et al., 1988).

It is clear that in most cases, loanword types were either always marked or never marked. However, approximately one third of loanword types in the corpus were variably marked, occurring with differing rates of markedness. To better analyse them, these types were placed on a continuum of markedness, by measuring the percentage of their tokens that were marked. These can be seen in Figure 36, where total loanword frequencies are also given in brackets.

In almost all cases, the variably marked loanwords with a very high overall frequency in the corpus (20 or more tokens) were only marked in less than five percent of their total occurrences. Exceptions to this were *kura* (with 30 tokens; 7% of which were marked), *manaakitanga* (21 tokens, 33% marked) and *taonga* (23 tokens; 35% marked). While *manaakitanga* is not considered to be a commonly known loanword, given that it only occurred in one year and was subject to thematic attraction (see Section 5.4.4.1.), *kura* and *taonga* were marked more than expected. On inspection, the seven percent of marked tokens in *kura* only amounted to two tokens, and were possible redundancies, as (22) and (23) show:

⁵¹ See Appendix A for an expanded definition.

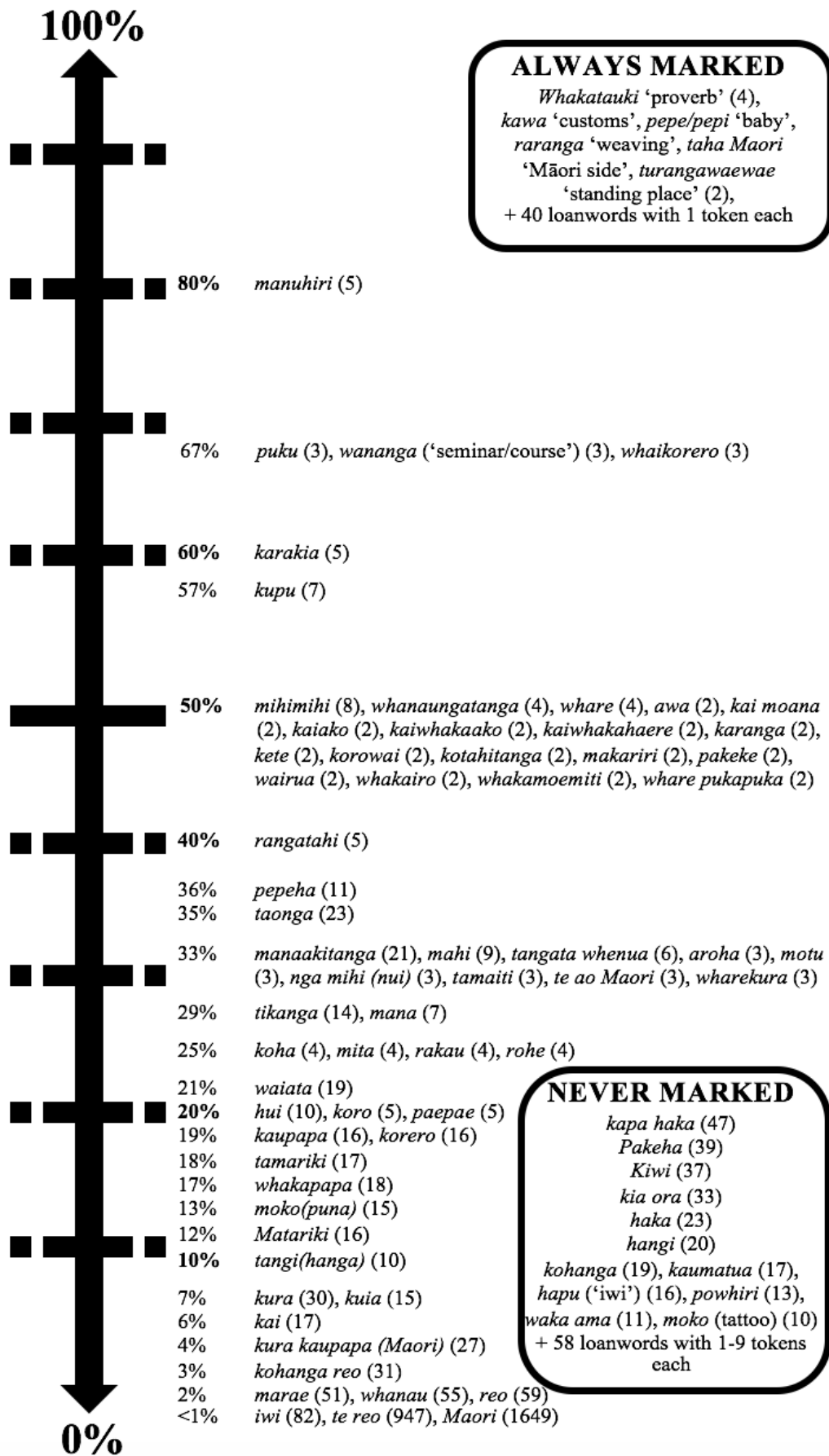


Figure 36: Percentages of marked and unmarked loanwords

(22) "...kohanga graduates who come through to the **kura school** and then to the performing arts" (TW_2014_22_JUL_HBT_001).

(23) "While New Zealand already has 93 **kura schools** and 500 odd Maori language preschools..." (MLW_2014_21_JUL_HBT_002).

The other explanation for these tokens is that *kura* has undergone a semantic shift, and refers to *kura* as a short form of *kura kaupapa Maori*. However, this is not the case in the majority of cases in the corpus, where *kura* indicates a school, and information regarding language of teaching is appended to this. The fact that the only marked examples of *kura* were redundancies indicates marking is only seen as necessary when a semantic change is occurring.

As well as being a high frequency and listed loanword, *taonga* occurred in eight out of 10 years in the corpus. It is (anecdotally) considered to be well-understood in NZE. Thus, its high rate of markedness was a surprise. *Manuhiri* (visitor, guest) is another example of a mid-frequency listed loanword which was marked in almost all occurrences (80% of five tokens). It may be that *taonga* and *manuhiri* are being translated by authors who perceive them as being unknown, when in fact it is already well entrenched in NZE. However, it is more likely that loanword entrenchment does not entail loanword comprehension.

The remaining loanwords on the continuum in Figure 36 would ordinarily be expected to be in various stages of incorporation into NZE. However, Poplack and Dion (2012) have shown that loanwords become immediately entrenched in a receptor language on their first use. In addition, markedness has been stated to be an indicator of entrenchment (Sharp, 2007). The less a loanword is marked, the more entrenched it is into the receptor language (Sharp, 2007, p. 235). In light of these statements, a difficult question arises: what is to be made of mid- to high frequency loanwords which are both listed and frequently marked?

This question is pertinent to the highly unanticipated occurrence of markedness in the six most frequent loanwords in the data (*Maori*, *te reo*, *iwi*, *reo*, *whanau*, *marae*); albeit in two percent of their tokens or less. Three of these loanwords were noted to be never marked in Davies and Maclagan (2006, p. 90), which was interpreted as them being accepted and established in NZE.

In the present study, two of these loanword types occurred in the same article: *reo* and *marae*. Nor were they the only marked loanwords in that article, as the excerpt below shows:

(24) ““We learnt so much of the **kawa [customs]** and reo [language] and meaning behind the words," she said. "I loved the culture of the **wananga**, you could make mistakes and laugh at it, you can self-correct and just have a go. I loved the **noho marae [marae stay]**, I got to say my **pepeha [tribal belonging speech]** and I was shaking””
(TW_2014_23_JUL_TDP_003).

Here, five loanwords occur in quick succession, four of which are marked in what appears to be an inconsistent fashion. Defining *reo* seems redundant given that it is such a high frequency listed type, and a variant of *te reo*, the second most salient Māori loanwords in NZE. In contrast, although *marae* would be expected to be comprehended by a general readership, in the variant phrase ‘*noho marae*’ it is appropriately marked as an unfamiliar loanword. However, *wananga* (university in this context) is an unlisted mid-frequency loanword and was not marked. While it is not made clear in the article, it seems likely that these translations were added post-interview by the author. Their prolific, and yet inconsistent, use of marking indicates that they are taking the opportunity to teach the meanings of Māori loanwords they perceive unknown to their audience. If education is the goal of markedness, then it is also possible that marking is a global textual choice rather than a choice made on an individual loanword level (see Calude et al., Forthcoming).

Instances where *te reo* was subject to marking in the data appear to be further cases where the authors have decided to teach, regardless of the need for a definition or not:

(25) “**te reo _ the Maori language**” (in MLW_2009_01_AUG_BPT_005 and MLW_2009_31_JUL_BPT_007; same author).

(26) “**te reo Maori (Maori language)**” (MLW_2008_11_JUL_HBT_003).

The singular instance of the type *Maori* being marked was when it referred semantically to the Māori language, and was one of the rarer instances of English

to Māori directional marking. Again, as is shown in (27), it appears to be marked in order to teach:

(27) “A free-to-use, free-to-share educational resource offering access to **Maori language (te reo)** content is seeking more fluent speakers of te reo Maori”
(TW_2010_02_AUG_TNA_001).

Finally, *iwi* was the last of the unanticipated, frequent-but-also-marked loanwords. However, on investigation, this was due to its occurrence in a list of commonly known loanwords as part of a street poll. Thus, when taken in context, its markedness actually indicates that it is perceived as being well-established in NZE.

These examples demonstrate that while measuring lack of markedness in Māori loanwords may sometimes be a good indication of the level of loanword entrenchment, neither is the presence of marking a completely reliable indicator of low levels of entrenchment, as the decision to mark a loanword is subject to personal perceptions of the author. Authors appear to be more likely to mark a loanword, even if it is listed and/or frequently occurring in NZE, if they wish to use it to serve an educative or teaching purpose. Markedness must therefore be considered carefully in conjunction with the factors of frequency and listedness, as well as author intent and perception, before making judgments about Māori loanword entrenchment in NZE.

6.3.1. Diachronic frequencies of all marked loanwords

Marked loanword tokens were measured as they occurred each year in the corpus between 2008-2017. These were compared with frequencies of all loanword tokens, and all words in the entire corpus. Diachronic frequencies of marked loanwords per 1,000 words were also calculated. These data are all combined in Table 43.

Although the frequency of marked loanwords was higher in 2008 than in 2017, there was also much fluctuation in between, making any diachronic change unclear, even when visually represented (see Figure 37). To this end, a Kendall Tau analysis was made.

Although there was no significant result in the raw token frequencies per year, when marked tokens were normalised per 1,000 words per year (refer back to Table 43), there was a significant result ($\tau=-0.511$, $p=0.0491$). According to these

findings, there was a significant, decreasing diachronic trend in the use of marked loanwords between 2008-2017.

Table 43: Summary of diachronic marked loanword data

Year	Marked loanword tokens	All loanword tokens	Entire corpus tokens	Marked loanwords per 1,000 words
2008	26	286	7,334	3.55
2009	5	276	6,900	0.72
2010	10	166	8,586	1.16
2011	12	202	5,750	2.09
2012	27	440	10,679	2.53
2013	17	366	11,409	1.49
2014	27	602	14,623	1.85
2015	15	448	12,498	1.20
2016	25	596	16,536	1.51
2017	13	413	14,610	0.89
Totals	177	3,795	108,925	1.62

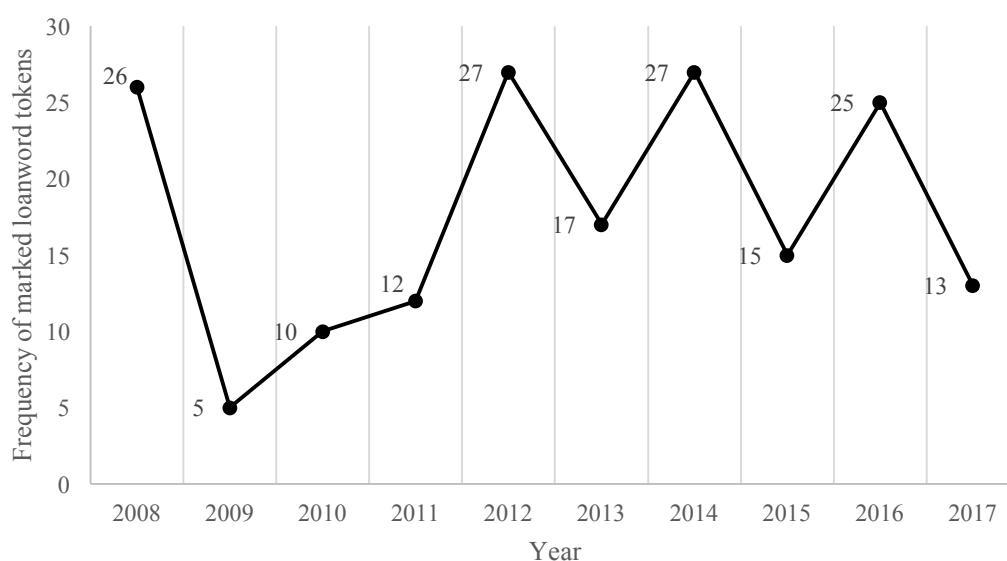


Figure 37: Marked loanword tokens by year

This decrease can be seen most clearly in Figure 38, which gives percentages of marked loanwords according to year. With the exceptions of the years 2012 and 2016, and a large anomalous decrease in 2009, markedness in loanwords has steadily decreased over the last 10 years, from around 9% in 2008, to just above 3% in 2017.

Conversely, there was also a significant increase in unmarked loanword use over the same period ($\tau = 0.511$, $p = 0.049098$).

