

The Use of Māori Words in National Science Challenge Online Discourse

Journal:	Journal of the Royal Society of New Zealand
Manuscript ID	NZJR-2019-0041.R1
Manuscript Type:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Calude, Andreea; University of Waikato, Stevenson, Louise; University of Waikato, School of General and Applied Linguistics Whaanga, Hemi; University of Waikato, Faculty of Maori and Indigenous Studies Keegan, Te Taka; University of Waikato, School of Mathematical and Computing Sciences
Keywords:	te reo Māori, linguistics, corpus linguistics, borrowings, National Science Challenge

SCHOLARONE™ Manuscripts

The Use of Māori Words in National Science Challenge Online Discourse

New Zealand English is well-known for its heavy borrowing of words from Māori. This lexical component, unique to New Zealand English alone has been studied intensely over the last 50 years, particularly in newspaper media. Current research suggests the use is still increasing today, primarily in Māori-related contexts. Here, we analyse a surprising and unexpected use of Māori loanwords in science digital discourse (neither genre being previously investigated), where we discover a strong presence of Māori borrowings in National Science Challenge website and Twitter content. Using corpus linguistics methods, we argue that the use of Māori loanwords in this genre functions as a national identity building tool, used by various authors to signal that the 'challenges' the country faces are uniquely New Zealand's 'challenges'.

Keywords: Māori loanwords, New Zealand English, National Science Challenge, corpus linguistics

Introduction – The National Science Challenges

In 2012, the New Zealand Government announced a new approach to funding public science by seeking input to decide the key scientific questions that New Zealand (NZ) needed to answer.¹ A number of strategic areas of scientific focus were developed which led to what became known as the National Science Challenges (NSCs). The NSCs are a set of 11 over-arching projects that concentrate NZ's scientific efforts on the major issues and opportunities facing the country. The NSCs have been allocated ten years of funding over two five-year periods. A review was undertaken after the first five-year period, to establish whether the NSCs were on track and warranted further funding. A summary of the 11 NSCs is given in Appendix S.1. (see Supplementary Materials).

_

¹ Source: https://www.beehive.govt.nz/release/govt-enlist-public-nz-science-challenges

A key difference in the NSC approach was that the alliances formed to address the mission-led goals resulted in scientists connecting and collaborating across a range of institutions, including universities and Crown Research Institutes (Daellenbach et al. 2017). Also, as the second tranche of funding was reliant on a positive outcome of the first five years of research, the NSCs were encouraged to disseminate their missions and intentions, their activities and their outputs. The most common vehicle for this dissemination was not only the NSC websites, but also their Twitter feeds, a unique and current discourse of the science issues currently facing New Zealand and the research being undertaken to address these issues. This paper is concerned with analysing that language.

Background to the language contact situation – why study loanwords?

It has long been argued that the most distinctive and salient feature of New Zealand English is its widespread use of words adopted from Te Reo Māori (Deverson, 1991). Not only has New Zealand English been actively and productively adopting Māori words ever since the first contact with British English took place, but their use is still increasing today in both frequency of existing adopted loanwords, and with regard to the import of new words from Māori (Macalister, 2006; Kennedy & Yamazaki, 1999; Calude et al., forthcoming). This increase is not a straight forward matter because research shows that the use of Māori loanwords (or borrowings) is not homogenous across people, with some people – most notably, Māori women (see Calude et al., 2018) – using more loanwords than others; nor is their use homogenous across topics of discourse, with some topics attracting more loanwords than others (specifically, topics pertaining to Māori people or Māori culture, cf. de Bres, 2006; Degani, 2010; Macalister, 2006; Davies & Maclagen, 2006). Interpreting the motivation and consequences of the use of Māori loans is similarly complex, with opinions remaining divided between (a) borrowings as indicators of positive attitudes towards Māori language and culture (Māori Language Commission), and (b) borrowings as perpetuators of negative stereotypes about the Māori world (Degani, 2010; Harlow, 2005).

Within the New Zealand context, the study of loanwords is important because it highlights the newly emerging voice that New Zealanders are choosing to represent their identity (or at least, part of it, for those who choose to use Māori borrowings). As the new dialect of English is coming into its final stage of separation from other English varieties

(Schneider, 2003), the time is ripe to consider the linguistic ways in which New Zealanders express distinctiveness, particularly in contrast to British English and Australian English, our closest related linguistic 'relatives'.

A second reason why the study of loanwords provides a fruitful endeavour relates to the fact that unlike other linguistic features (for instance, vowel mergers in pairs such as *air* and *ear*, see Maclagan & Gordon, 1996, Holmes & Bell, 1992 inter alia), Māori words and their use in New Zealand English represent both a salient, and a conscious, deliberate move. This move is embedded in a political and social perspective that strays from a neutral position. Evidence of the awareness and strength of such a linguistic move can be seen in strong comments from high-profile New Zealand figures found in the current New Zealand media, ranging from defiant and unwavering promotion of the use of loanwords (see quote 1 from Guyon Espiner), to strong opposition to their use (see quote 2 by Don Brash, former National Party leader).

(1) Quote from Guyon Espiner in favour of using Māori loanwords

"Armed with my tenuous grasp on the language I began to weave it into my radio presenting as co-host of RNZ's Morning Report. We've always done basic greetings in Māori on the show but I wanted to do more and began to extend the greetings and include basic information in Māori – such as the days, dates and temperatures for the main centres."

[source: My te reo Māori journey, 13/09/2017, The Spinoff]

(2) Quote from Don Brash against using Māori loanwords

"It's tokenism of the worst kind," he [Don Brash] says. Don Brash is continuing his attacks on the use of Te Reo on Radio New Zealand. The former National Party leader appeared on The AM Show on Wednesday, after saying he was "utterly sick" of the use of Aotearoa's indigenous language in broadcasting.

[source: "Don Brash's scathing statement on Te Reo usage", 29/11/2017, Newshub]

Thirdly, the situation of language contact in New Zealand provides a unique window into loanwords more generally, which is of interest to language researchers outside New Zealand. Within the wider international context, studying loanwords is itself not a new area of inquiry, nor is it linguistically diverse: studies often analyse English words which

are borrowed into other languages, such as, French, German, Brazilian Portuguese, and so on. Recently, a new methodological paradigm shift has breathed new light into this research topic (see a literature review summary of novel approaches to the phenomenon of "loanword success" in Calude et al., 2008). But the study of Māori loanwords specifically has drawn interest outside the New Zealand linguistic community because of the unique sociolinguistic context in which these words are found. Here, we have a world dominant language (English) adopting words from a non-dominant language which is (still) fighting for its survival (Te Reo Māori). Not only is this language contact scenario distinct from all others studied so far (to our knowledge), but counter to many other language contact scenarios (such as, those mentioned above), the adoption of Māori loanwords in New Zealand English is actively encouraged by the Māori Language Commission, as well as many other groups (see for instance, a recent piece about Radio NZ's choice to increase the use of Māori words throughout their English language broadcasts by RNZ chief executive Paul Thompson, RNZ 2017).

Data and Methods - National Science Challenge as a coherent discourse

Given the observations made by previous researchers in the area of Māori loanwords, we were interested to find out to what extent words of Māori origin were used in New Zealand English science discourse, and online sources – neither of which had previously received any attention. In this section, we outline the methods used to gather and code our data.

Over the last thirty years, the field of linguistics has experienced a dramatic paradigm shift towards a more empirical methodology, which rather than relying on internal introspection, is heavily based on collecting real language exchanges, be they spoken or written (see a review in McEnery & Hardie, 2011). While corpus studies still utilise a great deal of introspective thinking, for example, in defining the topic of research, and in designating research questions, the analyses ultimately rest on concrete and real occurring examples, which have the added advantage of being falsifiable.

We adopt a corpus linguistics framework here to analyse the use of Māori loanwords in the *National Science Challenge* online presence. To this end, we built a corpus termed the National Science Challenge Corpus (NSCC) in late 2017. The NSCC is made up of two types of datasets. The first and largest dataset comes from the internet

websites of the eleven National Science Challenges, which we gathered manually by copy-and-paste. The text came in different formats, usually in HTML code or as part of PDF documents and these were converted to plain text files, keeping the material from each Challenge separate. All graphical information was ignored. The websites appeared to share certain structural features, such as, (1) a paragraph summarising their main research questions and agenda, (2) a main content page detailing ongoing and current projects, and (3) other supplementary materials included as PDF file attachments (for some Challenges only).

The second dataset consisted of a body of tweets which were identified by following the hashtags of each Challenge. Ten from the eleven Challenges have active Twitter feeds and we used the open source software FireAnt (Anthony & Hardaker, 2016) to collect all the text included in these Twitter feeds. Regular expressions were used to 'clean up' the data and discard URLS and any other non-textual material (graphics). These materials gave us a corpus of approximately 1.5 million words (see Supplementary Materials, Table S.2. for details).

Once the corpus was collated, we used the open source concordancing software AntConc (Anthony, 2018) to extract the most frequently occurring loanwords in it and to perform keyness analyses, discussed in the following sections.

Results

This section is divided into three parts; the first deals with the general use of loanwords in the NSCC, reporting which loanwords appear in the NSCC and differences between the Twitter data and the website data. The second section considers the spread of loanwords across the eleven Challenges. The third section provides comparisons between the NSCC and other genres of New Zealand English with respect to loanwords using keyness measures. We conclude with a note on the use of the loanword *mātauranga*, which constitutes a novel importation from Māori and whose use in English is specifically linked to science discourse.

Use of loanwords in the NSCC

The NSCC comprises hundreds of Māori loanwords. Because there is (currently) no way to automate the detection of loanwords, manual inspection is required to identify

exhaustive lists of all Māori loanwords used in a given body of data.² Owing to the size of our corpus, such a task would have been extremely time-consuming and error-prone. We opted, instead, for using AntConc to build word frequency lists which allowed us to extract the *most frequently* occurring loanwords. We did this for the entire NSCC and for each Challenge separately, in turn. Let us first consider the Māori loanwords in the entire corpus. The twenty most frequent loanwords are given in Table 1.

Table 1. The 20 most frequent Māori loanwords in the entire NSCC corpus and their associated raw frequencies of occurrence

Rank frequency	Māori loanword	English equivalent	Raw frequency
1	MĀORI/ MAORI	(of) indigenous (origin)	5318 (with macron) +414 (without macron) =5732
2	MĀTAURANGA	traditional knowledge	1640
3	IWI	tribe	1348
4	OTAGO	Otago	904
5	HAPŪ	sub-tribe	547
6	WAIKATO	Waikato	409
7	WHENUA	land	382
8	WHĀNAU	extended family	340
9	KAIKŌURA	Kaikoura	336
10	KAITIAKITANGA	guardianship	335
11	NGĀTI	prefix to tribal name	334
12	KAI	food	324
14	KĀHUI	group	321
14	AOTEAROA	New Zealand	312
15	TAMARIKI	children	304
16	TIKANGA	custom	259
17	TANGAROA	God of the sea and fish	253
18	KAITIAKI	guardian	225
19	WHAI	to have, to pursue	214
20	ORA	health, to be alive	204

² It is theoretically possible to use machine learning methods to check individual words from a given body of text (say a corpus such as our data) against online Māori and English language dictionaries (both of which are available) in order to extract lists of Māori loanwords. However, unfortunately this method

quickly runs into trouble because a great majority of loanwords from Māori are proper nouns (personal names, names of organisations and place names) which do not appear in the dictionary. We believe that the problem of extracting and labelling Māori loanwords might necessitate a machine learning solution and leave this for future research to tackle.