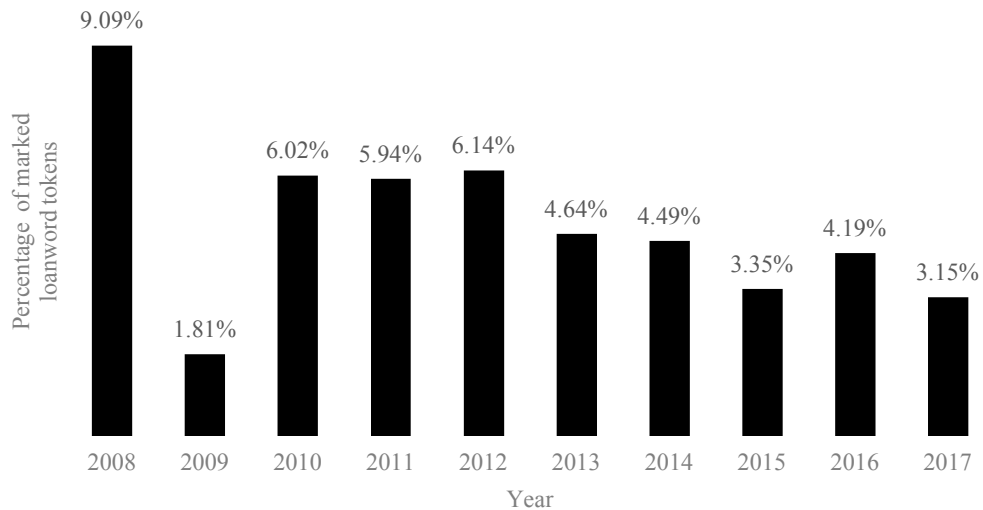


Figure 38: Percentage of marked loanwords by year

This trend away from marking Māori loanwords in NZE has been previously noted in Davies and Maclagan (2006, p. 96; termed ‘glossing’). They considered non-marked loanwords to be well established in NZE. In cases where some loanwords were marked in earlier years of their study but later dropped off, this indicated increasing “acceptance as a borrowing” and the fact that both newspaper authors and readers understand them (Davies & Maclagan, 2006, p. 80). It was noted that several frequent loanwords were never marked in their data⁵² (*hui*, *marae*, *whanau* and *iwi*), and inferred that they must therefore be well established in NZE (Davies & Maclagan, 2006, p. 90). This contradicts findings from the present study. All four of the aforementioned loanwords were marked to some extent in the data (refer back to Figure 36). Following Davies and Maclagan (2006), and other literature which holds that markedness indicates entrenchment (i.e., Sharp, 2007), the present findings would entail that these loanwords are not as established as first thought, or that they are becoming less entrenched. However, this is not the case, as evidenced by loanword frequency and listedness data.

There are two possible reasons for the decrease in marked loanwords over the last 10 years: 1) it could be that the marking of loanwords is becoming discouraged in journalism in New Zealand; or 2) fewer unknown loanwords and more known loanwords are being used by newspaper authors. To answer question 1), qualitative data from journalists are required, and these were not obtainable in

⁵² Data in Davies and Maclagan (2006) spanned 1997-2004.

the present research due to time constraints. In the case of question 2), it is necessary to consider other measurements of entrenchment, including loanword frequency, listedness, and morphological assimilation. These threads will all be woven together and examined in detail as one issue in the discussion section. However, before this is reached, there is one further aspect of entrenchment which has not yet been discussed. By its nature, markedness deals with implicit and covert perceptions of loanword comprehension. As such, the final issue to be dealt with is that of overt and explicit perceptions of knowledge of Māori loanwords in the corpus.

6.4. Author perception of loanword entrenchment

It has been suggested that after establishing the occurrence of a loanword in a corpus, interviews should be conducted in which speakers are asked about how widespread a loanword is, and when they use it, in order to gain a broader understanding of loanword entrenchment and use (Backus, 2014, p. 33). Although interviews of this nature were beyond the scope of this study, the articles themselves did yield some findings that allowed a similar, smaller scale investigation.

An interesting occurrence in the data is the small number of articles (10) in which the authors explicitly comment on the presence of Māori loanwords in NZE. In giving their knowledge and opinions on loanwords, these authors can be classified as “folk linguists” according to the definition given by Albury (2016, p. 292). Not only were the majority of the explicitly mentioned loanwords in the corpus not marked as belonging to a language other than English in the texts, but the authors argued that most New Zealanders would have no trouble recognising or understanding them. These data are interesting because of their anecdotal nature. Indeed, in certain cases, it would seem that the authors simply listed off the first few loanwords that came to mind. To gain a greater understanding of these loanwords, a comparison was made of how many times an author overtly described them as being part of NZE, with overall loanword frequencies in the corpus, and whether that loanword was listed in the *New Zealand Oxford Dictionary* (Deverson & Kennedy, 2005).

As Table 44 shows, a total of 50 loanword tokens and 41 types spread between 10 articles are mentioned by authors as being ‘well-known’ in NZE. These loanwords are ordered by frequency rank as they occurred in the corpus.

Table 44: Comparison of explicitly mentioned loanwords with corpus frequencies, listedness, and markedness

	Freq. rank	Loanword	Frequency mentioned explicitly	Corpus frequency	Listed or unlisted loanword	Marked or unmarked loanword	% Marked
High frequency loanwords	3	<i>iwi</i>	1	82	LISTED	MARKED	0.01
	5	<i>whanau</i>	2	55	LISTED	MARKED	0.02
	10	<i>kia ora</i>	1	33	LISTED	UNMARKED	0
	15	<i>haka</i>	1	23	LISTED	UNMARKED	0
	18	<i>hangi</i>	1	20	LISTED	UNMARKED	0
	20	<i>waiata</i>	1	19	LISTED	MARKED	0.21
	21	<i>whakapapa</i>	1	18	LISTED	MARKED	0.17
	22	<i>kai</i>	1	17	LISTED	MARKED	0.06
	24	<i>tamariki</i>	2	17	LISTED	MARKED	0.18
	29	<i>moko(puna)</i>	2	15	LISTED	MARKED	0.13
	31	<i>powhiri</i>	1	13	LISTED	UNMARKED	0
	34	<i>hui</i>	2	10	LISTED	MARKED	0.2
	36	<i>tangi</i>	1	10	LISTED	MARKED	0.1
Mid-frequency loanwords	41	<i>mana</i>	4	7	LISTED	MARKED	0.29
	46	<i>waka</i>	1	6	LISTED	UNMARKED	0
	48	<i>ka pai</i>	1	5	LISTED	UNMARKED	0
	55	<i>koha</i>	1	4	LISTED	MARKED	0.25
	63	<i>whare</i>	1	4	LISTED	MARKED	0.5
	66	<i>aroha</i>	1	3	LISTED	MARKED	0.33
	73	<i>wahine</i>	1	3	LISTED	UNMARKED	0
	80	<i>kai moana</i>	1	2	LISTED	MARKED	0.5
	101	<i>turangawaewae</i>	1	2	LISTED	MARKED	1
Low frequency loanwords	71	<i>puku</i>	3	3	LISTED	MARKED	0.67
	117	<i>e noho</i>	1	1	UNLISTED	MARKED	1
	118	<i>e tu</i>	1	1	UNLISTED	MARKED	1
	120	<i>hapu (pregnant)</i>	1	1	UNLISTED	MARKED	1
	123	<i>iwa</i>	1	1	UNLISTED	UNMARKED*	0
	129	<i>kauri</i>	1	1	LISTED	UNMARKED	0
	142	<i>morena</i>	1	1	UNLISTED	UNMARKED	0
	143	<i>ono</i>	1	1	UNLISTED	UNMARKED*	0
	152	<i>rima</i>	1	1	UNLISTED	UNMARKED*	0
	155	<i>rua</i>	1	1	UNLISTED	UNMARKED*	0
	158	<i>tahi</i>	1	1	UNLISTED	UNMARKED*	0
	162	<i>tapu</i>	1	1	LISTED	UNMARKED	0
	163	<i>taringa</i>	1	1	UNLISTED	MARKED	1
	166	<i>tekau</i>	1	1	UNLISTED	UNMARKED*	0
	173	<i>toru</i>	1	1	UNLISTED	UNMARKED*	0
	176	<i>utu</i>	1	1	LISTED	MARKED	1
	178	<i>waru</i>	1	1	UNLISTED	UNMARKED*	0
181	<i>wha</i>	1	1	UNLISTED	UNMARKED*	0	
186	<i>whitu</i>	1	1	UNLISTED	UNMARKED*	0	
		Total	50	389			

*Indicates a loanword which was only implicitly marked. Considered as unmarked.

What is immediately apparent is that almost half of all types (19) cited by authors were the only times they occurred in the data. These loanwords fall into the bottom band shown in Table 44, where they are misidentified by newspaper authors as high frequency loanwords. The majority of types in this band are unlisted, with only four being listed. Similarly, when unmarked tokens with an asterisk are removed (see note below table), the majority of bottom band loanwords are marked. In all cases of this category except one, markedness occurs at a rate of 100%. The exception to this is *puku*, of which all three of its tokens are mentioned explicitly in the data, and two of them are marked, at a rate of 67%. From this data, it appears that there is a correlation between low frequency loanwords in the corpus and unlistedness; however, markedness is more variable and seemingly indiscriminate.

In Section 5.1.1, high frequency loanwords are given as having a frequency of 10 or greater. Of the 36 high frequency types in the corpus, only 13 are correctly identified by authors. These can be seen in the top band of the table. Interestingly, only four high frequency types were identified as being such more than once: *whanau*, *tamariki*, *moko(puna)* and *hui*. Additionally, *iwi*, the third most frequent loanword in the corpus after *Maori* and *te reo*, was only correctly identified as being extremely salient in NZE by one author. In total, only one third (34%) of all tokens quoted by authors as being well known in NZE occur in this high frequency band. However, when a comparison is made with listedness, it is strikingly clear that all of these high frequency loanword types are listed. From this it can be inferred that these loanwords are in fact correctly identified as being highly frequent in NZE, as the authors claim. It is interesting, then, that the majority of these loanwords are marked. Given their high frequency and listedness, one would expect these to be rarely marked, if at all. Rates of markedness are generally low in this band, but do hit the 20% mark in some cases (*waiata*, *hui*).

This pattern of high rates of both listedness and markedness is repeated in data from the mid-frequency loanword band. Once again, the loanwords in this band are all listed, and the majority of them are marked. Only three loanwords here are unmarked: *waka* (6), *ka pai* (5) and *wahine* (3).

Of the 41 loanword types mentioned explicitly by authors, listed types were overwhelmingly more likely to be marked than unmarked; while unlisted and listed types were almost even in their distribution between markedness (see Table 45 below).

Table 45: Comparison of marked and listed types explicitly mentioned by authors

	Marked	Unmarked	Totals
Listed	17	9	26
Unlisted	4	11	15
	21	20	41

This reiterates findings established in previous sections that markedness is not functioning as an indicator of entrenchment of Māori loanwords as they pertain to NZE. If this were the case, listed loanwords would be less likely to be marked than unmarked, if ever. Further investigation is needed into whether this correlation exists outside of the slice of data considered here; however, from this, it appears that markedness in loanwords is doing different things in NZE than in other languages.

6.4.1. Author perception of loanword entrenchment: Diachronic effect

Another question regarding the loanwords explicitly mentioned in the data as being well and widely known, was whether there was any diachronic effect to this. Table 46 gives a list of all such tokens sorted according to their year of occurrence.

There is little to comment on this, beyond the fact that the year a given loanword was explicitly described as belonging to NZE seemed extremely variable. The highest number of loanwords were explicitly quoted in 2009, but so too did the highest number of articles occur then. 2008 had the highest number of loanwords drawn attention to in a single article, 19. However, it should also be noted that the majority of this article was in fact a list of words (including numerals) used in a street survey to test the knowledge of Māori loanwords in locals⁵³. Thus it is not a representative or reliable figure. Nor can it be said that the number of loanwords commented on is decreasing or increasing. In fact, the most commonly cited example of a well-known borrowing from Māori, *mana* occurred in 2008 and 2009, and then twice in 2014. The only common factor here is genre of article. Nine articles out of 10 could be described as opinion pieces, be they in the form of a letter to the editor (initial or rebuttal of a previous letter), or a journalist's opinion piece

⁵³ This article also included a survey of opinions on whether the Māori language should be compulsory in school, the responses to which were frustratingly not included in the article!

Table 46: Explicitly mentioned loanwords by year

Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Total
No. of tokens	19	15		5			4	4	3		50
No. of articles	1	4		1			2	1	1		10
Loanwords explicitly mentioned more than once		<i>hui</i>		<i>hui</i>							
	<i>mana</i>	<i>mana</i>					<i>mana</i> x2				
		<i>moko(puna)</i>		<i>moko(puna)</i>							
	<i>puku</i>			<i>puku</i>				<i>puku</i>			
		<i>tamariki</i>		<i>tamariki</i>							
		<i>whanau</i>							<i>whanau</i>		
Loanwords mentioned explicitly once only	<i>waiata</i> <i>turangawaewae</i> <i>whakapapa</i> <i>kai moana</i> <i>iwi</i> <i>whare</i> <i>hapu</i> (pregnant) numerals 1-10	<i>ka pai</i> <i>aroha</i> <i>kia ora</i> <i>waka</i> <i>haka</i> <i>hangi</i> <i>morena</i> <i>powhiri</i> <i>koha</i> <i>tangi</i>		<i>kai</i>			<i>utu</i> <i>tapu</i>	<i>taringa</i> <i>e tu</i> <i>e noho</i>	<i>kauri</i> <i>wahine</i>		

on the topic of the revitalisation of the Māori language. More specifically, the opinions given on the Māori language within the articles were universally positive. Thus, it can be stated that in cases where a positive opinion is being voiced about the Māori language in a news article, authors may choose to use their own perceptions to illustrate what they believe are frequent and well-entrenched Māori loanwords in NZE. Universally agreed upon known loanwords in the corpus include *puku*, *mana*, *moko(puna)*, *whanau*, *hui* and *tamariki*. However, this list is purely subjective. Genre appears to be functioning (logically) as an attractor for explicitly mentioned loanwords. In opinion pieces on the subject of the Māori language, where the sentiment is positive, it is more likely that this will occur; however, how many and which loanwords will be commented on is unpredictable.

6.4.2. Continuum of perceived loanword entrenchment

It has been shown that author perceptions of which loanwords are frequent and entrenched into NZE are extremely variable. However, what has not yet been discussed is exactly what those opinions are, as to as to how established a given loanword is. These can be seen by placing author quotes regarding the loanwords they mention on a continuum of perceived entrenchment, as in Figure 39.

There is a clear awareness in the corpus that not all borrowings from Māori are equally entrenched: some may be “regulars” (iii) or used by “the majority of the population” (v), while others “have migrated into New Zealand English to the point where no explanation is required” (ix). In the case of example (x), the claim is made that “all New Zealander[s] use” certain words, such as *mana* and *tapu*. While this is not the case in this corpus, both loanwords do occur in the top 50 most frequently occurring loanwords in spoken NZE⁵⁴ (Kennedy, 2001, pp. 70-71). Similarly, in a study of three different newspaper corpora by Degani (2010), *mana* was found to occur at minimum 10.48 and maximum 36.5 words per million words (p. 179).

New Zealanders are aware that not all words borrowed from Māori are embedded in NZE to the same extent, but in the few examples given in the corpus, there is not always agreement about this. Two of the articles from Figure 39 explicitly mention markedness in Māori loanwords. One (ix) mentions brackets as a measure of how well a loanword has become part of NZE, while the other (iv)

⁵⁴ But not in written NZE (Kennedy, 2001, p. 71).

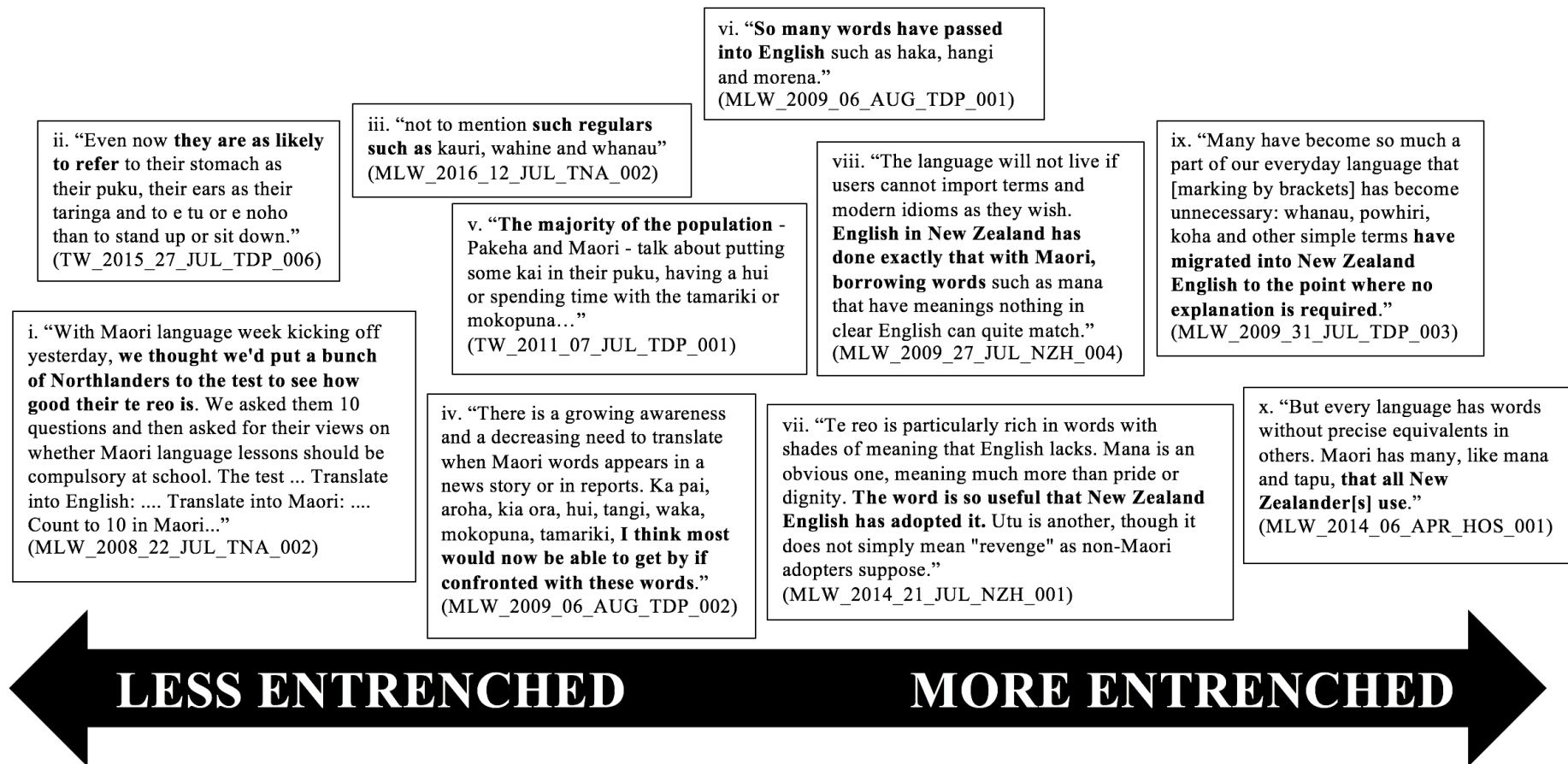


Figure 39: Continuum of anecdotes of author perceptions of loanword entrenchment

claims that “there is a decreasing need to translate when Maori words appears in a news story or in reports”. And yet despite commenting on the same issue, the loanword examples they provide have no overlap. A third, not included in the continuum as it does not name specific loanwords, is as follows:

(28) “The Rotorua Daily Post regularly features articles that include Maori words and phrases... while I think I can sometimes glean the meaning from context, that is both unreliable and potentially inaccurate or lacking the subtlety of distinction... Could columnists translate all Maori language references each time they are used? They introduce such beautiful imagery into each article, which can foster respect and interest in the language, and certainly shine as examples of ‘multi-cultural inclusion’. Even though many words or phrases become commonly understood over time, there is always the potential for new readers to be seeing them for the first time”
(TW_2017_14_SEPT_TDP_006).

While these data are anecdotal by nature, the hit-or-miss identification of common loanwords in NZE by the authors in the corpus indicates that personal perceptions of what constitutes an entrenched loanword are extremely variable and just as often inaccurate as accurate. Moreover, it has also been demonstrated that a corpus may not necessarily reflect actual loanword use. This is yet another indication that the loanwords harvested in this study are clustering around a certain topic.

Perception of known loanwords is an area that has been overlooked in Māori loanword research heretofore, and this small slice of data indicates that there is much to be found in it. It has been shown that commentaries of perceptions of loanword frequencies are strongly linked to article genre, and suggests that further research in the area of folk linguistic Māori loanword usage, specifically in opinion pieces in the media, could be highly fruitful.

7. Discussion

Entrenchment of Māori loanwords in NZE is an issue with a history as long as that of contact between the two languages. This research has attempted to situate Māori loanword entrenchment through a wide-spanning, in-depth and quantitative exploration of three interrelated concerns as viewed through the lens of a topically constrained newspaper corpus: loanword frequency of use, listedness, and markedness. Morphological assimilation, although only touched upon in the form of pluralisation, has also offered a fourth element of analysis. Perceptions of Māori loanword use have also been considered: both implicit perceptions, through markedness, and explicit perception, through directly expressed quotes in the corpus. Though by no means exhaustive, in centring the corpus upon the topic of Māori Language Week over a ten-year period, this research affords a diachronic snapshot of attitudes towards Māori loanwords in NZE, and by extension, the Māori language in New Zealand. Historical measurements of loanword entrenchment are also considered and a ‘one method fits all’ approach in loanword studies is argued against.

7.1. Māori loanword use as an expression of political attitude

For the most part, attitudes to Māori loanwords and the Māori language expressed explicitly in the corpus were strongly positive. In fact, there were very few articles in which negative sentiment was expressed, and when it was, this was usually done so cautiously. This is no surprise, as the attitude to Māori loanwords and the Māori language is linked to attitudes towards people of Māori ethnicity and the Māori culture. Māori loanword use is a documented marker of identity (Davies & Maclagan, 2006, p. 97). It has been suggested that “the use of loanwords carries the implication of solidarity towards Māori language and by implication, towards Māori people and culture” (Calude et al., Forthcoming, p. 16). This can be further applied to views on Māori cultural and political issues. An attitude to Māori loanword use is extrapolated into attitude towards the revitalisation of the Māori language, opinion on it being compulsory in schools, the rights of Māori, the Treaty of Waitangi, *tino rangatiratanga* (Māori sovereignty) and so on. If Māori language rights are recognised, then other, non-linguistic rights must also be recognised. It seems contradictory, then, that Māori loanword use in NZE is overtly perceived as

prestigious, while the revitalisation of the Māori language is a polarising issue: for example, whether or not it should be compulsory in schools⁵⁵. This is because attitude towards the Māori language cannot be extricated from any of its associated issues and considered on its own. This also applies to Māori loanwords. To use a Māori loanword in NZE (at least, one that is saliently of Māori origin; not, for instance, *Kiwi*) is to step onto the stage of political opinion.

Albury (2016, p. 299) suggests that the revitalisation of the Māori language is often reduced to a choice between culture and economy. If this is true, then it follows that Māori loanwords become a question of politics. As a result, Māori loanword use in NZE, which is an extension of language revitalisation, although not always overtly politicised, is a highly political issue. Several of the articles in the corpus were concerned with low voting rates in the Māori population, and Māori Language Week (and consequently the Māori language) was seen as an opportunity to address this. This can be seen in (29) and (30) below:

(29) “The Electoral Enrolment Centre hopes to use Maori Language Week to encourage almost 47,000 Maori to update their details and sign back on to the electoral roll before the coming general election” (MLW_2008_23_JUL_HBT_004).

(30) “Mr Te Kani said there was apathy among Maori choosing not to vote and Maori were the hardest voters to mobilise.

‘If you are eligible, enrol and exercise your democratic right. Don't let your voice become marginalised.’

Mr Te Kani said it was vitally important for Maori to get information in both Maori and English.

‘To have a bilingual approach in a general election is a very positive step to ensure Maori will vote in this election’”(MLW_2011_07_JUL_TDP_004).

The use of loanwords, whether conscious or not, is just as loaded in the political sphere. A variety of political opinions are on display in one article, in which seven MPs from the Rotorua region are interviewed regarding their stance on making the Māori language compulsory in schools. Six of the seven MPs are of Māori descent, and all declare a positive opinion of the Māori language, regardless of their political stance.

⁵⁵ For evidence of this, one need only search the phrase “compulsory Māori language” online to be assailed by a wide spectrum of vocal opinions.

Interestingly, the politician with the highest use of loanwords is Fletcher Tabuteau (New Zealand First), an MP of Māori descent. Although Mr Tabuteau expresses strong support for the Māori language, he is also extremely non-committal in his response and does not give a yes or no answer in regards to his stance on making it compulsory in schools. His full response is given below⁵⁶.

(31) “Fletcher Tabuteau – New Zealand First candidate for Rotorua.

I strongly assert that **te reo Maori** is a **taonga** of Aotearoa and it is for all of us to be agents for its continuance and growth. In all of our mainstream schools here in Rotorua teachers are always trying to integrate **Maori** culture and **te reo** into everyday learning in all aspects of the curriculum. NZ First would welcome a review around the **kaiako** resource support for the teaching and learning of **te reo** in in the mainstream schools, but also ensuring quality learning for all students in full immersion **kura**”

(TW_2014_26_JUL_TDP_011).

The above excerpt has eight loanword tokens and five types, as well as three proper noun place names, in just 100 words of text. In fact, 8% of the text is loanwords. Neither does Mr Tabuteau mark any of his loanwords, a move which could be interpreted as an appeal to his audience that he fully comprehends every loanword he has used and so can be trusted on such a matter. In comparison to the other politicians in the same article, it appears that his high use of loanwords is overcompensating for his lack of a firm answer, to emphasise his positive and supportive stance towards the Māori language, when otherwise it might be called into question. Regardless of his undisclosed (covert) political stance, Mr Tabuteau utilises a large number of loanwords to express his overt support for and positive attitude towards the Māori language, and to lessen his likelihood of being perceived as being negative and disrespectful.

Another article in which higher loanword frequency expresses a potentially thorny opinion can be seen in the below letter to the editor, in which efforts to revive the Māori language are described as a waste of money:

(32) “Setting aside confusion about the actual cost, the estimated multi-millions allocated for **te reo Maori** over many years have delivered next to no return for **Maori** or the

⁵⁶ Māori loanwords included in the corpus frequency counts are in bold, while proper nouns are underlined.

taxpayer. The number of people speaking fluent **Maori** has not increased, but rather has dramatically and inexorably declined since the inception of the Maori Language Act and Maori Language Commission in 1987. Minority languages worldwide are declining and are forecast to disappear. Even Rome was unable to save its language, Latin. To continue taxpayer subsidisation of **te reo Maori** is counter-productive and a waste of desperately needed capital. The Government should withdraw funding, and the money saved could be invested in more productive outcomes to help raise the standard of living for everyone, including **Maori**” (MLW_2011_08_JUL_NZH_003B).

Although the rate of loanword use here is not quite as high as that of the previous excerpt, it is close, at eight loanwords and two proper nouns from 123 words, including a title loanword (“Teaching Maori”). The difference is that this writer expresses a negative view towards the revitalisation of the Māori language. Interestingly, the writer uses ‘*te reo Maori*’ twice, when either *Maori* or *te reo* alone would have been sufficient. They appear to use a higher number of loanwords as an appeal to fact; however, their loanword types are restricted to two. This may be due to the possibility that a person with a negative opinion of the Māori language is more likely to have a narrower loanword vocabulary than someone with a positive opinion, yet may still use a high number of loanwords as a means of attitudinal expression or to show knowledge on the subject. Such a possibility warrants further investigation.