Normalising their frequency over 10,000 words gives us the following distribution, see Figure 1.

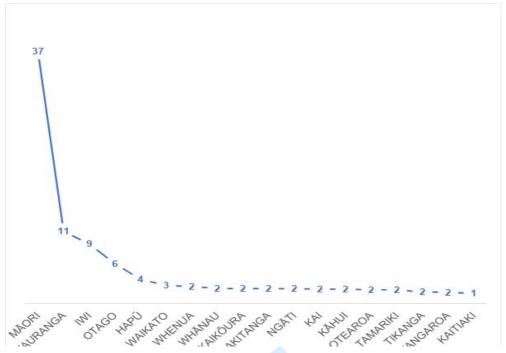


Figure 1. The 20 most frequent Māori loanwords in the entire NSCC.

Unsurprisingly, the word *Māori* is itself by far the most widely used loanword. The next most frequently used loan in this data is a relatively new lexical borrowing, namely *mātauranga* 'knowledge, wisdom'. No other New Zealand English corpus analysed so far appears to exhibit this loanword. It is noteworthy that the use of the loanword has gained momentum in New Zealand English and its use has come to prominence in Māori education and research circles in recent times (Mead, 2012; Royal, 2012; Smith et al., 2016). As noted by Hirini Moko Mead (2012, p. 9):

It was not so long ago that the term 'mātauranga Māori' was rarely ever mentioned in education circles Over the last two decades the term 'mātauranga Māori' has become increasingly important as more and more people are engaged in efforts to understand what it means. Put simply, the term refers to Māori knowledge.

Recent research and writing has highlighted the growing awareness and importance of mātauranga Māori (Awatere, 2008; Hikuroa, 2017; Mercier, 2013; Mika, 2012; Royal, 2009; Smith, 2012; Smith et al., 2016). Moreover, at a national research scale, mātauranga

Māori has been framed by the 'Vision Mātauranga Policy' (MoRST, 2007), whose mission is to "unlock the innovation potential of Māori knowledge, resources and people to assist New Zealand to create a better future" (p. 2). The policy focuses on four broad themes:

- indigenous innovation contributing to economic growth through distinctive science and innovation
- taiao/environment achieving environmental sustainability through iwi and hapū relationships with land and sea
- hauora/health improving health and social wellbeing
- mātauranga exploring indigenous knowledge and science and innovation.

The 'Vision Mātauranga Policy' has become infused across the government's research, science and technology investment programme and as such each of the Challenge holders were asked to address this aspect in their proposal for funding.

Comparing the data from the Twitter and the Web components of the NSCC, we find that there is a surprising amount of variability in the loans used between the two. *Māori* is used in both, and as expected, is highly frequent. *Mātauranga* also figures in both data sets. However, the only other loanwords consistently shared across the various Challenges appear to be place names, such as, *Otago*, *Waikato*, *Kaikoura* and *Aotearoa*, and the content words for tribe *iwi*, and guardianship *kaitiakitanga*. The differences could potentially arise from a change in authorship (Twitter feeds are likely to be produced by different writers to the web component). A second difference comes from the fact that Twitter exhibits a number of hashtags: #moturesearch, #hlkorero, #biohui #kaurirescue. These are either derived from a combination of an English word (*research*, *bio*, *rescue*) and a Māori word (*motu*, *hui*, *kauri*), or from the abbreviation of the Challenge name and a Māori word, such as in the case of #*hlkorero* (hl= healthier lives).

Table 2. A comparison between web and Twitter components in the NSCC with respect to the 20 most frequent Māori loanwords.

Rank	Māori loanword	English	Māori loanword	English
frequency	WEB	equivalent	TWITTER	equivalent
1	MĀORI/ MAORI	(of) indigenous (origin)	MĀORI/ MAORI	(of) indigenous (origin)
2	MĀTAURANGA	traditional knowledge	#MOTURESEARCH	hashtag
3	IWI	tribe	OTAGO	Otago
4	OTAGO	Otago	MĀTAURANGA	
5	HAPŪ	sub-tribe	WAIKATO	Waikato
6	WAIKATO	Waikato	#HLKORERO	hashtag
7	WHENUA	land	KŌRERO	talk/ a talk
8	WHĀNAU	extended family	KAURI	
9	KAIKŌURA	Kaikoura	TAHI	one
10	NGĀTI	tribe	KAIKŌURA	Kaikoura
11	KAITIAKITANGA	guardianship	#BIOHUI	hashtag
12	KAI	food	AOTEAROA	New Zealand
13	KĀHUI	group	IWI	tribe
14	TAMARIKI	children	ORA	health
15	AOTEAROA	New Zealand	MIHI	introduction
16	TIKANGA	custom	KIWIS	New Zealanders, flightless bird
17	TANGAROA	God of the sea and fish	NGAITAHU	the tribe of Ngāi Tahu
18	KAITIAKI	guardian	#KAURIRESCUE	
19	WHAI	to have, to pursue	KAITIAKITANGA	guardianship
20	TIPU	growth	KIWI	native bird, New Zealander

Hybrid expressions comprising an English word and a Māori word are not new to New Zealand English. A study by Onysko and Degani (2010) shows that possibly owing to the popularity of compounding as a strategy for creating new words in English, New Zealand English exhibits a number of productive and recurring compounds formed with English words and frequently occurring Māori loanwords, particularly with the words Māori, kiwi, iwi and Pākehā (Māori wars, Māori world, Māori people, Māori language, kiwi dollar, kiwi team, kiwi company, kiwi ingenuity, iwi leaders, iwi member, iwi organization, iwi consultation, Pākehā world, Pākehā language, Pākehā New Zealanders, Pākehā men, Pākehā women). Twitter provides a platform where such hybrids have the potential of being more common due to the word formation principles exploited by hashtags (which mimic compounding).

As this is the first study of Māori loanwords used in Twitter discourse, it is difficult to ascertain whether hybrid hashtags are innovations of the National Science Challenge websites or whether they are used more frequently in this type of data compared to other types of Twitter discourse. When comparing loanwords in the NSCC website contents and the NSCC associated Twitter feeds, we found a 50% overlap between the most frequently occurring loanwords, and this overlap was mostly made up by place names and proper nouns (5 loans: *Māori, Waikato, Otago, Aotearoa, Kaikoura*), rather than content words (3 loans: *kiwi, matauranga, iwi*).

Macalister (2006) classifies Māori loanwords into four broad categories: proper nouns, flora and fauna, material culture, and social culture terms. The first two classes are self-evident. The latter two involve concepts which are either concrete and "touchable", such as whare 'house', kai 'food', and pā 'settlement', grouped in the material culture set; or more abstract, socially-determined concepts, such as, kinship terms like koroua 'old man', kuia 'old woman, granny', or other abstract notions like mana 'respect, power', and aroha 'love', which come under the label of social culture. As mentioned in the introduction, the use of Māori loanwords is deemed to be increasing in New Zealand (English), cf. Macalister (2006), Kennedy & Yamazaki (1999), Calude et al (in press), with the fastest growing group being the social culture terms, according to Macalister (2006: p. 18). The NSCC also corroborates this finding and we see that despite the fact that five out of the eleven Challenges have an environmental theme and thus raise the expectation that a great majority of the Māori loanwords used might express flora and fauna concepts, it turns out that most loans denote social culture terms, see Table 3. The use of social culture terms suggests that rather than filling a lexical gap related to what is readily detectable in the physical environment, the words which are flowing from Māori into New Zealand English are shifting towards a more cultural perspective, incorporating cultural traditions and concepts which are more specialized and abstract, requiring a longer period of intimate contact between cultures to absorb.

Table 3. The 50 most frequently-occurring loanwords in the National Science Challenge Corpus, by semantic class (excluding hashtags, and the grammatical words *ka*, *o* and *te*)

PROPER NAMES	FLORA & FAUNA	MATERIAL CULTURE	SOCIAL CULTURE
MĀORI	KAURI	WHENUA	MĀTAURANGA
OTAGO	KIWIFRUIT	KAI	IWI
WAIKATO		MOANA	HĀPU
AOTEAROA		WĀNANGA ³	WHĀNAU
KAIKŌURA		MARAE	KAITIAKITANGA
TANGAROA		WAI	KĀHUI
WAIKATO		TAI	TAMARIKI/TAMAITI
TAURANGA		MOTU	TIKANGA
TARANAKI			KAITIAKI
TAHU			WHAI
RAUKAWA			TIPU
IHU			PAINGA
MOTU			REA
			TANGATA
			MAURI
			KAUPAPA
			WĀNANGA
	(MANA
			HUI
			TAONGA
			TAIAO
			MANAAKI
		7	TAU
			HAUORA
			TAHI
			KĀINGA
			ORANGA
			AO
TOTALS	TOTALS	TOTALS	TOTALS
TYPES 13	TYPES 2	TYPES 8	TYPES 28

³ Sometimes *wānanga* refers to a specific institution, and sometimes to a particular type of activity, so it can be classed in two different ways (as material or social culture), depending on context.

Homogeneity across Challenges

The overall frequency list produced by AntConc shows that in comparing the most frequent non-function words (that is, excluding words like *the*, *a*, *and*) in the corpus for each Challenge separately, we find that roughly half of the Challenges exhibit a Māori loanword within the ten most frequent words (the loanword *Māori*); specifically, these are: *A Better Start*, *New Zealand's Biological Heritage*, *Sustainable Seas*, *Aging Well*, and *Building Better Homes*, *Towns and Cities* (which also exhibits the loan *kāinga* 'home').

Focusing on the ten most frequent loanwords in each individual Challenge, we find that there is a great variety of loanword types used, and altogether, we identified a set of 69 distinct loanword types in the eleven Challenges (if there was 100% overlap, we would expect 10 loanwords; with no overlap at all, we would have 110 loanwords). Ranking these types by breadth of occurrence throughout the Challenges, we see that from the 69 types, only 15 loanwords are not unique to a single Challenge, (see Supplementary Materials, Table S4 and Figure S1). In other words, the great majority of loanwords are specific to a particular Challenge.

We can probe the issue of unique loanword use within individual Challenges further by looking at collocations, that is, recurrent groups of loanwords. Because we do not believe that loanwords occur in isolation from other loanwords, we analysed the most frequent loanword collocates of the word *Māori* (a loanword which is present in all of the eleven Challenges), with the help of Lancsbox (Brezina et al., 2015).

Figure 2 (A and B) shows the network plots derived for each Challenge, using MI scores with a threshold of 6 or higher, by considering a span of five words (a threshold of 3 or above is considered statistically significant, but because the language we are dealing with here seems to be rather formulaic in places, a higher threshold was used). The lines in the graphs are proportional to the strength of collocation (shorter lines correlate with stronger collocation). The network plots indicate that three Challenges (*The Deep South, Healthier Lives* and *High-Value Nutrition*) do not contain unique loanwords within the top ten most frequently-occurring collocates for the word *Māori*. For some few Challenges, the uniquely-identifying loanwords are topically predictable: *kāinga* ('home') for *Building Better Homes, Towns and Cities, tirohanga* ('view/sight') for *NZ Biological Heritage*; but for a significant portion of the Challenges, this is not the case (it is unclear why *ao* 'world' and *Aotearoa* 'New Zealand' should be unique collocates for

Māori in the *Science for Technological Innovation Challenge*, or why *Waitangi* should be a unique collocate for *A Better Start*). The fact that the loanwords identified are not topically related to the theme of the Challenges in which they occur points to individual stylistic preferences of the authors who produce the online discourse included for each website. In other words, the loans have stronger ties to the individual who uses them, rather than to the topic of discourse within which they appear.

The network plots also show that in some Challenges, the word *Māori* acts as a centre of discourse gravity, attracting many distinct collocates, for instance, *Sustainable Seas*, *NZ Biological Heritage* and *Our Land and Water*; while for other Challenges, the word *Māori* has a more limited scope, i.e., *Building Better Homes, Towns and Cities* and *Science and Technological Innovation*.

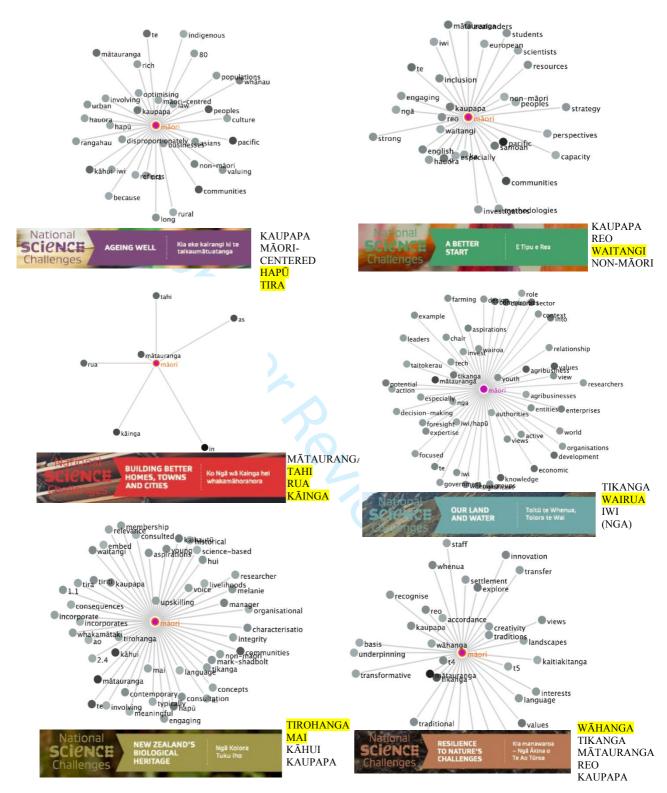


Figure 2A. Network graphs of the most frequency collocates for the loanword Māori in the eleven National Science Challenge websites (MI score, span or 5<>5, threshold=6).

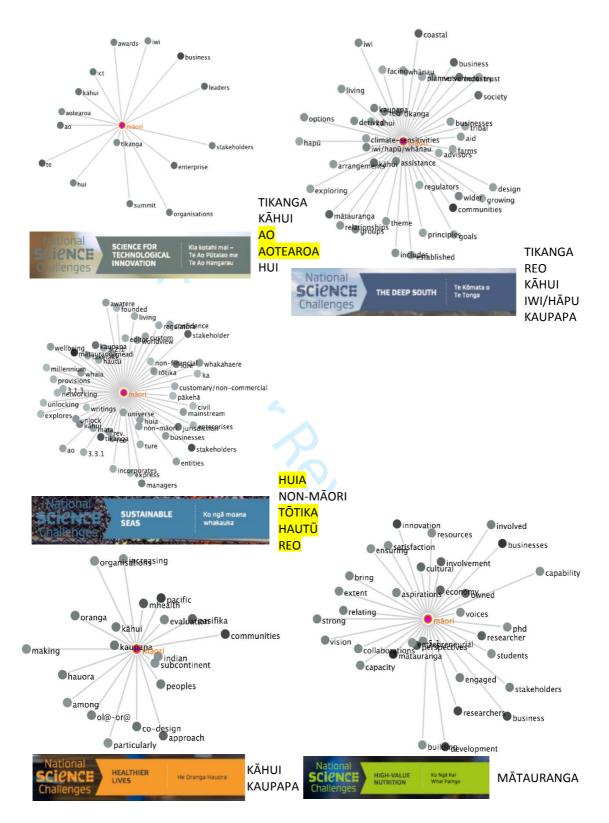


Figure 2B. Network graphs of the most frequency collocates for the loanword Māori in the eleven National Science Challenge websites (MI score, span or 5<>5, threshold=6).

Comparisons with other genres of New Zealand English

Given the unexpected presence of Māori loanwords which do not denote concepts other than flora and fauna in the National Science Challenge Corpus, it might be tempting to assume that the use of Māori loanwords is motivated by the fact that New Zealand English exhibits a steady increase of use of Māori loanwords more generally, and that perhaps these words make up a body of lexical items which are entrenched in typical New Zealand English usage. Put another way, topic is not the motivator for the use of these words, rather, their use is motivated by the variety of English itself. To test this hypothesis, we analysed keyness by comparing various other corpora of New Zealand English currently available.

Keyness was developed in the mid-1990s as part of a toolkit incorporated in a widely used piece of corpus linguistics software, Wordsmith Tools (Scott 1996). The author of Wordsmith Tools, Mike Scott defined his coined new term 'key word' as "a word which occurs with unusual frequency in a given text [...] by comparison with a reference corpus of some kind" (Scott, 1997, p. 236, as cited in Gabrielatos, 2018, p. 225). As the task of identifying a text's topic remains a notoriously difficult problem with low consensus in the linguistics community, linguists have looked to more concrete concepts, such as keyness, to get at the heart of a given portion of discourse, somewhat awkwardly referred to as the "aboutness" of a text. Having a quantitatively defined method for probing "aboutness" allows for comparisons to be made between texts, corpora, authors, and genres. We believe that keyness could be a useful notion in the study of Māori loanwords because previous work on the subject has pointed to a non-homogenous and highly skewed use of Māori loanwords in New Zealand English (de Bres, 2006; Degani, 2010; Calude et al., 2018).

First, we compared the National Science Challenge Corpus against the older, written Wellington Corpus of New Zealand English (Bauer, 1993). AntConc was used to extract key loanwords between the corpora in order to see which loanwords might be used with higher or lower than expected frequency in the NSCC. As above, keyness measures are calculated by comparing a corpus of interest (in our case, the NSCC), with a reference corpus (the Wellington Corpus), and measuring Log Ratio Odds (Hardie, 2014), comparing expected rates arising from the reference corpus, with the observed rates from the NSCC. Here, we are interested in loanwords which occur with higher rates than expected (rather than the lower rates) because we want to know which loanwords appear

more frequently in the NSCC compared to the Wellington Corpus (the working assumption is that the Wellington corpus will contain many more loanword types just by virtue of the wider topic of discussion). These loanwords are given in Table 4, together with their keyness rank, raw frequency of occurrence in the National Science Challenge Corpus, keyness score and log ratio scores. Keyness ranks provide an ordering of the key words identified in order of importance (with the first ranked word having the highest keyness score). Raw frequency counts are included because they give an impression of how important the word in question might be in the corpus of interest. Keyness scores are calculated using the default AntConc Log likelihood procedure. Finally, log ratio scores provide a measure of the effect size, such that with every increase in log ratio odds, there is an observed doubling of rate of occurrence in the corpus of interest compared to the reference corpus (see Hardie, 2014; and Gabrielatos, 2017 for more details). The table shows for instance that the loanword Māori has a log ratio of 2.1783 which means that it was used (just over) 4 times more frequently in the National Science Challenge data than it was in the Wellington Written Corpus (even after normalizing). Similarly, Mātauranga has a log ratio of 8.7109, which means it occurred in the National Science Challenge roughly 265 times more frequently than in the Wellington Written Corpus (interestingly, *Mātauranga* occurs 3 times in the Wellington Written Corpus).

Table 4. Loanwords which are significantly more frequent in the NSCC compared to the Wellington Corpus of Written NZE. The table gives the Māori loanwords in the top 500 keywords words, in the order of keyness, excluding place-names and function words (all macrons were removed from the corpora for this analysis)

Loanword	Wellington Written Corpus (Holmes et al 1998) score			
	Rank	Raw Freq (NSCC)	Keyness score	Log Ratio
MĀORI	11	5591	2521.07	2.1783
MĀTAURANGA	22	1659	1833.35	8.7109
IWI	36	1334	1429.54	7.1739
HAPU	134	579	604.08	6.4552
WHĀNAU	223	380	407.06	7.1696
KĀHUI	229	356	401.62	9.0755
KAITIAKITANGA	253	326	367.77	8.9485
TAMARIKI	269	303	341.82	8.8429
TIKANGA	322	258	291.05	8.611
WHENUA	329	377	284	3.5732
KAITIAKI	375	221	249.31	8.3876
KAI	391	319	237.59	3.5248

PAINGA	408	201	226.75	8.2508
REA	409	201	226.75	8.2508
TIPU	422	203	218.05	7.2651
WHAI	439	214	208.23	5.3412
NGATI	453	381	199.4	2.4729
KAUPAPA	474	169	190.65	8.0006
MOANA	484	189	187.35	5.577
MAURI	488	193	185.36	5.1922

Next, we repeated the process with diachronic corpora compiled on the basis of topic selection, the first was a set of newspaper articles on the topic of Matariki /Māori New Year (Calude et al., forthcoming), and the second, a corpus of newspaper articles on the topic of Te wiki o te reo Māori/ Māori Language Week (Levendis & Calude, 2018, 2019). Both of these corpora comprise articles over a period of 10 years, from 2007-2018. The corpora were precisely put together to fruitfully inform the study of loanwords over that period, given the association of Māori loanwords with Māori topics. We reasoned that loanwords which are significantly more frequent in the NSCC compared to these corpora would be noteworthy, given the already high expectation of frequent loanword use in these topically-constrained written corpora. The keyness analysis was repeated in turn with each corpus and the results are given in Tables 5A and 5B.

Table 5A. Loanwords which are significantly more frequent in the NSCC compared to the *Matariki Corpus* (all loanwords, excluding place-names and function words)

Loanword	Matariki Corpus (Calude et al in press)				
	Rank	Raw Freq (NSCC)	Keyness score	Log Ratio	
MĀTAURANGA	43	1659	165.62	4.005	
HAPŪ	281	579	45.93	2.9718	
IWI	297	1334	44.04	1.4104	
KAITIAKITANGA	306	326	42.74	5.465	
TIPU	476	203	26.62	4.7816	
KĀHUI	514	356	24.78	2.592	

Table 5B. Loanwords which are significantly more frequent in the NSCC compared the *Māori Language Week Corpus*.