A final example of Māori loanwords being used, in this case misleadingly, to further a political agenda is given in Mutu (1994), who describes an event at a Board of Inquiry into the New Zealand Coastal Policy Statement where “a rather exceptional few local authorities and interest groups...quite openly used the confusion to take the [Māori] terms and redefine them for their own political purposes, thus adding to the confusion” (p. 4). It is clear from this that use of loanwords is no longer, if it ever was, restricted to only those who identify with the Māori language or even respect it. Nor is high frequency loanword use.

Although this is somewhat unexpected, it is not unprecedented. In a study on codeswitching in the Jersey Islands, it was found that “although frequency of use is probably not without significance here, a speaker’s attitude towards Jèrriais seems to be the factor which has the most bearing on the extent to which they code-switch” (Jones, 2005, p. 18). It appears that this is also the case with Māori

loanwords in NZE. Frequency certainly cannot be discounted; however, it is attitude which is of most importance to loanword use.

This correlation can be explained by Calude et al.'s (Forthcoming) "global textual choice", in which loanword use is not done on an individual lexical level, but on a global level "to use a group of loanwords or none at all" (p. 16). Loanword use is a global textual choice precisely because it expresses attitudes to Māori language and political issues. Those who wish to demonstrate support for the Māori language can be expected to use more loanwords, and more types, as a vehicle for doing so. However, so may those who wish to be seen favourably despite their neutral or negative attitude to the Māori language. Either way, a global choice to express a given attitude to Māori issues has been made.

In discussing the role of Māori loanwords as differentiating it from other varieties of English, NZE has been described as "the language of a politically dominant majority made distinctive by the incorporation of words from the language of the indigenous minority" (Metge, 2009, p. 2). If this is the case, then the use of loanwords by the political minority becomes in fact a method of asserting political volition. This is not a new concept. As Zuckermann (2011) writes, "language revival does not only do historical justice and address inequality but can also result in the empowerment of people who have lost their heritage and purpose in life" (p. 113). Using Māori loanwords in NZE, then, is one symbolic means of redressing this inequality. However, it seems to be first and foremost a means of political expression, of which Māori language revitalisation is a substantial but not unique concern.

It must be qualified that an author or speaker doesn't necessarily use any or every loanword, for example *kai* (food), to make a political statement. It may also be variously employed as a marker of ethnic or New Zealand identity; both intentionally or unconsciously. Nonetheless, and regardless of intention, doing so is equated with a political positioning.

If this is true, then we must return to the initial research question of why Māori loanword frequencies in NZE matter. It has already been established that loanword use is significantly higher than it has been in the past. But why should we care? The answer is that this increase has extremely important ramifications for all New Zealanders. Following the assumptions that:

- a) People who use Māori loanwords support (or wish to be seen as supporters of) the Māori language
- b) People who support the Māori language support (or wish to be seen as supporters of) the Māori people and culture⁵⁷
- c) People who support the Māori people support (or wish to be seen as supporters of) X Māori political issue

Then we must conclude that the increase in use of Māori loanwords reflects shifting overt attitudes to Māori language revitalisation, Māori political issues, and cultural relations in New Zealand. This assessment offers a more hopeful interpretation of Harlow's (2005) claim that the use of Māori loanwords in NZE, "far from contributing in any way to the status of Māori as a language [only] serves to reinforce the monolingually-based status quo" (p. 140). Rather, loanwords are a vital means of expressing a wide range of attitudes to some of the most salient socio-cultural and political issues in New Zealand today. This is not to the exclusion of loanwords as markers of ethnicity and identity; nor does it entirely preclude the overt positive, but covert negative attitudes, or tokenism mentioned in Harlow (2005). However, on the whole in the corpus, increased diachronic loanword use is overwhelmingly reflected in an overt willingness to preserve the Māori language; and to identify with and express solidarity with a range of Māori cultural and political issues.

Given this conclusion, it is regrettable that attitude to the Māori language was not coded for in this study. More investigation is needed in order to confirm whether there is any statistically significant correlation between attitude and loanword frequency. Further research into corpora focussed on the topic of the Māori language in general may well prove fruitful in resolving this issue; and a possible means of doing so would be to focus on opinion articles in newspapers to compare loanword use with articulated and implicit attitudes, both positive and negative. It could be that the topic of Māori Language Week served as an attractor of chiefly positive sentiments, and that lower frequencies of loanwords will be found where attitudes are not as strongly positive. However, for now, it is an open question. The fact remains that overtly positive attitudes to the Māori language are

⁵⁷ See Calude et al. (Forthcoming) for assumptions a) and b)

increasing, at least in newspaper media on the topic of Māori Language Week, and increased frequencies of Māori loanwords therein are both a means of expressing such attitudes, as well as measuring them.

7.2. Measuring entrenchment of loanwords from Māori as a threatened language

This research has considered conventional methods of measuring loanword entrenchment; however, they were only found to be effective to a certain point. Where loanword use is above the level of consciousness, as Māori loanwords are in NZE, and where they are being used (intentionally or otherwise) as a marker of attitude and to express political ideologies, indicators such as frequency, morphological assimilation, listedness and markedness are not always reliable.

It has just been established in this work that loanword entrenchment is more reliant on attitude than on frequency. Although level of entrenchment is usually predicated on loanword frequency (Backus, 2014), when loanwords are utilised as a positioning marker of attitude to a language, culture, people or other political issues, as is the case with Māori and NZE, frequencies of a loanword are less important than the usefulness of a particular loanword to express the attitude of the user.

In the same way, morphological assimilation has generally been considered to be a good indication of the entrenchment of a loanword. Turpin (1998) writes that “the more a loanword is diffused (and accepted) in the linguistic community, the more it is integrated into the recipient language” (p. 231). English loanwords in Welsh are more likely to be adapted to Welsh grammar the more they are frequent (Stammers & Deuchar, 2012). Adaptation to a receptor language’s grammar is also usually immediate and abrupt (Poplack & Dion, 2012). One reason for this is that users seek to minimise the salience of a loanword by fully assimilating it into the receptor language morphology (Poplack & Dion, 2012, p. 309).

However, the opposite is true of Māori loanwords in NZE. Pluralisation is one of the only means of measuring such morphological assimilation, and it has already been noted that in recent years, active steps have been taken to cease this (Davies & Maclagan, 2006). Regardless of whether it is linguistically appropriate, the prevailing attitude is now that “not imposing the grammatical rules of the dominant language in our culture reveals a growing acceptance of te reo Māori, and

reflects the social and political status of Māori in general” (Davies & Maclagan, 2006, p. 89). Enforcing English plurals on Māori loanwords is seen as disrespectful and just another act of colonisation, while not using English pluralisation is a marker of respect. This is evidenced in the complete lack of English pluralisation in non-reported speech loanwords in this corpus⁵⁸. In contrast to other languages, newspaper media in New Zealand universally seek to make Māori loanwords as salient as possible in their texts.

The singular exception to this is *Kiwi*, which always has an ‘s’ suffix when occurring in the plural. This loanword is both high frequency, listed, and totally conformant to English pluralisation rules; as such, it is the only loanword in the corpus interpretable according to the previously cited literature. Based on this evidence alone, however, it would be ridiculous to claim that *Kiwi* is the only loanword that is entrenched in NZE. What it does reveal is that there has been a social movement to linguistically ‘dis-entrench’ Māori loanwords, and that while the majority of loanwords are collected by this, there are some which were so well-established prior to this shift that they are no longer considered loanwords, and so are resistant to dis-entrenching. It is speculated that loanwords resistant to the shift to dis-entrench would include predominantly flora and fauna loanwords and material culture loanwords likely to have entered NZE in Macalister’s (2006b) first wave of borrowing, prior to the 1970s.

In the same way that some “phonological criteria are not reliable indicators of loanword integration” (Poplack & Dion, 2012, p. 284), the same may be said of morphological assimilation in cases where the act itself is above the level of awareness and has attendant socio-political and cultural motives. In fact, morphological assimilation as a hypercorrection is just one indicator of a ‘change from above’. Changes from above “take place at a relatively high level of social consciousness, show a higher rate of occurrence in formal styles, are often subject to hypercorrection, and sometimes show overt stereotypes” (Labov, 1990, p. 213). Although stereotyping in loanword use was not explored here, all other conditions of such a change are met in the data. This offers strong evidence that increased Māori loanword use in NZE, at least in the more formal text of newspapers, is undergoing a change from above. It also suggests that morphological conformity is

⁵⁸ Excluding *Kiwi*, and, of course, reported speech, as discussed in Section 5.1.4.

only an indicator of loanword entrenchment insofar as a loanword exists below the level of social consciousness. In light of these findings, Māori loanword entrenchment cannot be measured according to standard morphological conventions, because of the conscious social movement to morphologically dis-entrench them. As this movement is shaped by overtly held attitudes to the Māori language, studies of the morphological entrenchment of loanwords must therefore take into consideration social factors, attitudes to and the status of the source language, and cannot solely rely on morphological factors.

The next measurement of entrenchment to be discussed is that of listedness. This research has shown that there was a significant, diachronic increase in the frequencies of listed loanwords between 2008-2017. In contrast, there was no such significant result in unlisted loanwords.

While attestation offers a good starting point for determining loanword entrenchment, it does not consider that in a topic-bound corpus the loanwords themselves are constrained to certain topics, as demonstrated in Sharp (2007). The majority of borrowings in previous studied corpora are usually attested (Barffour, 2016, p. 85; Poplack et al., 1988, p. 59; Poplack & Dion, 2012, p. 286; Stammers & Deuchar, 2012, p. 637). Moreover, in newspaper media, it is standard for more conservative language to occur (Davies & Maclagan, 2006). It can therefore be assumed that there will always be a high proportion of listed loanwords in newspapers. This is reflected in the data, with over 94% of loanword tokens being listed.

Another issue in relying on attestation for entrenchment levels is that attested loanwords will undoubtedly vary between dictionaries. Listed loanwords in this study were taken from the most recent version of the *New Zealand Oxford Dictionary*, a dictionary which was published in 2005. It is likely that a more recent, revised edition would have an increased number of listed loanwords⁵⁹, and this would only increase the number of listed results in this study. Neither is any question asked in regards to which loanwords were chosen to be listed in said dictionary. In the case of NZE, it may be that certain Māori words were selected not on the basis of frequency or on common comprehension, but on the grounds

⁵⁹ Although it was hoped that a figure might be available for the number of Māori loanwords in the *New Zealand Oxford Dictionary*, after chasing many leads, a dead end was reached. At this juncture, counting the loanwords manually appears to be the only method to gaining such information.

that they are culturally important concepts which are integral to the understanding of Māori, and therefore to Māori-Pākehā cultural relations, rather than on linguistic grounds. It is likely that listed loanwords themselves are highly subject to author perception of such issues. An example of this is the loan *turangawaewae* (standing place), a listed word which occurred twice in the data, but was both times marked by a definition.

Despite this, listedness still has benefits to offer in the research area of entrenchment. High frequencies in listed loanwords can certainly suggest which loanword types are considered to be well-established; while growing frequencies in unlisted loanwords may reveal potentially novel or recent borrowings, as well as increased frequency of use therein, such as in the loanwords *Matariki* (Māori New Year), *manaakitanga* (hospitality), *waka ama* (outrigger canoe sport) and *pepeha* (tribal saying or motto⁶⁰) in the present study. However, the high number of listed loanwords with extremely low frequencies in this study also shows that listedness cannot be considered alone when assessing entrenchment. It is rather a complementary measurement that can provide insight when taken in conjunction with other traditional measures of entrenchment; one that is once again subject to perceptions and attitudes to the Māori language which were present in the construction of the dictionary.

The final measurement of loanword entrenchment used in the present study is markedness. As it has been shown that a low number of flagged (marked) loanwords is not unusual in a topic-bound corpus because of assumed knowledge in topic-specific discourse (Sharp, 2007), the low frequency of marked loanwords in this study (<5%) could be seen as unremarkable. Nonetheless, there was a significant diachronic decrease in marked loanwords between 2008-2017, which ran in tandem to a significant increase in overall loanword use. The most immediate inference one might make following this is that, on the surface, Māori loanwords are becoming more entrenched in NZE. However, within this decrease in markedness, it is clear that loanword marking is being utilised in two distinctly different ways.

The first reason for marking loanwords in the corpus appears to be conventional. Loanwords are marked when authors perceive a lack of sufficient

⁶⁰ See Moorfield (2018) for a fuller definition.

information in their audience. The two newspapers with the lowest proportional resident Māori populations were the *NZH* (a nationwide publication) and the *BPT* (a regional publication from the city of Tauranga). These newspapers also had the lowest frequencies of loanwords in general, as well as the lowest percentages of marked loanwords. In these newspapers, comprehension appears to be the goal. Fewer Māori loanwords are used than in other newspapers (by both type and token), and types which are considered to be more easily comprehensible are preferred. This results in a low percentage of marked loanwords, and those which are marked could be considered less well entrenched than co-occurring unmarked loanwords. This supposition is interpretable according to previous literature that consider markedness to be “indicative of the unintegrated status of the...items used” (Sharp, 2007, p. 235); to be inversely correlated with loanword entrenchment (Grant-Russell & Beaudet, 1999; Sharp, 2007); and that lack of marking indicates strong entrenchment (Davies & Maclagan, 2006; Degani, 2010).

However, not all newspapers utilise markedness to the same end. Barffour (2016) states that “flagging a word or not in journalistic writing may be motivated by other subjective reasons” (p. 15). The three newspapers with the highest proportion of Māori residents in their readership predictably had the highest frequencies of Māori loanword use. Yet, despite the fact that it can be assumed that readers in these regions will be more familiar with loanwords, these same newspapers counterintuitively also have the highest percentages of marked loanwords in the corpus. The *HBT* (from the Hawke’s Bay) and *TNA* (from Northland) have the highest percentages of markedness in their loanwords: 6.95% and 5.02% respectively. There is no reason to think that less loanwords are known in these regions than in regions with lower Māori demographics. Thus, it can be inferred that they are using markedness to teach loanwords and their meanings in a move supportive of Māori language revitalisation. Moreover, the newspaper *TDP* (from Rotorua) takes this a step further: it is the only newspaper to have more marked loanword types than unmarked types. Located in the first bilingual city in New Zealand, *TDP* appears to be not only teaching the meanings of loanwords they perceive to be unknown, but actively and intentionally teaching new Māori vocabulary. Loanwords are thus likely being deliberately used to help regenerate the Māori language in Rotorua.