Loanword	Māori Language Week Corpus (Calude and Levendis in prep)				
	Rank Raw Freq Keyness score Log Ratio (NSCC)				
MĀTAURANGA	20	1659	158.73	7.3442	
KAITIAKITANGA	284	326	31.18	4.9969	

One final observation concerns loanwords which were found to be especially infrequent in the NSCC, compared to the Matariki and the Māori Language Week Corpora. Given their Māori focus and wider authorship breadth, the latter corpora are expected to contain many more loanwords that are absent from the NSCC data. However, the loanwords Pākehā and Kiwi were both conspicuously absent from the NSCC which caused some surprise given that the National Science Challenges are meant to represent both Māori and Pākehā interests and concerns. It appears that the word Pākehā is becoming widely replaced by the word non-Māori, a use which has previously not featured on lists compiled to document the various lexical ways of referring to New Zealanders of European descent (Daly, 2017, p. 4). It remains to be seen to what extent *non-Māori* overlaps with the semantic content denoted by the term *Pākehā*. The examples below indicate that *non-Māori* is sometimes used to refer to New Zealanders who are not Māori (Pākehā but also New Zealanders of other nationalities living in New Zealand), as in example (1), but at other times, *non-Māori* represents a semantic opposition to the term New Zealand European, thereby implying a closer correspondence to Pākehā, see example (2).

- (1) Professor Kerse's research areas include promoting activity and function in residential care, residential care organisational culture and outcomes, promoting physical activity in community dwelling older people, activity for depression in the very old, staying upright (preventing falls and injury) in older people in all settings, improving prescribing in primary care, and a large cohort of Māori and **non-Māori** in advanced age. (from the website of the *Aging Well* NSC).
- (2) The obvious Vision Mātauranga theme that the Tūhonohono project comes under is "Taiao: Achieving Environmental Sustainability through Iwi and Hapū relationships with land and sea." Tūhonohono is about Māori and **non-Māori** aspiring to live in sustainable communities dwelling in healthy marine and coastal area environments. Distinctive environmental research arising in Māori communities relates to the expression of iwi and hapū tikanga and mātauranga knowledge, culture and experience over the takutai moana area such as rahui, kaitiakitanga, mana whenua, mana moana, and matāwhanga in the sustainable use and restoration of the marine environment. We will identify these concepts 'in practice' through approaches that are based on whakapapa and 'place based'

values and perspectives which will unlock the science and innovation potential of tikanga and mātauranga Māori knowledge, resources and peoples over the takutai moana area which should lead to building synergistic environmental sustainability by combining the best of both founding cultures – Māori and **New Zealand European** – legal systems, philosophies, values, rules and institutions.

(from the website of the Sustainable Seas NSC)

In so far as the term *non-Māori* refers to Pākehā (or New Zealanders of European descent), its morphological derivation from the word *Māori* has the symbolic effect of construing the distribution of populations in New Zealand as either Māori or non-Māori thereby, legitimizing the historic perspective of the indigenous people of the country (Māori becomes the 'default' group).

Case-study: the use of Mātauranga

Given the pervasiveness and importance of the loanword *Mātauranga* and the importance of the Vision Mātauranga Policy, we investigated its use in the National Science Challenge Corpus more closely. The total number of 1652 uses of the loanword across the corpus is unevenly spread between the Challenges, with some using it more than others, and two only using it in passing (*Building Better Homes, Towns and Cities* and *Healthier Lives*). The loanword is spelt both with a macron (*Mātauranga* n=1610) and without a macron (*Matauranga* n=42), and interestingly, most Challenges include both spellings (notable exceptions being the two Challenges which do not use the word extensively, *Building Better Homes, Towns and Cities, Healthier Lives*, along with *Aging Well*).

Figure 3 shows the overall use of *Mātauranga* in each Challenge, normalized per 10,000 words (see Supplementary Materials, Figure S2 for raw numbers of occurrence in each Challenge). This plot shows that the highest use of the loanword *Mātauranga* comes from the following Challenges: *Science for Technological Innovation*, *Sustainable Seas* and *NZ's Biological Heritage*.

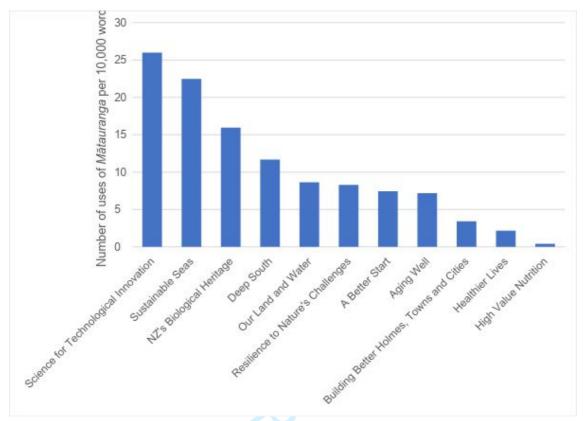


Figure 3. Frequency of the use of *Mātauranga* in each Challenge (normalized per 10,000 words)

Next, we looked at how the loanword *Mātauranga* is used in the text and to this end, we extracted its most frequent collocates by searching neighbouring words (occurring both, to the left and the right of it).⁴ Perhaps unsurprisingly, by far the strongest collocate was the word *vision* (*Vision Mātauranga*). Grouping collocates by word-class, we found that *Mātauranga* occurred with the nouns *Māori*, *programme(s)*, *policy*, *science*, *theme(s)*, *knowledge*, *principles*, *research*, *approaches*, *alignment*, *framework*, *project(s)*, *workshops*, *objectives*, *strategy*, *clarification*, *capability* and *leader*. Most of these occurred to the right of the loanword, but the nouns *science* and *knowledge* also occurred to the left of it (i.e., *science Mātauranga* and *Mātauranga science*). Looking at verbs which occurred with *Mātauranga*, we found the verbs: *reframe*, *employ*, *incorporate*, *examine*, *integrate*, *explored*, *need*, *include* occurring to the left of the loanword, and the verb *address(es)* occurring to its right (in addition to the most generic *is/are*, *can*, *will*). Widening the context of collocation illuminated the following frequent

_

⁴ We ignored the use of the macron for this analysis.

phrasal units: Mātauranga Māori and, Mātauranga and tikanga (Māori), Mātauranga Māori and science, Mātauranga Māori research, Mātauranga Māori and kaitiakitanga, principles of Vision Mātauranga, and systems derived from Mātauranga. Studying collocational patterns hints at the ways in which writers construe the concept of Mātauranga and how they refer to it, namely, it is a type of science, knowledge, principle, research, approach and framework, which is to be reframed, employed, incorporated, examined, integrated, included, and aligned to.

Conclusion

In conclusion, we found a strong presence of Māori borrowings in National Science Challenge web and Twitter feeds which may be reflective of the prominence given to the Vision Mātauranga Policy across the government's research, science and technology investment programme. Using corpus linguistics methods, we highlight the growing use of Māori loanwords as a national identity building tool in this genre, in addition to being one of the many mechanisms to signal a growing cultural awareness of Māori traditions and concepts and the many 'challenges' that the country is currently facing.

Acknowledgements

(Removed for anonymity purposes.)

References

- Anthony L 2018. AntConc (Version 3.5.7) [Computer Software] Tokyo, Japan: Japan Waseda University. Accessible from: http://www.laurenceanthony.net/
- Anthony, L., & Hardaker, C. (2016). FireAnt (Version 1.0) [Computer Software] Tokyo, Japan: Japan Waseda University. Accessible from: http://www.laurenceanthony.net/
- Bauer, L. (1993). Manual to accompany the Wellington Corpus of Written New Zealand English. Department of Linguistics, Victoria University of Wellington, New Zealand.

- 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., Harper, S., Whaanga, H., & Miller, S. (forthcoming) Loanwords as Textual Collocates: Investigating Māori Loanwords related to Matariki in New Zealand English newspapers. *Asia-Pacific Language Variation Journal*
- Calude, A., Pagel, M., & Miller, S. (2018). Modelling Borrowing Success a quantitative study of Maori loanwords. *Journal of Theoretical Linguistics and Corpus Linguistics*, 15, 2-38. https://doi.org/10.1515/cllt-2017-0010
- Daellenbach, U., Davenport, S., & Ruckstuhl, K. (2017). Developing absorptive capacity for midstream science in open innovation contexts. *International Journal of Technology Transfer and Commercialisation*, *15*(4), 447–462. https://doi.org/10.1504/IJTTC.2017.089674
- Davies, C., & Maclagan, M. (2006). Māori Words Read all about it: testing the presence of 13 Māori words in four New Zealand newspapers from 1997 to 2004. *Te Reo*, 49, 73–99.
- Daly, N. (2017). Pākehā–Māori, European–Native: Ethnic Labeling in the Dorothy Neal White Collection. *New Review of Children's Literature and Librarianship, 23*(1), 1–12. https://doi.org/10.1080/13614541.2017.1280335
- 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 Facchinetti, R., Crystal, D., & Seidlhofer, B. (Eds.), *From International to Local English and Back Again*. (pp. 165–196). Bern, Switzerland: Peter Lang
- Degani, M., Onykso, A. (2010). Hybrid Compounding in New Zealand English. *World Englishes*, 29(2), 209–233
- Deverson, T. (1991). New Zealand English Lexis: the Māori Dimension. *English Today* 28, 18–25
- Gabrielatos, C. (2018). Keyness Analysis: Nature, Metrics and Techniques. In Taylor, C., & Marchi, A. (Eds.), *Corpus Approaches to Discourse: A Critical Review.* (pp. 225–258.). Oxon, England: Routledge

- Hardie, A. (2014). Log Ratio an Informal Introduction. Post on the website of the ERSC Centre for Corpus Approaches to Social Science CASS. Accessible from http://cass.lancs.as.uk/?p=1133 [Accessed 8 January 2019].
- Harlow, R. (2005). Covert Attitudes to Māori. *International Journal of the Sociology of Language 172*, 133–147
- Hikuroa, D. (2017). Mātauranga Māori—the ūkaipō of knowledge in New Zealand. *Journal of the Royal Society of New Zealand*, 47(1), 5–10
- Holmes, J., & Bell, A. (1992). On shear markers and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English. *Language Variation and Change*, 4(3), 251–271
- Kennedy, G., Yamazaki, S. (1999). The influence of Māori on the New Zealand English lexicon. In Kirk, J. (Ed.), *Corpora Galore: Analyses and Techniques in Describing English*. (pp. 33–44). Amsterdam, The Netherlands: Rodopi
- Kingi, T.K., Durie, M., Elder, H., Tapsell, R., Lawrence, M., & Bennett, S. (2018). *Maea te Toi Ora: Māori Health Transformations*. Wellington, New Zealand: Huia
- Levendis, K., & Calude, A. (forthcoming 2019). Flagging loanwords and what they can tell us a case study from New Zealand English. *Ampersand* (Special Issue on Loanwords and Identity)
- Levendis, K., & Calude, A. (2018, November). 'When you don't know what you know': Use of Māori loanwords in a diachronic corpus of New Zealand English. Paper presented at the New Zealand Linguistics Society Conference, Victoria University of Wellington, New Zealand
- Macalister, J. (2008). Tracking changes in familiarity with borrowings from te reo Māori. *Te Reo 51*, 75–97
- Macalister, J. (2006). The Maori presence in the New Zealand English lexicon, 1850–2000: Evidence from a corpus-based study. *English World-Wide* 27, 1–24
- Maclagan, M., & Gordon, E. (1996). Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change* 8(1), 125–147
- Māori Language Commission, (2018). *Te Taura Whiri I te Reo Māori*. http://www.tetaurawhiri.govt.nz/ [accessed January 2019]
- Mead, H.M. (2012). "Understanding Mātauranga Māori Conversations on Mātauranga Māori". In Bean, D., Black, T., Collings, W., & Nuku, W. (Eds). *Conversations*

- on Mātauranga Māori. (9–14). Wellington, New Zealand: New Zealand Qualifications Authority
- Mercier, O.R. (2013). Indigenous knowledge and science. A new representation of the interface between indigenous and eurocentric ways of knowing. *He Pukenga Korero*, 8(2)
- Ministry of Research Science and Technology. (2007). Vision Mātauranga: Unlocking the innovation potential of Māori knowledge, resources and people. Wellington, New Zealand: New Zealand Ministry of Research Science and Technology (MoRST)
- McEnery, T. & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge, England: Cambridge University Press
- Royal, T.A.C. (2009). *Mātauranga Māori*. Porirua, Wellington: Mauriora-ki-te-Ao/Living Universe Ltd
- Royal, T.A.C. (2012). Politics and knowledge: Kaupapa Māori and mātauranga Maori. New Zealand Journal of Educational Studies, 47(2), 30–37
- RNZ (Radio New Zealand) (2017). *Why RNZ is dedicated to the use of te reo*. Published 1 December 2017, https://www.radionz.co.nz/news/on-the-inside/345170/why-rnz-is-dedicated-to-the-use-of-te-reo [accessed January 2019]
- Schneider, E. (2003). The Dynamics of New Englishes: from Identity Construction to Dialect Birth. *Language* 79(2), 233–281
- Smith, L.T. (2012). *Decolonising methodologies: Research and Indigenous peoples*. London, UK: Zed Books
- Smith, L.T., Maxwell, T.K., Puke, H., & Temara, P. (2016). Indigenous knowledge, methodology and mayhem: What is the role of methodology in producing indigenous insights? A discussion from Mātauranga Maori. *Knowledge Cultures*, 4(3), 131–156

The Use of Māori Words in National Science Challenge Online Discourse

Andreea Calude^{a*}, Louise Stevenson^b, Hēmi Whaanga^c and Te Taka Keegan^d

^a Department of General and Applied Linguistics, University of Waikato, Hamilton, New Zealand, email: andreea.calude@waikato.ac.nz;

^b Department of General and Applied Linguistics, University of Waikato, Hamilton, New Zealand;

^c School of Māori and Indigenous Studies, University of Waikato, Hamilton, New Zealand;

d School of Mathematical and Computing Sciences, University of Waikato, Hamilton, New Zealand.

The Use of Māori Words in National Science Challenge Online Discourse

New Zealand English is well-known for its heavy borrowing of words from Māori. This lexical component, unique to New Zealand English alone has been studied intensely over the last 50 years, particularly in newspaper media. Current research suggests the use is still increasing today, primarily in Māori-related contexts. Here, we analyse a surprising and unexpected use of Māori loanwords in science digital discourse (neither genre being previously investigated), where we discover a strong presence of Māori borrowings in National Science Challenge website and Twitter content. Using corpus linguistics methods, we argue that the use of Māori loanwords in this genre functions as a national identity building tool, used by various authors to signal that the 'challenges' the country faces are uniquely New Zealand's 'challenges'.

Keywords: Māori loanwords, New Zealand English, National Science Challenge, corpus linguistics

Introduction – The National Science Challenges

In 2012, the New Zealand Government announced a new approach to funding public science by seeking input to decide the key scientific questions that New Zealand (NZ) needed to answer.¹ A number of strategic areas of scientific focus were developed which led to what became known as the National Science Challenges (NSCs). The NSCs are a set of 11 over-arching projects that concentrate NZ's scientific efforts on the major issues and opportunities facing the country. The NSCs have been allocated ten years of funding over two five-year periods. A review was undertaken after the first five-year period, to establish whether the NSCs were on track and warranted further funding. A summary of the 11 NSCs is given in Table S.1. (see Supplementary Materials).

¹ Source: https://www.beehive.govt.nz/release/govt-enlist-public-nz-science-challenges

A key difference in the NSC approach compared to other funding opportunities was that the alliances formed to address the mission-led goals resulted in scientists connecting and collaborating across a range of institutions, including universities and Crown Research Institutes (Daellenbach, et al. 2017). Also, as the second tranche of funding was reliant on a positive outcome of the first five years of research, the NSCs were encouraged to disseminate their missions and intentions, their activities and their outputs. The most common vehicle for this dissemination was not only the NSC websites, but also their Twitter feeds, a unique and current discourse of the science issues currently facing New Zealand and the research being undertaken to address these issues. This paper is concerned with analysing that language.

Background to the language contact situation – why study loanwords?

It has long been argued that the most distinctive and salient feature of New Zealand English is its widespread use of words adopted from Te Reo Māori (Calude et al 2018, Deverson, 1991Degani 2010, Deverson, 1991, Kennedy & Yamazaki 1999, Macalister 2006). Not only has (what eventually become known as) New Zealand English been actively and productively adopting Māori words ever since the first contact with British English took place, but their use is still increasing today in both frequency of existing adopted loanwords, and with regard to the import of new words from Māori (Macalister, 2006; Kennedy & Yamazaki, 1999; Calude et al., forthcoming).

This increase is not a straight forward matter because research shows that the use of Māori loanwords² (or borrowings) is not homogenous across people; with some people, most notably, Māori women (see Calude et al., 2018)—, useing more loanwords than others.; nNor is their use of loanwords homogenous across topics of discourse; with some topics attracting more loanwords than others, (specifically, topics pertaining to Māori people or Māori culture, (-cf. de Bres, 2006; Degani, 2010; Macalister, 2006; Davies & Maclagen, 2006). The use of loanwords has been interpreted by some as reflecting the identity of speakers (or writers), for instance by Davies & Maclagen, 2006 and Macalister (2007), but it might also be reflecting the speaker's attuning of their language to their audience (see Audience Design Theory, Bell 1984, Coupland 2007). Interpreting the motivation and consequences of the use of Māori loans is similarly

² The term loanword (or borrowing) is used here to denote the productive (reucrrent) use of words from a donor language (in our case, Māori) in a receving language (English).

complex, with opinions remaining divided between (a) borrowings as indicators of positive attitudes towards Māori language and culture (Māori Language Commission), and (b) borrowings as perpetuators of negative stereotypes about the Māori world (Degani, 2010; Harlow, 2005).

Within the New Zealand context, the study of loanwords is important because it highlights the newly emerging voice that New Zealanders are choosing to represent their identity (or at least, part of it, for those who choose to use Māori borrowings). As the new dialect of English is coming into its final stage of separation from other English varieties (Schneider, 2003), the time is ripe to consider the linguistic ways in which New Zealanders express distinctiveness, particularly in contrast to British English and Australian English, our closest related linguistic 'relatives'.

A second reason why the study of loanwords provides a fruitful endeavour relates to the fact that unlike other linguistic features (for instance, vowel mergers in pairs such as *air* and *ear*, see Maclagan & Gordon, 1996, Holmes & Bell, 1992 inter alia), Māori words and their use in New Zealand English represent both a salient, and a conscious, deliberate move. This move is embedded in a political and social perspective that strays from a neutral position. Evidence of the awareness and strength of such a linguistic move can be seen in strong comments from high-profile New Zealand figures found in the current New Zealand media, ranging from defiant and unwavering promotion of the use of loanwords (see quote 1 from Guyon Espiner), to strong opposition to their use (see quote 2 by Don Brash, former National Party leader).

(1) Quote from Guyon Espiner in favour of using Māori loanwords

"Armed with my tenuous grasp on the language I began to weave it into my radio presenting as co-host of RNZ's Morning Report. We've always done basic greetings in Māori on the show but I wanted to do more and began to extend the greetings and include basic information in Māori – such as the days, dates and temperatures for the main centres."

[source: My te reo Māori journey, 13/09/2017, The Spinoff]

(2) Quote from Don Brash against using Māori loanwords

"It's tokenism of the worst kind," he [Don Brash] says. Don Brash is continuing his attacks on the use of Te Reo on Radio New Zealand. The former National Party leader

appeared on The AM Show on Wednesday, after saying he was "utterly sick" of the use of Aotearoa's indigenous language in broadcasting.

[source: "Don Brash's scathing statement on Te Reo usage", 29/11/2017, Newshub]

Thirdly, the situation of language contact in New Zealand provides a unique window into loanwords more generally, which is of interest to language researchers outside New Zealand. Within the wider international context, studying loanwords is itself not a new area of inquiry, nor is it linguistically diverse: studies often analyse English words which are borrowed into other languages, such as, French, German, Brazilian Portuguese, and so on. Recently, a new methodological paradigm shift has breathed new light into this research topic (see a literature review summary of novel approaches to the phenomenon of "loanword success" in Calude et al., 2008). But the study of Māori loanwords specifically has drawn interest outside the New Zealand linguistic community because of the unique sociolinguistic context in which these words are found. Here, we have a world dominant language (English) adopting words from a non-dominant language which is (still) fighting for its survival (Te Reo Māori). Not only is this language contact scenario distinct from all others studied so far (to our knowledge), but counter to many other language contact scenarios (such as, those mentioned above), the adoption of Māori loanwords in New Zealand English is actively encouraged by the Māori Language Commission, as well as many other groups (see for instance, a recent piece about Radio NZ's choice to increase the use of Māori words throughout their English language broadcasts by RNZ chief executive Paul Thompson, RNZ 2017).

Data and Methods – National Science Challenge as a coherent discourse

Given the observations made by previous researchers in the area of Māori loanwords, we were interested to find out to what extent words of Māori origin were used in New Zealand English science discourse, and online sources – neither of which had previously received any attention. In this section, we outline the methods used to gather and code our data.

Over the last thirty years, the field of linguistics has experienced a dramatic paradigm shift towards a more empirical methodology, which rather than relying on internal introspection alone (and without backing from real data), is heavily based on

collecting real language exchanges, be they spoken or written (see a review in McEnery & Hardie, 2011). While corpus studies still utilise a great deal of introspective thinking, for example, in defining the topic of research, and in designating research questions, the analyses ultimately rest on concrete and real occurring examples, which have the added advantage of being falsifiable. That said, the corpus linguistics analysis remains driven by the researcher and the interpretation of the findings is done within the researcher's chosen theoretical framework.