This has important implications for how loanword entrenchment is measured in all languages. If markedness is not only a method for clarifying the unknown, but also serves an educative function, and by extension, plays a potentially indirect role in language revitalisation, then markedness can no longer be taken as a straightforward rule of entrenchment. Furthermore, markedness itself is dependent on author perception of what might or might not be understood by their audience. It has been noted that “typographic flagging can be purely arbitrary” (Grant-Russell & Beaudet, 1999, p. 26). This research has gone some way to distinguishing motivations behind marking loanwords, but much remains to be discovered. Therefore, before markedness can even be considered as a reliable gauge of how established a loanword is in a receptor language, function and author-intended purpose of any given marking must be established. How this may be done in practice will be difficult, if not impossible, but it cannot be ignored.

7.3. Māori loanwords in NZE cannot be evaluated by the same criteria as loanwords originating in non-threatened languages

The Māori language is in a substratal relationship with NZE; that is, it “has a position of lower social and cultural prestige vis-à-vis another” (Durkin, 2014, p. 13)⁶¹. While Māori is a substrate language in this context, NZE is a superstrate language with higher prestige (Durkin, 2014). It was unexpectedly difficult to find research which considered borrowings donated from or received into substrate languages; the majority seem only to consider the superstratal languages in loanwords studies.

One of the only examples of this was found in Stammers and Deuchar (2012). Unfortunately, their results are not directly comparable with those of the present study for a number of reasons⁶². However, their findings do offer an interesting possibility for the behaviour of Māori loanwords in NZE, as they are the only study to challenge the Nonce Borrowing Hypothesis, which they define as follows: “there is no difference between frequent and infrequent donor-language

⁶¹ It should be noted that in some social situations in New Zealand, the reverse is true and the Māori language has higher prestige (it is a superstrate).

⁶² Following Stammers and Deuchar (2012), the more frequent a Māori loanword is, the more conformant it should be to the morphology and grammar of NZE, and unlisted Māori loanwords should be less well entrenched in NZE. However, they consider the mutation of English verb loanwords borrowed into Welsh. Māori verb loanwords are extremely rare in NZE, and their morphological conformance in NZE is not generally measurable even in noun loanwords.

items in terms of their degree of integration” (Stammers & Deuchar, 2012, p. 632)⁶³. I propose that this difference is not a repudiation of that hypothesis; but rather, it shows that the same criteria measuring loanword entrenchment cannot be applied to all languages as if they were equal. Of all the studies they compared (see Table 1 in Stammers & Deuchar, 2012, p. 634), theirs is the only one which includes a vulnerable language (Welsh). However, the direction of borrowing is inverse to that of the present case. Loanwords are being borrowed from a superstrate (dominant) language, English, into a substrate, vulnerable language, Welsh. Additionally, no comment is made on the language status of Welsh. In fact, language status appears to be rarely mentioned in loanword studies, let alone the possibility that it has an influence on entrenchment.

Very few loanword studies have focussed on loanwords from non-dominant substrate languages, let alone those which are threatened like Māori (Simons & Fennig, 2018). Jones (2005) writes that “the motivation for code-switching in situations of language death is not identical to that which occurs in ‘healthy’ languages” (p. 3). This can certainly be seen in New Zealand. In order to revitalise the Māori language, Māori loanword use by New Zealanders is actively encouraged by Te Taura Whiri i te Reo Māori, regardless of “whether they speak the language or not” (Calude, 2017). This stands in stark contrast to the borrowing of linguistic resources from other indigenous language minorities, such as Australian Aboriginals, who consider their heritage language to be a matter of intellectual property (Simpson, 2006). It has been shown that “some Aboriginal people distinguish between usership and ownership” (Zuckermann, 2011, p. 113), whereas Te Taura Whiri i te Reo Māori has stated that “the [Māori] language belongs to New Zealand and all New Zealanders” and not just those of Māori descent (Tahana, 2008). However, differences in use of loanwords between Māori and non-Māori (de Bres, 2006; Calude et al., 2017; Kennedy, 2001; Onysko & Calude, 2014) show that this ideal is not necessarily a reality, and that there may indeed be a perceived distinction between ownership and the right-to-use of Māori loanwords. New Zealanders may have been told that they have an equal right to use the Māori language regardless of their ethnicity or heritage, but this is not how they are behaving.

⁶³ It has already been shown that this hypothesis cannot be applied to the present study as social awareness has disrupted traditional measurements of entrenchment.

Further confusing this issue is the fact that the Māori language is both a *taonga* (treasure) and a *koha* (gift⁶⁴). The Māori language is widely “perceived as a *taonga*...rather than a commodity” (Calude et al., Forthcoming, p. 14). One ramification of this is that “the right to access the language is not equal for everyone, and therefore, care and sensitivity is exercised by speakers” (Calude et al., Forthcoming, p. 14). If there are perceptions of differences in right-to-use of the Māori language because it must be treasured, then it follows that the frequencies of use and choice of loanwords themselves will be influenced by user perceptions of their entitlement to them, above and beyond any other factor. Similarly, if the Māori language is a *koha*, then so too, by extension, is the Māori lexicon. Metge (2009) writes that:

From the beginning of their campaign to revive the Maori [sic] language, champions of te reo Maori [sic] have identified it as a taonga, an identification which places it squarely in the context of gift exchange, to be given away in expectation of a return. (p. 2)

Metge (2009) goes on to state that the acceptance of Māori loanwords into NZE, “however flawed, fulfils the ultimate purpose of gift exchange, the reinforcement of social and political relations” (p. 2). It is thus only logical that Māori loanwords in this corpus are being used as a means of revitalising a threatened language, and as an expression of political attitude and empowerment. Likewise, there is a growing recognition of the Māori language as a *koha*, in that a gift given should be used, based on the diachronic increase in recent years of Māori loanwords. The problem, it seems, is not that the *koha* is not being embraced; but rather, that not all New Zealanders are convinced that they have a right to use the *koha* in its entirety. This highlights a gap between reason-to-use and perceived right-to-use of loanwords. If Māori loanwords can express ethnic or national identity (Davies & Maclagan, 2006; Kennedy, 2001), then it is clear that those who more strongly self-identify as either Māori and/or ‘Kiwi’ will be more likely to perceive themselves as having a right to use certain Māori loanwords. This begs the question of which loanwords express what identity, and if there is any difference. However, there can

⁶⁴ As with so many Māori loanwords, the meaning of *koha* in NZE has been largely reduced from its traditional sense of a gift exchange that aids in “maintaining social relationships and has connotations of reciprocity” (Moorfield, 2018). See Moorfield (2018) for further explanation.

be no doubt that if a user has no understanding or connection to a loanword, then it is far less likely that they will use it.

Lack of restriction surrounding language ownership is one of many factors attributed to the revitalisation of the Hebrew language (Zuckermann, 2011). Thus, it is speculated that if the revitalisation of the Māori language is to continue successfully, users themselves must feel that there is no impediment to loanword use. Given that Māori loanword use in NZE is often culturally and politically loaded, it is anticipated that this will be a difficult obstacle to surmount. It is predicted that loanword frequencies will continue to grow in NZE, but only in areas where there is a reason for their use, and where users feel their right-to-use is assured. In this regard, there may be dramatic increases of more widely known loanwords in culturally specific domains that may have a trickle-down effect in other areas, while nonce borrowings or more contextually specific loanwords will be neglected by all but a few due to there being a lack of self-perceived ownership, entitlement and comprehension of them. If Māori loanword use is truly a means for revitalising the Māori language, then perceptions of right-to-use must be tackled. *Te Taura Whiri i te Reo Māori* has gone some way to addressing this; however, until non-Māori New Zealanders perceive their right-to-use as being equal to that of Māori speakers, then there will continue to be a discrepancy between the desire to use loanwords as a means to language revitalisation, and actual loanword use.

In light of this, it follows that a comparison of the use of Māori loanwords in NZE must be made with other substrate borrowings in superstrate languages, where there is a strong social movement to revitalise the substrate language. Links have been made between the revitalisation of the Māori language and that of Sámi in Norway (Albury, 2015), Mapudungun in Chile (Gallegos, Murray, & Evans, 2010) and Gaelic in Scotland (Timms, 2013). Although none of these studies consider loanwords, they do highlight the many similarities between Māori and their respective indigenous languages. It is logical that research on loanwords borrowed from indigenous languages in the process of revitalisation would be more beneficial to studies in NZE than on those from languages that are flourishing unaided. In this regard, studies on loanwords in Welsh, Irish and Scottish English might be the first port of call. Yet, this researcher proposes that little is to be gained here, as English has been borrowing from the Celtic languages for over 1,000 years (Durkin, 2014). If like is to be truly compared with like, language status and history

of colonisation are not enough. Colonisation and language contact must be more recent, at least within range of the less than 300 years of contact⁶⁵ between Māori and English. It has already been stated that it is Māori lexical items which make NZE unique from all other varieties of English (Gordon, 2005). If the presence of Māori words is what makes NZE distinct (Macalister, 2006b), then comparisons must be made with those superstrate languages in which loanwords from a substrate indigenous language make that variety distinct.

One possible comparison could be the case of Maya⁶⁶ loanwords in Yucatán Spanish. The arrival of the Spanish to Mexico in 1492 (Zamora, 1982), although preceding Captain Cook's arrival to New Zealand in 1773 (Macalister, 2006b) is much more recent than the history of language contact between Celtic and English. Moreover, loanwords from Maya are widely used by monolingual Spanish speakers, and account for the uniqueness of Yucatán Spanish from other Mexican Spanish varieties (Pfeiler, 2014). This has been attributed to the extreme geographical isolation of the Yucatán Peninsular (Pfeiler, 2014, p. 206); a fact not dissimilar to New Zealand's own geographical isolation as a South Pacific nation. Further parallels to New Zealand may be seen in that Maya has been displaced historically by Spanish, but is now moving towards "a reaffirmation of Maya language and culture" (Pfeiler, 2014, p. 220). Finally, as has been shown in the use of Māori loanwords as both an overt and covert means of expressing social, cultural and political attitudes, in Yucatán Spanish, "social dynamics are more powerful than governmental initiatives when it comes to determining the extent and frequency of language use in Yucatán state" (Pfeiler, 2014, p. 221). Thus, it is clear that Māori loanwords cannot be compared with loanwords in other languages unless they meet a more rigorous set of comparable criteria. It is suggested these be:

- Loanwords are borrowed from a substrate into a superstrate language
- Substrate language is vulnerable or at-risk, and superstrate language is healthy and non-threatened
- There is a history of colonisation by the superstrate language (within a similar timeframe) in the country concerned

⁶⁵ See Macalister (2006b).

⁶⁶ Both Pfeiler (2014) and Zamora (1982) use 'Maya' to refer to the language of the Maya and not 'Mayan'; this convention is followed here.

- Substrate language use has been historically discouraged or prohibited in certain/all social spheres
- Substrate language is undergoing language revitalisation efforts
- Loanwords are used by both bilingual and monolingual speakers
- Substrate language lexical items make the superstrate language distinct from all other varieties of that language
- Social awareness of loanword use
- Reason-to-use and right-to-use of loanwords is socio-politically motivated
- Geographical isolation of country concerned

It must be qualified that, although recommended, these criteria are by no means comprehensive or prescriptive. Neither should they preclude comparison of Māori loanwords with those of other languages simply because they are not all met. However, they demonstrate that factors influencing loanword use are far more complex than previous research suggests. Inter-lingual loanword study should not be a simple case of side-by-side comparison, as this risks the legitimacy of results. In this regard, Māori loanword use in NZE offers many possibilities for furthering our understanding of loanword theory-of-use around the world.

8. Conclusion

8.1. Limitations and future suggestions

It must be acknowledged that there were several factors which limited the scope of this study, both in the corpus construction and in the coding of the data. Certain loanword types presented particular challenges. Nonetheless, this study has also illuminated several possible avenues for potential future loanword research.

The most obvious limitation in constructing the corpus was the database from which newspaper articles were drawn. As the *New Zealand Herald* website had undergone changes making a replicable search impossible, *Newztext* was instead used. However, *Newztext* did not include any *New Zealand Herald* subsidiary newspapers from the South Island, a fact which was not clear until late in the data collection process. As a result, conclusions regarding loanword use in NZE over the ten-year period of this study are representative of certain New Zealand regions only. There was also an uneven yearly distribution of articles, primarily due to the recentness of one of the search terms used to net articles. Articles did not always have authors given (where they did, gender and ethnicity were often impossible to discern), and they also lacked information on page numbers and newspaper section which might have provided correlative data with loanword frequencies. Future construction of newspaper corpora would benefit from a more rigorous selection process when choosing databases from which newspaper articles are sourced, to ensure they have essential metadata, and a more even diachronic and geographical distribution. One solution to the author information problem is only including articles in which author details are given, however this risks building the corpus around a bias. Another suggestion is that studies of Māori loanwords in NZE newspapers could consider media subsidiaries other than the *New Zealand Herald*, such as *Stuff*. Care must also be taken when constructing corpora constrained to a single topic, as some topics may only have a limited amount of data available to them. More topic-bound corpora are needed for future loanword research in both newspapers and other media genres in order to ascertain if Māori loanword frequencies are indeed increasing diachronically in NZE, are simply the result of topical attraction, or both. A larger data set would also enable diachronic tracking of individual loanwords, to observe if they are increasing in use over time, or if their use is restricted to certain topics.