We adopt a corpus linguistics framework here to analyse the use of Māori loanwords in the *National Science Challenge* online presence. To this end, we built a corpus termed the National Science Challenge Corpus (NSCC) in late 2017. The NSCC is made up of two types of datasets. The first and largest dataset comes from the internet websites of the eleven National Science Challenges, which we gathered manually by copy-and-paste. The text came in different formats, usually in HTML code or as part of PDF documents and these were converted to plain text files, keeping the material from each Challenge separate. All graphical information was ignored. The websites appeared to share certain structural features, such as, (1) a paragraph summarising their main research questions and agenda, (2) a main content page detailing ongoing and current projects, and (3) other supplementary materials included as PDF file attachments (for some Challenges only).

The second dataset consisted of a body of tweets which were identified by following the hashtags of each Challenge. Ten from the eleven Challenges have active Twitter feeds and we used the open source software FireAnt (Anthony & Hardaker, 2016) to collect all the text included in these Twitter feeds. Regular expressions were used to 'clean up' the data and discard URLS and any other non-textual material (graphics). These materials gave us a corpus of approximately 1.5 million words (see Supplementary Materials, Table S.2. for details).

Once the corpus was collated, we used the open source concordancing software AntConc (Anthony, 2018) to extract the most frequently occurring loanwords in it and to perform keyness analyses, discussed in the following sections.

Results

This section is divided into three parts; the first deals with the general use of loanwords in the NSCC, reporting which loanwords appear in the NSCC and differences between

the Twitter data and the website data. The second section considers the spread of loanwords across the eleven Challenges. The third section provides comparisons between the NSCC and other genres of New Zealand English with respect to loanwords using keyness measures. We conclude with a note on the use of the loanword *mātauranga*, which constitutes a novel importation from Māori and whose use in English is specifically linked to science discourse.

Use of loanwords in the NSCC

The NSCC comprises hundreds of Māori loanwords. Because there is (currently) no way to automate the detection of loanwords, manual inspection is required to identify exhaustive lists of all Māori loanwords used in a given body of data.³ Owing to the size of our corpus, such a task would have been extremely time-consuming and error-prone. We opted, instead, for using AntConc to build word frequency lists which allowed us to extract the *most frequently* occurring loanwords. We did this for the entire NSCC and for each Challenge separately, in turn. Let us first consider the Māori loanwords in the entire corpus. The twenty most frequent loanwords are given in Table 1 (a normalised plot of this data is provided in Supplementary Materials, Figure S1).

Table 1. The 20 most frequent Māori loanwords in the entire NSCC corpus and their associated raw frequencies of occurrence

Rank frequency	Māori loanword	English equivalent	Raw frequency
1	MĀORI/ MAORI	(of) indigenous (origin)	5318 (with macron) +414 (without macron) =5732
2	MĀTAURANGA	traditional knowledge	1640
3	IWI	tribe	1348
4	OTAGO ⁴	Otago	904
5	HAPŪ	sub-tribe	547

³ It is theoretically possible to use machine learning methods to check individual words from a given body of text (say a corpus such as our data) against online Māori and English language dictionaries (both of which are available) in order to extract lists of Māori loanwords. However, unfortunately this method quickly runs into trouble because a great majority of loanwords from Māori are proper nouns (personal names, names of organisations and place names) which do not appear in the dictionary. We believe that the problem of extracting and labelling Māori loanwords might necessitate a machine learning solution and leave this for future research to tackle.

⁴ The native Māori word had the form "Ōtākou", which was later changed to "Otago" to fit with Anglicized spelling and phonology. Despite this change, we note the Māori origin of the word and count it as a loanword here.

6	WAIKATO	Waikato	409
7	WHENUA	land	382
8	WHĀNAU	extended family	340
9	KAIKŌURA	Kaikoura	336
10	KAITIAKITANGA	guardianship	335
11	NGĀTI	prefix to tribal name	334
12	KAI	food	324
14	KĀHUI	group	321
14	AOTEAROA	New Zealand	312
15	TAMARIKI	children	304
16	TIKANGA	custom	259
17	TANGAROA	God of the sea and fish	253
18	KAITIAKI	guardian	225
19	WHAI	to have, to pursue	214
20	ORA	health, to be alive	204

Unsurprisingly, the word *Māori* is itself by far the most widely used loanword. The next most frequently used loan in this data is a relatively new lexical borrowing, namely *mātauranga* 'knowledge, wisdom'. No other New Zealand English corpus analysed so far appears to exhibit this loanword. It is noteworthy that the use of the loanword has gained momentum in New Zealand English and its use has come to prominence in Māori education and research circles in recent times (Mead, 2012; Royal, 2012; Smith et al., 2016). As noted by Hirini Moko Mead (2012, p. 9):

It was not so long ago that the term 'mātauranga Māori' was rarely ever mentioned in education circles Over the last two decades the term 'mātauranga Māori' has become increasingly important as more and more people are engaged in efforts to understand what it means. Put simply, the term refers to Māori knowledge.

Recent research and writing has highlighted the growing awareness and importance of mātauranga Māori (Awatere, 2008; Hikuroa, 2017; Mercier, 2013; Mika, 2012; Royal, 2009; Smith, 2012; Smith et al., 2016). Moreover, at a national research scale, mātauranga Māori has been framed by the 'Vision Mātauranga Policy' (MoRST, 2007), whose mission is to "unlock the innovation potential of Māori knowledge, resources and people to assist New Zealand to create a better future" (p. 2). The policy focuses on four broad themes:

- indigenous innovation contributing to economic growth through distinctive science and innovation
- taiao/environment achieving environmental sustainability through iwi and hapū relationships with land and sea
- hauora/health improving health and social wellbeing
- mātauranga exploring indigenous knowledge and science and innovation.

The 'Vision Mātauranga Policy' has become infused across the government's research, science and technology investment programme and as such each of the Challenge holders were asked to address this aspect in their proposal for funding.

Comparing the data from the Twitter and the Web components of the NSCC, we find that there is a surprising amount of variability in the loans used between the two. *Māori* is used in both, and as expected, is highly frequent. *Mātauranga* also figures in both data sets. However, the only other loanwords consistently shared across the various Challenges appear to be place names, such as, *Otago*, *Waikato*, *Kaikoura* and *Aotearoa*, and the content words for tribe *iwi*, and guardianship *kaitiakitanga*. The differences could potentially arise from a change in authorship (Twitter feeds are likely to be produced by different writers to the web component). A second difference comes from the fact that Twitter exhibits a number of hashtags: #moturesearch, #hlkorero, #biohui #kaurirescue. These are either derived from a combination of an English word (*research*, *bio*, *rescue*) and a Māori word (*motu*, *hui*, *kauri*), or from the abbreviation of the Challenge name and a Māori word, such as in the case of #hlkorero (hl= healthier lives).

Table 2. A comparison between web and Twitter components in the NSCC with respect to the 20 most frequent Māori loanwords.

Rank	Māori loanword	English	Māori loanword	English
frequency	WEB	equivalent	TWITTER	equivalent
1	MĀORI/ MAORI	(of) indigenous	MĀORI/ MAORI	(of) indigenous
1	WAOKI/ WAOKI	(origin)	WAORI/ WAORI	(origin)
2	MĀTAURANGA	traditional	#MOTURESEARCH	hashtag
2	WATAUKANGA	knowledge	#WOTURESEARCH	nasmag
3	IWI	tribe	OTAGO	Otago
4	OTAGO	Otago	MĀTAURANGA	
5	HAPŪ	sub-tribe	WAIKATO	Waikato
6	WAIKATO	Waikato	#HLKORERO	hashtag
7	WHENUA	land	KŌRERO	talk/ a talk
8	WHĀNAU	extended family	KAURI	

9	KAIKŌURA	Kaikoura	TAHI	one
10	NGĀTI	tribe	KAIKŌURA	Kaikoura
11	KAITIAKITANGA	guardianship	#BIOHUI	hashtag
12	KAI	food	AOTEAROA	New Zealand
13	KĀHUI	group	IWI	tribe
14	TAMARIKI	children	ORA	health
15	AOTEAROA	New Zealand	MIHI	introduction
16	TIKANGA	custom	KIWIS	New Zealanders, flightless bird
17	TANGAROA	God of the sea and fish	NGAITAHU	the tribe of Ngāi Tahu
18	KAITIAKI	guardian	#KAURIRESCUE	
19	WHAI	to have, to pursue	KAITIAKITANGA	guardianship
20	TIPU	growth	KIWI	native bird, New Zealander

Hybrid expressions comprising an English word and a Māori word are not new to New Zealand English. A study by Onysko and Degani (2010) shows that possibly owing to the popularity of compounding as a strategy for creating new words in English, New Zealand English exhibits a number of productive and recurring compounds formed with English words and frequently occurring Māori loanwords, particularly with the words Māori, kiwi, iwi and Pākehā (Māori wars, Māori world, Māori people, Māori language, kiwi dollar, kiwi team, kiwi company, kiwi ingenuity, iwi leaders, iwi member, iwi organization, iwi consultation, Pākehā world, Pākehā language, Pākehā New Zealanders, Pākehā men, Pākehā women). Twitter provides a platform where such hybrids have the potential of being more common due to the word formation principles exploited by hashtags (which mimic compounding).

As this is the first study of Māori loanwords used in Twitter discourse, it is difficult to ascertain whether hybrid hashtags are innovations of the National Science Challenge websites or whether they are used more frequently in this type of data compared to other types of Twitter discourse. When comparing loanwords in the NSCC website contents and the NSCC associated Twitter feeds, we found a 50% overlap between the most frequently occurring loanwords, and this overlap was mostly made up by place names and proper nouns (5 loans: *Māori, Waikato, Otago, Aotearoa, Kaikoura*), rather than content words (3 loans: *kiwi, matauranga, iwi*).

Macalister (2006) classifies Māori loanwords into four broad categories: proper nouns, flora and fauna, material culture, and social culture terms. The first two classes are self-evident. The latter two involve concepts which are either concrete and

"touchable", such as whare 'house', kai 'food', and pā 'settlement', grouped in the material culture set; or more abstract, socially-determined concepts, such as, kinship terms like koroua 'old man', kuia 'old woman, granny', or other abstract notions like mana 'respect, power', and aroha 'love', which come under the label of social culture. As mentioned in the introduction, the use of Māori loanwords is deemed to be increasing in New Zealand (English), cf. Macalister (2006), Kennedy & Yamazaki (1999), Calude et al (in press), with the fastest growing group being the social culture terms, according to Macalister (2006: p. 18). The NSCC also corroborates this finding and we see that despite the fact that five out of the eleven Challenges have an environmental theme and thus raise the expectation that a great majority of the Māori loanwords used might express flora and fauna concepts, it turns out that most loans denote social culture terms, see Table 3. The use of social culture terms suggests that rather than filling a lexical gap related to what is readily detectable in the physical environment, the words which are flowing from Māori into New Zealand English are shifting towards a more cultural perspective, incorporating cultural traditions and concepts which are more specialized and abstract, requiring a longer period of intimate contact between cultures to absorb.

Table 3. The 50 most frequently-occurring loanwords in the National Science Challenge Corpus, by semantic class (excluding hashtags, and the grammatical words ka, o and te)

PROPER	FLORA & FAUNA	MATERIAL CULTURE	SOCIAL CULTURE
NAMES NOUNS			
MĀORI	KAURI	WHENUA	MĀTAURANGA
OTAGO	KIWIFRUIT	KAI	IWI
WAIKATO		MOANA	HĀPU
AOTEAROA		WĀNANGA ⁵	WHĀNAU
KAIKŌURA		MARAE	KAITIAKITANGA
TANGAROA		WAI	KĀHUI
WAIKATO		TAI	TAMARIKI/TAMAITI
TAURANGA		MOTU	TIKANGA
TARANAKI			KAITIAKI
TAHU			WHAI
RAUKAWA			TIPU
IHU			PAINGA
MOTU			REA
			TANGATA

⁵ Sometimes wānanga refers to a specific institution, and sometimes to a particular type of activity, so it can be classed in two different ways (as material or social culture), depending on context.

			MAURI
			KAUPAPA
			WĀNANGA
			MANA
			HUI
			TAONGA
			TAIAO
			MANAAKI
			TAU
			HAUORA
			TAHI
			KĀINGA
	U,		ORANGA
			AO
TOTALS	TOTALS	TOTALS	TOTALS
TYPES 13	TYPES 2	TYPES 8	TYPES 28

Homogeneity across Challenges

The overall frequency list produced by AntConc shows that in comparing the most frequent non-function words (that is, excluding words like *the*, *a*, *and*) in the corpus for each Challenge separately, we find that roughly half of the Challenges exhibit a Māori loanword within the ten most frequent words (the loanword *Māori*); specifically, these are: *A Better Start*, *New Zealand's Biological Heritage*, *Sustainable Seas*, *Aging Well*, and *Building Better Homes*, *Towns and Cities* (which also exhibits the loan *kāinga* 'home').

Focusing on the ten most frequent loanwords in each individual Challenge, we find that there is a great variety of loanword types used, and altogether, we identified a set of 69 distinct loanword types in the eleven Challenges (if there was 100% overlap, we would expect 10 loanwords; with no overlap at all, we would have 110 loanwords). Ranking these types by breadth of occurrence throughout the Challenges, we see that from the 69 types, only 15 loanwords are not unique to a single Challenge, (see Supplementary Materials, Table S4 and Figure S2). In other words, the great majority of loanwords are specific to a particular Challenge.

We can probe the issue of unique loanword use within individual Challenges further by looking at collocations, that is, recurrent groups of loanwords. Because we do not believe that loanwords occur in isolation from other loanwords, we analysed the most

frequent loanword collocates of the word *Māori* (a loanword which is present in all of the eleven Challenges), with the help of Lancsbox (Brezina et al., 2015).

Figure 1 (A and B) shows the network plots derived for each Challenge, using MI scores with a threshold of 6 or higher, by considering a span of five words (a threshold of 3 or above is considered statistically significant, but because the language we are dealing with here seems to be rather formulaic in places, a higher threshold was used). The lines in the graphs are proportional to the strength of collocation (shorter lines correlate with stronger collocation). The network plots indicate that three Challenges (*The Deep South*, Healthier Lives and High-Value Nutrition) do not contain unique loanwords within the top ten most frequently-occurring collocates for the word Māori. For some few Challenges, the uniquely-identifying loanwords are topically predictable: kāinga ('home') for Building Better Homes, Towns and Cities, tirohanga ('view/sight') for NZ Biological Heritage; but for a significant portion of the Challenges, this is not the case (it is unclear why *ao* 'world' and *Aotearoa* 'New Zealand' should be unique collocates for Māori in the Science for Technological Innovation Challenge, or why Waitangi should be a unique collocate for A Better Start). The fact that the loanwords identified are not topically related to the theme of the Challenges in which they occur points to individual stylistic preferences of the authors who produce the online discourse included for each website. In other words, the loans have stronger ties to the individual who uses them, rather than to the topic of discourse within which they appear.

The network plots also show that in some Challenges, the word *Māori* acts as a centre of discourse gravity, attracting many distinct collocates, for instance, *Sustainable Seas*, *NZ Biological Heritage* and *Our Land and Water*; while for other Challenges, the word *Māori* has a more limited scope, i.e., *Building Better Homes, Towns and Cities* and *Science and Technological Innovation*.

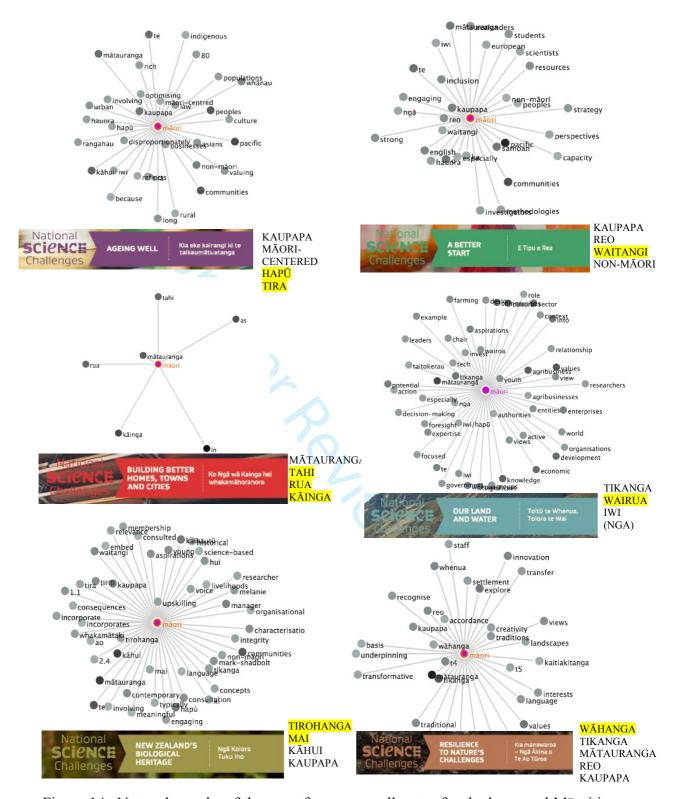


Figure 1A. Network graphs of the most frequency collocates for the loanword Māori in the eleven National Science Challenge websites (MI score, span or 5<>5, threshold=6).

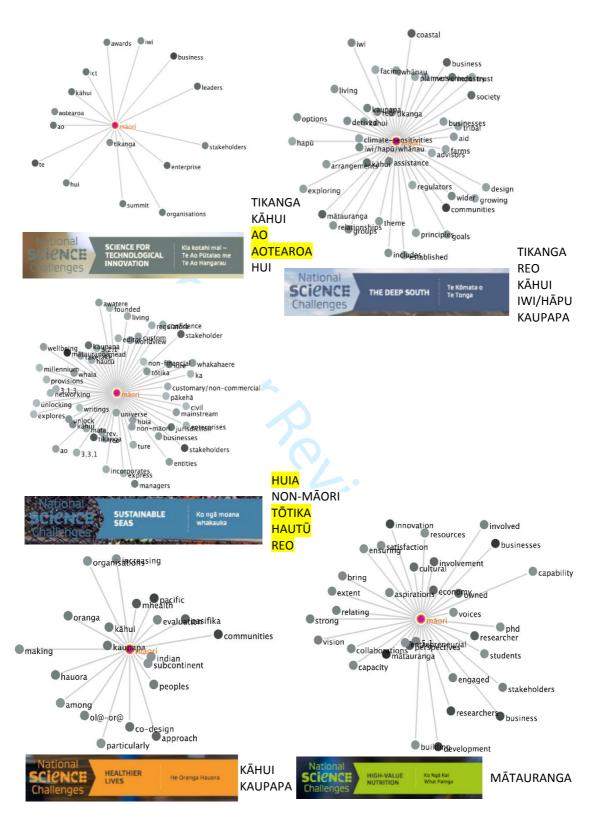


Figure 1B. Network graphs of the most frequency collocates for the loanword Māori in the eleven National Science Challenge websites (MI score, span or 5<>5, threshold=6).

Comparisons with other genres of New Zealand English

Given the unexpected presence of Māori loanwords which do not denote concepts other than flora and fauna in the National Science Challenge Corpus, it might be tempting to assume that the use of Māori loanwords is motivated by the fact that New Zealand English exhibits a steady increase of use of Māori loanwords more generally, and that perhaps these words make up a body of lexical items which are entrenched in typical New Zealand English usage. Put another way, topic is not the motivator for the use of these words, rather, their use is motivated by the variety of English itself. To test this hypothesis, we analysed keyness by comparing various other corpora of New Zealand English currently available.

Keyness was developed in the mid-1990s as part of a toolkit incorporated in a widely used piece of corpus linguistics software, Wordsmith Tools (Scott 1996). The author of Wordsmith Tools, Mike Scott, defined his coined new term 'key word' as "a word which occurs with unusual frequency in a given text [...] by comparison with a reference corpus of some kind" (Scott, 1997, p. 236, as cited in Gabrielatos, 2018, p. 225). As the task of identifying a text's topic remains a notoriously difficult problem with low consensus in the linguistics community, linguists have looked to more concrete concepts, such as keyness, to get at the heart of a given portion of discourse, somewhat awkwardly referred to as the "aboutness" of a text. Having a quantitatively defined method for probing "aboutness" allows for comparisons to be made between texts, corpora, authors, and genres. We believe that keyness could be a useful notion in the study of Māori loanwords because previous work on the subject has pointed to a non-homogenous and highly skewed use of Māori loanwords in New Zealand English (de Bres, 2006; Degani, 2010; Calude et al., 2018).

First, we compared the National Science Challenge Corpus against the older, written Wellington Corpus of New Zealand English (Bauer, 1993). Although the Wellington Corpus of New Zealand English is nearly 25 years old, and contains a diverse set of discourse genres (letters, fiction novels, transactions, and so on), we chose this data because we wanted to investigate the use of Māori loanwords against a "baseline", "average" New Zealand English (though, of course, that is almost impossible to capture), and we wanted to track changes over time. AntConc was used to extract key loanwords between the corpora in order to see which loanwords might be used with higher or lower than expected frequency in the NSCC. As above, keyness measures are calculated by

comparing a corpus of interest (in our case, the NSCC), with a reference corpus (the Wellington Corpus), and measuring Log Ratio Odds (Hardie, 2014), comparing expected rates arising from the reference corpus, with the observed rates from the NSCC. Here, we are interested in loanwords which occur with higher rates than expected (rather than the lower rates) because we want to know which loanwords appear more frequently in the NSCC compared to the Wellington Corpus (the working assumption is that the Wellington corpus will contain many more loanword types just by virtue of the wider topic of discussion). These loanwords are given in Table 4, together with their keyness rank, raw frequency of occurrence in the National Science Challenge Corpus, keyness score and log ratio scores. Keyness ranks provide an ordering of the key words identified in order of importance (with the first ranked word having the highest keyness score). Raw frequency counts are included because they give an impression of how important the word in question might be in the corpus of interest. Keyness scores are calculated using the default AntConc Log likelihood procedure. Finally, log ratio scores provide a measure of the effect size, such that with every increase in log ratio odds, there is an observed doubling of rate of occurrence in the corpus of interest compared to the reference corpus (see Hardie, 2014; and Gabrielatos, 2017 for more details). The table shows for instance that the loanword *Māori* has a log ratio of 2.1783 which means that it was used (just over) 4 times more frequently in the National Science Challenge data than it was in the Wellington Written Corpus (even after normalizing). Similarly, *Mātauranga* has a log ratio of 8.7109, which means it occurred in the National Science Challenge roughly 265 times more frequently than in the Wellington Written Corpus (interestingly, Mātauranga occurs 3 times in the Wellington Written Corpus).