Coding the data also presented limitations in interpreting results. It has been demonstrated that in New Zealand attitudes towards the Māori language are equated with attitudes on Māori political issues. Use of Māori loanwords is therefore not only a positioning of identity, but it is also a declaration of social, cultural and political values. It has also been hypothesised that a correlation exists between attitude and loanword frequency. It is regrettable, therefore, that author attitudes to the Māori language were not coded for within the data. Attitudes and perceptions of loanword users, to both Māori loanwords and the Māori language, will be essential in future research. Although these are by nature problematic to measure, especially covert attitudes, qualitative data such as that provided by sociolinguistic interviews (Backus, 2014) or folk linguistics (Albury, 2016) would be a good place to start. Opinion pieces and editorial sections in newspapers could also provide such information. These data could then be compared to quantitative loanword data, including frequencies. Coding for article topic (whether within a topically constrained or unconstrained corpus) would also be beneficial in this area as it might draw out ‘topics within a topic’ which relate to particular expressions of attitude, such as the positive support of the Māori language in Māori Language Week articles. Although such coding was attempted in this study, it was eventually abandoned due to ambiguity in topic classification⁶⁷. A rigorous methodology would need to be adopted to avoid this happening again in the future.

Another issue this study was confronted with was that of proper nouns and high frequency loanword types. While the inclusion of the proper nouns *Maori*, *Pakeha*, *Matariki* and *Kiwi* gave some important insights into the behaviour of Māori loanwords in NZE, they also proved problematic in their high propensity for semantic variation. All of these proper nouns are polysemous. In particular, the loanword *Maori*, which made up more than half of all tokens in the corpus, was a challenge, as analysis had to be done manually. This was extremely time consuming and often semantic distinctions were unclear. To include it in future studies, *Maori* must be rigorously coded for its semantic meanings. The question must also be asked that, if *Maori* is such a high frequency outlier as has been documented here

⁶⁷ Initially, articles in the corpus were coded according to their perceived topic. Categories included: Māori language revitalisation, events occurring in Māori Language Week, history of the Māori Language Week, etc. However, there was strong and blurry overlap between topics in the majority of articles, and almost none had clear topical distinctions.

and elsewhere, might it be better put aside, to avoid skewing the data? *Te reo* is another loanword outlier that may need to be reconsidered; however, more research is first needed into whether the significant diachronic increase of *te reo* was not just the result of topical attraction in a topic-bound corpus. A more appropriate solution would be to devise a means of automating the semantic tagging of Māori loanwords.

In the meantime, our understanding of how and why Māori loanwords are used in NZE may be furthered by several means. Network analysis offers exciting possibilities in loanword attraction (see Calude et al., Forthcoming). Indications in the data are that Māori loanword frequencies are near-Zipfian in distribution; an observation that has been made before regarding loanwords in Welsh but left unexplored (Stammers & Deuchar, 2012, p. 640). Similarly, Māori loanword use as a change from above could be investigated using sociolinguistic methodologies (Labov, 1990).

Nor should the issues of entrenchment be neglected: frequency, morphological assimilation, listedness and markedness. All of these threads belong to the same cord of entrenchment, and cannot be considered alone. Frequencies may be more accurately measured by semantically tagging polysemous loanwords. The social movement to linguistically dis-entrench Māori loanwords by removing plural markers could be investigated in other areas of morphology, where possible. For example, it would also be interesting to consider pronunciation of spoken Māori loanwords from the standpoint of phonological dis-entrenchment. Regarding listedness, a comparison of listed Māori loanwords across NZE dictionaries, as well as investigating author criteria for including Māori loanwords therein, could be useful. A robust categorisation system of markedness, including author purpose of marking, as well as type of marking, is needed to corroborate claims that different modes of marking are used for different reasons. It may be that frequencies of markedness in loanwords will be higher in corpora not constrained to a single topic (see Sharp, 2007). More diachronic data on markedness is also required. This will necessitate a larger volume of data than was treated with here, and may prove difficult, as marked loanwords will need to be manually extracted.

Finally, future research into Māori loanwords in NZE cannot overlook the status of the Māori language as a vulnerable, substrate source of loanwords into the superstrate NZE. Attention must be given to language status and relationship, as

well as geographical and socio-political considerations, not just historical and cultural contexts, when comparing Māori loanword use with that of loanwords in other languages. The use of Maya loanwords in Yucatán Spanish is just one suggestion for doing so. However, it is clear that progress in studies on Māori loanwords, as well as loanword use in general, will only continue so far if they are treated as counterparts regardless of context.

8.2. Concluding remarks

This research has attempted to give a broad, quantitative overview of Māori loanwords in NZE as they occur in a corpus constrained to the topic of a yearly event which is becoming ever more prominent in the New Zealand consciousness: Māori Language Week. By positioning loanword frequencies within the lenses of traditional measures of loanword entrenchment – frequency, morphological assimilation, listedness and markedness – it has been shown that use of Māori loanwords in NZE cannot be evaluated by the same measures as loanwords from other, healthier, superstrate languages. There are a number of reasons for this. First of all, loanword use in New Zealand exists above the level of consciousness. There is a high level of social awareness of Māori loanwords, and an active and intentional movement to ‘dis-entrench’ them morphologically that does not exist in other studies on loanwords. Frequency of use is influenced by such complications as perceived right to use, reason to use, and topical attraction. Listedness is less helpful in establishing individual loanword entrenchment than in revealing perceptions of loanword use; while markedness has been appropriated as a teaching resource by some newspapers as a language revitalisation aid, muddying its accuracy as an objective gauge of entrenchment. These conventional methods of measuring entrenchment cannot be used reliably or accurately where loanword use is above the level of social consciousness, as many of those same methods are being disrupted to further social agendas, and being utilised intentionally to express political ideologies. In order to be effective, all four measures are required to attempt to position the entrenchment of a loanword, and even then, they are indicative only. Because Māori loanword use in NZE is a means of expressing political attitude – conscious and unconscious, overt and covert – towards all Māori issues, not only Māori language revitalisation, it is therefore a socio-political issue, and measures of entrenchment cannot ignore this.

Secondly, as the Māori language is perceived to be a *taonga* and a *koha*, not a commodity as is the case of loanwords in so many other languages, it cannot be evaluated as if it were. Māori loanwords are a vital tool in the political expression of a minority, substrate language community in a superstrate language; a *taonga* shared by all New Zealanders; a marker of ethnic and national identity; and a means of expressing cultural attitudes. There is also some evidence that increasing frequencies of Māori loanword use in NZE is the result of a change from above. All of these issues need to be taken into consideration when they are studied, and new methods are required in order to better research the entrenchment of Māori loanwords in NZE. However, there can be no doubt as to the importance of the issue of entrenchment. Neither is there any question that Māori loanwords are used in NZE in a manner so far unlike that of loanwords in any other language, and the research thereof is not only of importance for all New Zealanders, but also for research into loanword use in languages worldwide.

9. References

- Albury, N.J. (2015). Your language or ours? Inclusion and exclusion of non-indigenous majorities in Māori and Sámi language revitalization policy. *Current Issues in Language Planning*, 16(3), 315–334.
- Albury, N.J. (2016). Defining Māori language revitalisation: A project in folk linguistics. *Journal of Sociolinguistics*, 20(3), 287–311.
- An ‘inspirational’ day as Rotorua declared bilingual. (2017, August 11). *Rotorua Daily Post*. Retrieved from <http://www.nzherald.co.nz>
- Anderson, A., Binney, J., & Harris, A. (2014). *Tangata whenua: An illustrated history*. Wellington, New Zealand: Bridget Williams Books.
- Backus, A. (2014). A usage-based approach to borrowability. In E. Zenner & G. Kristiansen (Eds.), *New perspectives on lexical borrowing: Onomasiological, methodological and phraseological innovations* (pp. 19–39). Berlin, Germany: de Gruyter Mouton.
- Barffour, A.A. (2016). A study of English loanwords in French written texts and advertisements and the perceptions and attitudes of the Francophone readership (Doctoral dissertation, University of Alabama). Retrieved from https://ir.ua.edu/bitstream/handle/123456789/2685/file_1.pdf?sequence=1&isAllowed=y
- Bauer, L. (1993a). *Manual of information to accompany the ‘Wellington corpus of written New Zealand English’*. Wellington, New Zealand: Victoria University of Wellington.
- Bauer, L. (1993b). *The Wellington corpus of written New Zealand English*. Wellington, New Zealand: Victoria University of Wellington.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13, 145–204.
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139–173.
- Calude, A.S. (2017, September 27). Kia ora: how Māori borrowings shape New Zealand English [Blog message]. Retrieved from <http://theconversation.com/kia-ora-how-maori-borrowings-shape-new-zealand-english-82683>

- Calude, A.S., Miller, S., Harper, S., & Whaanga, H. (Forthcoming). Loanwords as textual collocates: Investigating Māori loanwords related to Matariki in New Zealand English newspapers. 1–21.
- Calude, A.S., Miller, S., & Pagel, M. (2017). Modelling loanword success – A sociolinguistic quantitative study of Māori loanwords in New Zealand English. *Corpus Linguistics and Linguistic Theory*, 1–38.
doi: 10.1515/cllt-2017-0010.
- Daly, N. (2007). Kūkupa, koro and kai: The use of Māori vocabulary items in New Zealand English children’s picture books. *New Zealand English Journal*, 21, 20–33.
- Daly, N. (2009). Overhearing tangi, tangaroa and taniwha: The reported effects of Māori loanwords in children’s picture books on language use and cultural knowledge of adult readers. *Te Reo*, 52, 3–16.
- Daly, N. (2013). Belonging and differentiating: Aspects of New Zealand national identity reflected in the New Zealand Picture Book Collection (NZPBC). *Birdbook*, 51(1), 73–79.
- Davies, C., & Maclagan, M. (2006). Māori words – Read all about it: Testing the presence of 13 Māori words in four New Zealand newspapers from 1999 to 2004. *Te Reo*, 49, 73–99.
- de Bres, J. (2006). Māori lexical items in the mainstream television news in New Zealand. *New Zealand English Journal*, 20, 17–34.
- Degani, M. (2010). The Pakeha myth of one New Zealand/Aotearoa: An exploration in the use of Maori loanwords in New Zealand English. In R. Facchinetti, D. Crystal, & B. Seidlhofer (Eds.), *From International to Local English – And Back Again* (pp.165–196). Bern, Switzerland: Peter Lang.
- Degani, M. (2017). Cultural conceptualisations in stories of Māori-English bilinguals: The cultural schema of marae. In F. Sharidian (Ed.), *Advances in Cultural Linguistic*, (pp. 661–682). https://doi:10.1007/978-981-10-4056-6_29
- Degani, M., & Onysko, A. (2010). Hybrid compounding in New Zealand English. *World Englishes*, 29(2), 209–233.
- Deverson, T., & Kennedy, G. (2005). *The New Zealand Oxford Dictionary*. Melbourne, Australia: Oxford University Press.

doi: 10.1093/acref/9780195584516.001.0001

- Durkin, P. (2014). *Borrowed words: A history of loanwords in English*. Oxford, England: Oxford University Press.
- Education Review Office. (2016). *Part 4: ERO's framework for review - Ngā pou here*. Retrieved from <http://www.ero.govt.nz/publications/he-pou-tataki-how-ero-reviews-early-childhood-services/part-4-eros-framework-for-review-nga-pou-here/>
- Fagan, S., & Gençay, R. (2011). An introduction to textual econometrics. In A. Ullah & D.E.A. Giles (Eds.), *Handbook of empirical economics and finance* (pp.133–153). Boca Raton, FL: Chapman & Hall/CRC.
- Gallegos, C., Murray, W.E., & Evans, M. (2010). Research note: Comparing indigenous language revitalisation: Te reo Māori in Aotearoa New Zealand and Mupudungun in Chile. *Asia Pacific Viewpoint*, 51(1), 91–104.
- Gardiner, H. (2016, July 4). Māori Language Week parade attracts thousands. *Māori Television*. Retrieved from <http://www.maoritelevision.com>
- Gordon, E. (2005, January 1). Macmillan Brown Lecture 2 [Radio broadcast]. In *The Macmillan Brown Lectures*. Retrieved from <http://www.radionz.co.nz>
- Grant, L.E. (2012). Culturally motivated lexis in New Zealand English. *World Englishes*, 31(2), 162–176.
- Grant-Russell, P., & Beaudet, C. (1999). Lexical borrowings from French in written Quebec English. *University of Pennsylvania Working Papers in Linguistics*, 6(2), 17–33. Selected papers from NWA V 27.
- Harlow, R. (2005). Covert attitudes to Māori. *International Journal of the Sociology of Language*, 172, 133–147.
- Holmes, J., Vine, B., & Johnson, G. (1998). *The Wellington corpus of spoken New Zealand English*. Wellington, New Zealand: Victoria University of Wellington.
- Jones, M.C. (2005). Some structural and social correlates of single word intrasentential code-switching in Jersey Norman French. *French Language Studies*, 15, 1–23. doi:10.1017/S0959269505001894
- Kachru, B.B. (1988). The sacred cows of English. *English Today*, 4, 3–8. doi: 10.1017/S0266078400000973

- Kennedy, G. (2001). Lexical borrowing from Maori in New Zealand English. In B. Moore (Ed.), *Who's centric now? The present state of post-colonial Englishes* (pp. 59–81). Canberra, Australia: Oxford University Press.
- Kennedy, G., & Yamazaki, S. (2000). The influence of Maori on the New Zealand English lexicon. In J.M. Kirk (Ed.), *Corpora galore: Analyses and techniques in describing English: Papers from the nineteenth international conference on English language research on computerised corpora (ICAME 1998)* [Conference proceedings] (pp. 33–44). Amsterdam, The Netherlands: Editions Rodopi B.V.
- King, M. (2007). *The Penguin History of New Zealand Illustrated*. North Shore, New Zealand: Penguin.
- Kruger, H. (2012). *Postcolonial polysystems: The production and reception of translated children's literature in South Africa*. Retrieved from <http://ebookcentral.proquest.com>
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205–254.
- Macalister, J. (2001). The transformation of the Kiwi. *English in Aotearoa*, 43, 20–22.
- Macalister, J. (2004). A survey of Maori word knowledge. *English in Aotearoa*, 52, 69–73.
- Macalister, J. (2006a). The Maori lexical presence in New Zealand English: Constructing a corpus for diachronic change. *Corpora*, 1(1), 85–98.
- Macalister, J. (2006b). The Maori presence in the New Zealand English lexicon, 1850–2000: Evidence from a corpus-based study. *English World-Wide*, 27(1), 1–24.
- Macalister, J. (2007). Revisiting weka and waiata: Familiarity with Maori words among older speakers of New Zealand English. *New Zealand English Journal*, 21, 41–51.
- Macalister, J. (2008). Tracking changes in familiarity with borrowings from te reo Māori. *Te Reo*, 51, 75–97.
- Macalister, J. (2009). Investigating the changing use of te reo. *NZ Words*, 13, 3–4.
- Mead, H. M. (2003). *Tikanga Māori: Living by Māori values*. Wellington, New Zealand: Huia Publishers.

- Metge, J. (2009). Words escape us. *NZ Words*, 13, 1–2.
- Ministry for Culture and Heritage. (2017). *History of the Māori language*. Retrieved from <https://nzhistory.govt.nz/culture/maori-language-week/history-of-the-maori-language>, updated 10-Oct-2017
- Moon, P. (2013). *Turning Points: Events that changed the course of New Zealand History*. Auckland, New Zealand: New Holland Publishers Ltd.
- Moorfield, J.C. (2018). *Te aka online Māori dictionary*. Retrieved from <http://www.maoridictionary.co.nz>
- Mutu, M. (1994, February). *The use and meaning of Maori words borrowed into English for discussing resource management and conservation in Aotearoa/New Zealand*. Paper presented at the Conservation Board Chairpersons Conference. New Zealand Conservation Authority.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Retrieved from <https://www.proquest.com/>
- Newztext Newspapers [Database]. (2018). Retrieved from <https://www-knowledge-basket-co.nz.ezproxy.waikato.ac.nz/databases/newztext/search-newztext/>
- NZME. (2018). *Our brands*. Retrieved from <http://www.nzme.co.nz/about-us/our-brands/>
- Onysko, A., & Calude, A. (2014). Comparing the usage of Māori loans in spoken and written New Zealand English: A case study of Maori, Pakeha, and Kiwi. In E. Zenner & G. Kristiansen (Eds.), *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations* (pp.143–169). Berlin, Germany: de Gruyter Mouton.
- Pawley, A. (1989). Can the Maori language survive? *Hurupaa: Undergrowth: Student Journal of Anthropology*, 10, 12–23.
- Pfeiler, B. (2014). Maya and Spanish in Yucatán: An example of continuity and change. In Salikoko S. Mufwene [Ed.], *Iberian imperialism and language evolution in Latin America* (pp. 205-224). Chicago, IL: The University of Chicago Press.
- Poplack, S., & Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24, 279–315.
- Poplack, S., Sankoff, D., & Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, 47–104.

- Poplack, S., Wheeler, S., & Westwood, A. (1989). Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. *World Englishes*, 8(3), 389–406.
- R Core Team. (2018). R: A language and environment for statistical computing (Version 3.5.1) [Computer software]. Retrieved from <https://www.R-project.org>
- Sharp, H. (2007). Swedish–English language mixing. *World Englishes*, 26(2), 224–240.
- Simons, G.F., & Fennig, C.D. (Eds.). (2018). *Ethnologue: Languages of the world* (21st ed.). Retrieved from <http://www.ethnologue.com>
- Simpson, J. (2006, November 25). Sovereignty over languages and land [Blog message]. Retrieved from the University of Sydney website: http://blogs.usyd.edu.au/elac/2006/11/sovereignty_over_languages_and_1.html
- Stammers, J. R., & Deuchar, M. (2012). Testing the nonce borrowing hypothesis: Counter-evidence from English-origin verbs in Welsh. *Language and Cognition*, 15(3), 630–643.
- Statistics New Zealand. (2018). [Regional statistics from 2013 census data]. Retrieved from <https://www.stats.govt.nz>
- Tahana, Y. (2008, July 23). Sky's the limit under new pact with national airline. *New Zealand Herald*. Retrieved from <http://www.nzherald.co.nz>
- Te Taura Whiri i te Reo Māori. (2007). *Annual report of Māori Language Commission for the year ended 30 June 2007*. Retrieved from https://www.parliament.nz/resource/mi-NZ/48DBHOH_PAP16067_1/7a807184c3921135f2c6a79a08416982061b3f84
- Te Taura Whiri i te Reo Māori. (2012). *Guidelines for Māori language orthography*. Retrieved from <http://www.tetaurawhiri.govt.nz/assets/Uploads/translators-docs/Guidelines-for-Maori-Language-Orthography.pdf>
- Te Wānanga o Aotearoa. (2018). *Te reo rangatira*. Retrieved from <https://www.twoa.ac.nz/Nga-Akoranga-Our-Programmes/Te-Reo-Maori-Maori-Language>

- Timms, C.E. (2013). Indigenous language revitalisation in Aotearoa New Zealand and Alba Scotland (Doctoral thesis, University of Otago, Dunedin, New Zealand). Retrieved from <http://hdl.handle.net/10523/3851>
- Turpin, D. (1998). 'Le français, c'est le last frontier': The status of English-origin nouns in Acadian French. *The International Journal of Bilingualism*, 2(2), 221–233.
- United Nations Department of Economic and Social Affairs. (2017). *World Population Prospects: The 2017 Revision*. Retrieved from <https://esa.un.org/unpd/wpp/Publications/>
- Zamora, J.C. (1982). Amerindian loanwords in general and local varieties of American Spanish. *Word*, 33(1–2), 159–171.
doi: 10.1080/00437956.1982.11435730
- Zenner, E., & Kristiansen, G. (2014). Introduction: Onomasiological, methodological and phraseological perspectives on lexical borrowing. In E. Zenner & G. Kristiansen (Eds.), *New perspectives on lexical borrowing: Onomasiological, methodological and phraseological innovations* (pp. 1–17). Berlin, Germany: de Gruyter Mouton.
- Zuckermann, G., & Walsh, M. (2011). Stop, revive, survive: Lessons from the Hebrew revival applicable to the reclamation, maintenance and empowerment of Aboriginal languages and cultures. *Australian Journal of Linguistics*, 31(1), 111–127. doi: 10.1080/07268602.2011.532859

Appendix A: A complete list of the loanwords

The following is a complete list of all Māori loanwords that appeared in the corpus, along with supplementary information on definitions, frequencies, listedness⁶⁸, markedness, morphological dis-entrenching, and macronisation (given in brackets where applicable). Definitions in this appendix were taken from Moorfield (2018) and adapted with permission. In cases where definitions here differ from those given in Moorfield (2018), or were unavailable therein, alternative sources were consulted. These sources are cited in the definition column. In the course of compiling this appendix, it was noted that there was often a discrepancy in meaning between Māori words and Māori loanwords, usually in the form of meaning reduction in loanwords. While every effort was made to give true, accurate and concise definitions of Māori loanwords, any mistakes are the author's own.

Loanword frequency	Loanword	Definition	Listedness	Frequency marked	Notes on morphological dis-entrenching
1	<i>ae</i>	(<i>āe</i>) yes, aye, OK	listed	0	
1	<i>akina</i>	definition in corpus: “akina means to promote, to push forward” (file TW_2016_04_JUL_BPT_001)	unlisted	1	
1	<i>akonga</i>	(<i>ākonga</i>) student	unlisted	1	
3	<i>apopo</i>	(<i>āpōpō</i>) tomorrow	unlisted	0	
3	<i>aroa</i>	love	listed	0.33	
1	<i>ataahua</i>	(<i>ātaahua</i>) beautiful, handsome	unlisted	1	

⁶⁸ Although definitions of loanwords were taken from Moorfield (2018) for best accuracy, listedness was measured from the *New Zealand Oxford Dictionary* (Deverson & Kennedy, 2005), as established in Section 6.1. A loanword was considered listed if it appeared in Deverson and Kennedy (2005) as a headword.

1	<i>ae</i>	(<i>auē</i>) expression of surprise, astonishment or distress, i.e., ‘heck’ or ‘oh dear’	listed	0	
2	<i>awa</i>	river, stream, creek	unlisted	0.5	
1	<i>e noho</i>	use in corpus: ‘sit down’ (file TW_2015_27_JUL_TDP_006)	unlisted	1	
1	<i>e tu</i>	(<i>e tū</i>) use in corpus: ‘stand up’ (file TW_2015_27_JUL_TDP_006)	unlisted	1	
23	<i>haka</i>	vigorous posture dance with actions and rhythmically shouted words	listed	0	
1	<i>hakeka</i>	cloud ear fungus, <i>Auricularia polytricha</i>	unlisted	1	
20	<i>hangi</i>	(<i>hāngī</i>) earth oven, food cooked in an earth oven	listed	0	
16	<i>hapu</i> (sub-tribe)	(<i>hapū</i>) sub-tribe. See Moorfield (2018) for full explanation of concept	listed	0	
1	<i>hapu</i> (pregnant)	(<i>hapū</i>) pregnant	unlisted	1	
2	<i>harakeke</i>	New Zealand flax, <i>Phormium tenax</i>	listed	0	
1	<i>hiko</i>	lightning	unlisted	1	
6	<i>hikoi</i>	(<i>hīkoi</i>) step, march, journey	listed	0	
3	<i>hori</i>	a Māori person (colloquial, offensive; definition taken from Deverson & Kennedy, 2005)	listed	0	
1	<i>horopito</i>	pepper tree, <i>Pseudowintera axillaris</i> and <i>Pseudowintera colorata</i>	listed	1	
10	<i>hui</i>	meeting, gathering, assembly	listed	0.2	
1	<i>iwa</i>	nine, 9	unlisted	0	
82	<i>iwi</i>	tribe, extended kinship group	listed	0.02	
1	<i>ka kite</i>	‘I’ll see you again’. See Moorfield (2018) for explanation of use	unlisted	0	1 token given ‘s’ plural suffix; an in-text example of Māori greetings
5	<i>ka pai</i>	good	listed	0	
1	<i>kaea</i>	leader of a <i>haka</i> or <i>waiata</i>	unlisted	1	
17	<i>kai</i>	food, meal	listed	0.06	

2	<i>kai moana</i>	seafood, shellfish	listed	0.5	
2	<i>kaiako</i>	teacher, instructor	listed	0.5	
1	<i>kaiarahi reo</i>	(<i>kaiārahi reo</i>) language guide, mentor. From <i>kaiārahi</i> (guide, escort, counsellor, conductor, escort, leader, mentor) and <i>reo</i> (see Moorfield, 2018)	unlisted	0	
1	<i>kaimahi</i>	worker, employee	listed	1	
1	<i>kaitiaki</i>	guardian, caregiver, trustee	listed	1	
2	<i>kaiwhakaako</i>	teacher, lecturer, coach, trainer, instructor	unlisted	0.5	
2	<i>kaiwhakahaere</i>	administrator, manager	unlisted	0.5	
47	<i>kapa haka</i>	<i>haka</i> group, Māori cultural performing group	listed	0	
5	<i>karakia</i>	use in corpus: “prayer” (file TW_2014_21_JUL_TDP_005). See Moorfield (2018) for full explanation of concept	listed	0.6	
2	<i>karanga</i>	a ceremonial call of welcome to visitors at the start of a <i>pōwhiri</i> . See Moorfield (2018) for full explanation of concept	listed	0.5	
17	<i>kaumatua</i>	an elder person of status within the <i>whānau</i>	listed	0	
16	<i>kaupapa</i> (not <i>kura kaupapa</i> , <i>kaupapa Maori</i>)	purpose, topic	listed	0.19	
1	<i>kaupapa Maori</i> (not <i>kura kaupapa Maori</i>)	(<i>kaupapa Māori</i>) Māori customary approach	unlisted	0	
1	<i>kauri</i>	large forest tree found only in the northern North Island, <i>Agathis australis</i>	listed	0	
2	<i>kawa</i>	<i>marae</i> protocol. Definition in corpus: “customs”, “lore” (files TW_2013_02_JUL_TDP_008; TW_2014_23_JUL_TDP_003)	listed	1	
1	<i>kawakawa</i>	pepper tree, <i>Macropiper excelsum</i>	listed	1	
1	<i>kereru</i>	(<i>kererū</i>) New Zealand pigeon, <i>Hemiphaga novaeseelandiae</i>	listed	0	
2	<i>kete</i>	basket	listed	0.5	
1	<i>ki</i>	(<i>kī</i>) round flax ball used in <i>kī-o-rahi</i>	unlisted	0	
2	<i>ki-o-rahi</i>	(<i>kī-o-rahi</i>) a traditional Māori ball game	unlisted	0	

1	<i>kia kaha</i>	be strong, get stuck in, keep going	listed	0	
33	<i>kia ora</i>	hello, cheers, good luck, best wishes	listed	0	1 token given ‘s’ plural suffix; an in-text example of Māori greetings
1	<i>kia tere</i>	definition in corpus: “hurry up” (file MLW_2012_25_JUL_TDP_011)	unlisted	1	
1	<i>kiwi</i> (bird)	flightless bird native to New Zealand, <i>Apteryx mantelli</i> , <i>Apteryx australis</i> and <i>Apteryx owenii</i>	listed	0	1 token given ‘s’ plural suffix
37	<i>Kiwi</i> (person)	a New Zealander (definition taken from Deverson & Kennedy, 2005)	listed	0	In all cases where <i>Kiwi</i> referred to more than one person (19 tokens), an ‘s’ plural suffix was added. This also occurred once in the compound <i>Kiwi-isms</i>
1	<i>koauau</i>	(<i>kōauau</i>) cross-blown flute	listed	1	
4	<i>koha</i>	gift, donation. See Moorfield (2018) for full explanation of concept	listed	0.25	
19	<i>kohanga</i>	see <i>kōhanga reo</i> ; Māori language preschool	listed	0	
32	<i>kohanga reo</i>	(<i>kōhanga reo, te</i>) Māori language preschool	listed	0.03	
16	<i>korero</i>	(<i>kōrero</i>) to speak, talk	listed	0.19	
5	<i>koro</i>	elderly man, grandfather	listed	0.2	
2	<i>koroua</i>	elderly man, grandfather	listed	0	
2	<i>korowai</i>	ornamental cloak made of <i>muka</i> (New Zealand flax fibre)	listed	0.5	
2	<i>kotahitanga</i>	unity, togetherness, solidarity, collective action	listed	0.5	
1	<i>kowhaiwhai</i>	(<i>kōwhaiwhai</i>) painted scroll ornamentation, commonly used on meeting house rafters	listed	1	
15	<i>kuia</i>	elderly woman, grandmother, female elder	listed	0.07	

2	<i>kumara</i>	(<i>kūmara</i>) sweet potato, <i>Ipomoea batatas</i>	listed	0	
7	<i>kupu</i>	word	unlisted	0.57	
30	<i>kura</i>	school	listed	0.07	
28	<i>kura kaupapa (Maori)</i>	Māori language primary school	listed	0.04	
9	<i>mahi</i>	work	unlisted	0.33	
2	<i>makariri</i>	cold	listed	0.5	
7	<i>mana</i>	prestige, authority, influence, charisma. See Moorfield (2018) for full explanation of concept	listed	0.29	
2	<i>mana whenua</i>	territorial rights, power from the land, authority or jurisdiction over land or territory	unlisted	0	
21	<i>manaakitanga</i>	hospitality, kindness, generosity, support, the process of showing respect, generosity and care for others	unlisted	0.33	
5	<i>manuhiri</i>	visitor, guest	listed	0.8	
1,649	<i>Maori</i>	(<i>Māori</i>) indigenous person of <i>Aotearoa</i> New Zealand. See Moorfield (2018) for further definitions and usages	listed	<0.01	1 token referring to 'Māori people' given 's' plural suffix; occurred in reported speech
6	<i>Maoridom</i>	(<i>Māoridom</i>) Maori people as a whole; the Maori world and culture (definition taken from Deverson & Kennedy, 2005)	listed	0	
2	<i>Maoritanga</i>	(<i>Māoritanga</i>) Māori culture, Māori practices and beliefs, Māoriness, Māori way of life	listed	0	
51	<i>marae</i>	courtyard, open area in front of the <i>wharenui</i> , where formal greetings and discussions take place. May also include the complex of buildings around the <i>marae</i>	listed	0.02	
1	<i>matapihi</i>	window	unlisted	1	
26	<i>Matariki</i>	1. The Pleiades, Messier 45 star constellation 2. Definition and main use in corpus: "Māori New Year" (e.g., file TW_2015_29_JUL_TDP_004)	unlisted	0.12	
1	<i>matauranga</i>	(<i>mātauranga</i>) knowledge, wisdom, understanding, skill	unlisted	0	

4	<i>matua</i>	father, parent, uncle	unlisted	0	
1	<i>mau raakau</i>	(<i>mau rākau</i>) Māori weaponry. Definition in corpus: “Maori martial art disciplines” (file MLW_2010_27_JUL_TNA_004)	unlisted	1	
1	<i>maunga</i>	mountain	unlisted	1	
1	<i>mauri</i>	life force. See Moorfield (2018) for full explanation of concept	listed	0	
8	<i>mihi(mihi)</i>	(<i>mihimihi, mihi</i>) speech of greeting, acknowledgement, tribute	listed	0.5	
4	<i>mita</i>	dialect	unlisted	0.25	
10	<i>moko</i> (tattoo)	traditional Māori tattoo. See also <i>tā moko</i> , which is only done on the face	listed	0	
15	<i>moko(puna)</i>	grandchild, grandchildren, descendant	listed	0.13	12 <i>moko(puna)</i> tokens referred to ‘grandchildren’ (not ‘grandchild’); 2 tokens of these gave an ‘s’ plural suffix
1	<i>morena</i>	(<i>mōrena</i>) good morning	unlisted	0	
3	<i>motu</i>	definition in corpus: “area” (file MLW_2014_04_NOV_TDP_002)	unlisted	0.33	
3	<i>nga mihi (nui)</i>	(<i>ngā mihi</i>) acknowledgements, greetings; definition in corpus: “thank you” (file TW_2017_14_SEPT_TDP_006)	unlisted	0.33	
1	<i>ono</i>	six, 6	unlisted	0	
1	<i>ora</i>	life, health, vitality	unlisted	1	
5	<i>paepae</i>	orators’ bench	listed	0.2	
39	<i>Pakeha</i>	(<i>Pākehā</i>) New Zealander of European descent	listed	0	
2	<i>pakeke</i>	adult. Definition in corpus: “senior” (file MLW_2012_28_JUL_HBT_001)	unlisted	0.5	
3	<i>papakāinga</i>	(<i>papa kāinga, papakāinga</i>) original home, home base, village, communal Māori land	listed	0	
1	<i>patiki</i>	(<i>pātiki</i>) flounder, a general term for flounder-type fish	listed	1	

1	<i>paua</i>	(<i>pāua</i>) abalone, sea ear, <i>Haliotis</i>	listed	0
1	<i>pene rakau</i>	(<i>pene rākau</i>) pencil	unlisted	1
2	<i>pepe/pepi</i>	(<i>pēpē, pēpi</i>) baby, infant	unlisted	1
11	<i>pepeha</i>	tribal saying, motto or proverb	unlisted	0.36
1	<i>pikopiko</i>	common shield fern, <i>Polystichum neozelandicum</i> subspecies <i>zerophyllum</i>	listed	1
1	<i>piripiri</i>	definition in corpus: “bush pepper”, <i>Gonocarpus micranthus</i> subsp. <i>Micranthus</i> (file TW 2010_26_JUL_TDP_004)	unlisted	1
8	<i>poi</i>	a light ball on a string which is swung or twirled rhythmically to sung accompaniment. See Moorfield (2018) for full explanation of concept	listed	0
1	<i>pouako</i>	teacher	unlisted	1
1	<i>pouarahi reo</i>	unclear meaning, possibly Māori language supervisor. “Pou ārahi refer to the people who ‘ārahi’ or provide guidance, supervision and direction to others” (Education Review Office, 2016)	unlisted	0
13	<i>powhiri</i>	(<i>pōwhiri</i>) welcome ceremony on a marae	listed	0
1	<i>puha</i>	(<i>pūhā</i>) sowthistle, <i>Sonchus arvensis</i> , <i>Sonchus oleraceus</i> , <i>Sonchus asper</i> , and <i>Sonchus kirkii</i>	listed	0
3	<i>puku</i>	stomach	listed	0.67
4	<i>rakau</i>	(<i>tī rākau</i>) stick game	listed	0.25
5	<i>rangatahi</i>	younger generation, youth.	listed	0.4
2	<i>raranga</i>	weaving	unlisted	1
59	<i>reo</i> (no preceding <i>te</i>)	language, dialect	listed	0.02
2	<i>reo irirangi</i> (<i>Maori</i>)	radio	unlisted	0
1	<i>rima</i>	five, 5	unlisted	0
4	<i>rohe</i>	district, region, territory, area	unlisted	0.25
2	<i>ropu</i>	(<i>rōpū</i>) group, association, committee	unlisted	0

1	<i>roro</i>	brain	unlisted	1
1	<i>rorohiko</i>	computer	unlisted	1
1	<i>rua</i>	two, 2	unlisted	0
1	<i>rumaki</i>	use in corpus: “immersion class” (file MLW_2013_01_JUL_HBT_002)	unlisted	0
1	<i>runanga</i>	(<i>rūnanga</i>) tribal council, board. See Moorfield (2018) for full explanation of concept	listed	0
2	<i>taha Maori</i>	(<i>taha Māori</i>) Māori identity, Māori side	listed	1
1	<i>tahi</i>	one, 1	unlisted	0
1	<i>taiaha</i>	long wooden weapon of hard wood with one end carved and often decorated with dogs’ hair	listed	0
1	<i>taiao</i>	world, Earth, natural world, environment, nature, country	unlisted	1
1	<i>taihoa</i>	use in corpus: “wait” (file MLW_2013_02_JUL_TDP_005)	listed	0
3	<i>tamaiti</i>	child, boy	listed	0.33
17	<i>tamariki</i>	children	listed	0.18
6	<i>tangata whenua</i>	local people, indigenous people, people born of the <i>whenua</i> (land)	listed	0.33
10	<i>tangi(hanga)</i>	(<i>tangihanga</i>) funeral. See Moorfield (2018) for full explanation of concept	listed	0.1
4	<i>taniwha</i>	water spirit, monster	listed	0
23	<i>taonga</i>	treasure. See Moorfield (2018) for full explanation of concept	listed	0.35
1	<i>tapu</i>	sacred. See Moorfield (2018) for full explanation of concept	listed	0
1	<i>taringa</i>	ear	unlisted	1
1	<i>taukumekume</i>	dispute, disagreement. Definition in corpus: “debate” (file TW_2008_23_JUL_HBT_001)	unlisted	1
1	<i>tautoko</i>	to support	unlisted	0
3	<i>te ao Maori</i>	definition in corpus: “the Māori world”, “Māori worldview” (file TW_2012_23_JUL_TDP_001)	unlisted	0.33

1	<i>te ara reo</i>	use in corpus: beginners' Māori language classes (see Te Wānanga o Aotearoa, 2018)	unlisted	0	
949	<i>te reo</i>	see <i>reo</i> ; use in corpus: “the language”, “the Māori language” e.g., file TW_2014_23_JUL_TDP_003)	listed	<0.01	
1	<i>tekau</i>	ten, 10	unlisted	0	
1	<i>tena koe</i>	(<i>tēnā koe</i>) hello! (speaking to one person), thank you	listed	0	
1	<i>tiaki</i>	to guard, keep	unlisted	1	
1	<i>tika</i>	correctly	listed	1	
14	<i>tikanga</i> (not <i>tikanga Maori</i>)	correct procedure, custom. See Moorfield (2018) for full explanation of concept	listed	0.29	
5	<i>tikanga Maori</i>	see <i>tikanga</i> ; correct Māori procedure, Māori custom	unlisted	0	
2	<i>tinu rangatiratanga</i>	self-determination, sovereignty, autonomy, self-government	listed	0	
7	<i>tipuna/tupuna</i>	1. (<i>tipuna, tupuna</i>) ancestor, grandparent 2. (<i>tīpuna, tūpuna</i>) ancestors, grandparents	listed	0	
1	<i>tititorea</i>	definition in corpus: “stick games” (file TW_2012_25_JUL_TNA_002)	unlisted	1	
1	<i>toa</i>	warrior, champion	listed	1	
1	<i>tohunga</i>	priest, healer	listed	1	
1	<i>toru</i>	three, 3.	unlisted	0	
1	<i>totara</i>	(<i>tōtara</i>) large forest trees found throughout New Zealand, <i>Podocarpus totara</i> , and <i>Podocarpus cunninghamii</i>	listed	0	
1	<i>tukutuku</i>	ornamental lattice-work used between carvings around meeting house walls	listed	0	
2	<i>turangawaewae</i>	(<i>tūrangawaewae</i>) standing place where one has rights of residence and belonging through kinship and <i>whakapapa</i>	listed	1	
1	<i>utu</i>	revenge. See Moorfield (2018) for full explanation of concept	listed	1	
3	<i>wahine</i>	1. (<i>wahine</i>) woman 2. (<i>wāhine</i>) women	listed	0	
1	<i>wai</i>	water, stream, creek, river	unlisted	1	

19	<i>waiata</i>	song	listed	0.21	
2	<i>wairua</i>	spirit, soul	listed	0.5	
6	<i>waka</i>	canoe	listed	0	
11	<i>waka ama</i>	outrigger canoe. Use in corpus: “outrigger canoe sport” (file TW_2012_27_JUL_HBT_001)	unlisted	0	
3	<i>wananga</i> (seminar/course)	(<i>wānanga</i>) seminar, conference	unlisted	0.67	
6	<i>wananga</i> (university)	(<i>wānanga</i>) tertiary institution that caters for Māori learning needs	unlisted	0	
1	<i>waru</i>	eight, 8	unlisted	0	
1	<i>wero</i>	challenge	listed	0	
1	<i>weta</i>	(<i>wētā</i>) large native New Zealand insect	listed	0	
1	<i>wha</i>	(<i>whā</i>) four, 4	unlisted	0	
2	<i>whaea</i>	mother, aunt. Definition in corpus: “teacher” (file MLW_2016_06_JUL_KN_001)	unlisted	0	
1	<i>whai</i>	cat's cradle, string game	unlisted	1	
3	<i>whaikorero</i>	(<i>whaikōrero</i>) oratory, formal speech	listed	0.67	
1	<i>whakaaro</i>	thought, opinion	unlisted	1	
2	<i>whakairo</i>	carving	listed	0.5	
1	<i>whakama</i>	(<i>whakamā</i>) to be shy, embarrassed. Definition in corpus: “to shy away” (file MLW_2009_01_AUG_BPT_005)	listed	1	
2	<i>whakamoemiti</i>	praise, thanks. Definition in corpus: “prayers” (file MLW_2012_23_JUL_HBT_002)	unlisted	0.5	
18	<i>whakapapa</i>	genealogy, lineage	listed	0.17	
4	<i>whakatauki</i>	(<i>whakataukī</i>) proverb, significant saying	unlisted	1	
55	<i>whanau</i>	(<i>whānau</i>) family	listed	0.02	
4	<i>whanaungatanga</i>	relationship, kinship, sense of family connection	listed	0.5	

4	<i>whare</i>	house	listed	0.5	
2	<i>whare pukapuka</i>	library	unlisted	0.5	
1	<i>wharekai</i>	dining hall	listed	1	
3	<i>wharekura</i>	Māori language secondary school	unlisted	0.33	
2	<i>wharenuī</i>	meeting house, main building of <i>marae</i>	listed	0	
1	<i>whitu</i>	seven, 7	unlisted	0	
3,795					