Table 4. Loanwords which are significantly more frequent in the NSCC compared to the Wellington Corpus of Written NZE. The table gives the Māori loanwords in the top 500 keywords words, in the order of keyness, excluding place-names and function words (all macrons were removed from the corpora for this analysis)

Loanword	Wellington Written Corpus (Holmes et al 1998) score				
	Rank	Raw Freq (NSCC)	Keyness score	Log Ratio	
MĀORI	11	5591	2521.07	2.1783	
MĀTAURANGA	22	1659	1833.35	8.7109	
IWI	36	1334	1429.54	7.1739	
HAPU	134	579	604.08	6.4552	
WHĀNAU	223	380	407.06	7.1696	

KĀHUI	229	356	401.62	9.0755
KAITIAKITANGA	253	326	367.77	8.9485
TAMARIKI	269	303	341.82	8.8429
TIKANGA	322	258	291.05	8.611
WHENUA	329	377	284	3.5732
KAITIAKI	375	221	249.31	8.3876
KAI	391	319	237.59	3.5248
PAINGA	408	201	226.75	8.2508
REA	409	201	226.75	8.2508
TIPU	422	203	218.05	7.2651
WHAI	439	214	208.23	5.3412
NGATI	453	381	199.4	2.4729
KAUPAPA	474	169	190.65	8.0006
MOANA	484	189	187.35	5.577
MAURI	488	193	185.36	5.1922

Next, we repeated the process with diachronic corpora compiled on the basis of topic selection, the first was a set of newspaper articles on the topic of Matariki /Māori New Year (Calude et al., forthcoming), and the second, a corpus of newspaper articles on the topic of Te wiki o te reo Māori/ Māori Language Week (Levendis & Calude, 2018, 2019). Both of these corpora comprise articles over a period of 10 years, from 2007-2018. The corpora were precisely put together to fruitfully inform the study of loanwords over that period, given the association of Māori loanwords with Māori topics. We reasoned that loanwords which are significantly more frequent in the NSCC compared to these corpora would be noteworthy, given the already high expectation of frequent loanword use in these topically-constrained written corpora. The keyness analysis was repeated in turn with each corpus and the results are given in Tables 5A and 5B.

Table 5A. Loanwords which are significantly more frequent in the NSCC compared to the *Matariki Corpus* (all loanwords, excluding place-names and function words)

Loanword		Matariki Corpus (Calude et al in press)				
	Rank	Raw Freq (NSCC)	Keyness score	Log Ratio		
MĀTAURANGA	43	1659	165.62	4.005		
HAPŪ	281	579	45.93	2.9718		
IWI	297	1334	44.04	1.4104		
KAITIAKITANGA	306	326	42.74	5.465		
TIPU	476	203	26.62	4.7816		
KĀHUI	514	356	24.78	2.592		

Table 5B. Loanwords which are significantly more frequent in the NSCC compared the

Māori Language Week Corpus.

Loanword	Māori Language Week Corpus (Calude and Levendis in prep)			
	Rank Raw Freq Keyness score Log Ratio (NSCC)			
MĀTAURANGA	20	1659	158.73	7.3442
KAITIAKITANGA	284	326	31.18	4.9969

One final observation concerns loanwords which were found to be especially infrequent in the NSCC, compared to the Matariki and the Māori Language Week Corpora. Given their Māori focus and wider authorship breadth, the latter corpora are expected to contain many more loanwords that are absent from the NSCC data. However, the loanwords Pākehā and Kiwi were both conspicuously absent from the NSCC which caused some surprise given that the National Science Challenges are meant to represent both Māori and Pākehā interests and concerns. It appears that the word Pākehā is becoming widely replaced by the word non-Māori, a use which has previously not featured on lists compiled to document the various lexical ways of referring to New Zealanders of European descent (Daly, 2017, p. 4). It remains to be seen to what extent *non-Māori* overlaps with the semantic content denoted by the term *Pākehā*. The examples below indicate that *non-Māori* is sometimes used to refer to any New Zealanders who are not Māori (Pākehā but also New Zealanders of other nationalities living in New Zealand), as in example (1), but at other times, non-Māori represents a semantic opposition to the term New Zealand European specifically, thereby implying a closer correspondence to *Pākehā*, see example (2) (and not including other ethnic groups).

- (1) Professor Kerse's research areas include promoting activity and function in residential care, residential care organisational culture and outcomes, promoting physical activity in community dwelling older people, activity for depression in the very old, staying upright (preventing falls and injury) in older people in all settings, improving prescribing in primary care, and a large cohort of Māori and **non-Māori** in advanced age. (from the website of the *Aging Well* NSC).
- (2) The obvious Vision Mātauranga theme that the Tūhonohono project comes under is "Taiao: Achieving Environmental Sustainability through Iwi and Hapū relationships with land and sea." Tūhonohono is about Māori and **non-Māori** aspiring to live in sustainable communities dwelling in healthy marine and coastal area environments. Distinctive

environmental research arising in Māori communities relates to the expression of iwi and hapū tikanga and mātauranga knowledge, culture and experience over the takutai moana area such as rahui, kaitiakitanga, mana whenua, mana moana, and matāwhanga in the sustainable use and restoration of the marine environment. We will identify these concepts 'in practice' through approaches that are based on whakapapa and 'place based' values and perspectives which will unlock the science and innovation potential of tikanga and mātauranga Māori knowledge, resources and peoples over the takutai moana area which should lead to building synergistic environmental sustainability by combining the best of both founding cultures – Māori and **New Zealand European** – legal systems, philosophies, values, rules and institutions.

(from the website of the *Sustainable Seas* NSC)

In so far as the term *non-Māori* refers to Pākehā (or New Zealanders of European descent), its morphological derivation from the word *Māori* has the symbolic effect of construing the distribution of populations in New Zealand as either Māori or non-Māori thereby, legitimizing the historic perspective of the indigenous people of the country (Māori becomes the 'default' group).

Case-study: the use of Mātauranga

Given the pervasiveness and importance of the loanword *Mātauranga* and the importance of the Vision Mātauranga Policy, we investigated its use in the National Science Challenge Corpus more closely. The total number of 1652 uses of the loanword across the corpus is unevenly spread between the Challenges, with some using it more than others, and two only using it in passing (*Building Better Homes, Towns and Cities* and *Healthier Lives*). The loanword is spelt both with a macron (*Mātauranga* n=1610) and without a macron (*Matauranga* n=42), and interestingly, most Challenges include both spellings (notable exceptions being the two Challenges which do not use the word extensively, *Building Better Homes, Towns and Cities, Healthier Lives*, along with *Aging Well*).

Figure 2 shows the overall use of *Mātauranga* in each Challenge, normalized per 10,000 words (see Supplementary Materials, Figure S2 for raw numbers of occurrence in each Challenge). This plot shows that the highest use of the loanword *Mātauranga* comes

from the following Challenges: *Science for Technological Innovation*, *Sustainable Seas* and *NZ's Biological Heritage*.

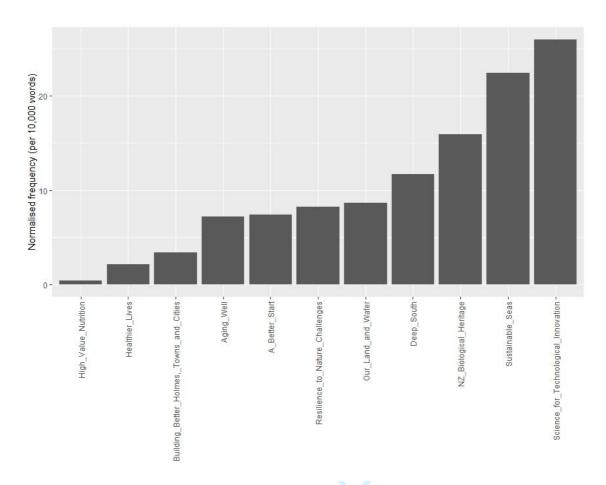


Figure 2⁶. Frequency of the use of *Mātauranga* in each Challenge (normalized per 10,000 words)

Next, we looked at how the loanword *Mātauranga* is used in the text and to this end, we extracted its most frequent collocates by searching neighbouring words (occurring both, to the left and the right of it). Perhaps unsurprisingly, by far the strongest collocate was the word *vision* (*Vision Mātauranga*). Grouping collocates by word-class, we found that *Mātauranga* occurred with the nouns *Māori*, *programme(s)*, *policy*, *science*, *theme(s)*, *knowledge*, *principles*, *research*, *approaches*, *alignment*, *framework*, *project(s)*, *workshops*, *objectives*, *strategy*, *clarification*, *capability* and *leader*. Most of these occurred to the right of the loanword, but the nouns *science* and *knowledge* also

-

⁶ This plot (and all plots in the Supplementary Materials) were drawn using R (R Core Team, 2014).

⁷ We ignored the use of the macron for this analysis.

occurred to the left of it (i.e., science Mātauranga and Mātauranga science). Looking at verbs which occurred with Mātauranga, we found the verbs: reframe, employ, incorporate, examine, integrate, explored, need, include occurring to the left of the loanword, and the verb address(es) occurring to its right (in addition to the most generic is/are, can, will). Widening the context of collocation illuminated the following frequent phrasal units: Mātauranga Māori and, Mātauranga and tikanga (Māori), Mātauranga Māori and science, Mātauranga Māori research, Mātauranga Māori and kaitiakitanga, principles of Vision Mātauranga, and systems derived from Mātauranga. Studying collocational patterns hints at the ways in which writers construe the concept of Mātauranga and how they refer to it, namely, it is a type of science, knowledge, principle, research, approach and framework, which is to be reframed, employed, incorporated, examined, integrated, included, and aligned to.

Conclusion

In conclusion, we found a strong presence of Māori borrowings in National Science Challenge web and Twitter feeds which may be reflective of the prominence given to the Vision Mātauranga Policy across the government's research, science and technology investment programme. Using corpus linguistics methods, we highlight the growing use of Māori loanwords as a national identity building tool in this genre, in addition to being one of the many mechanisms to signal a growing cultural awareness of Māori traditions and concepts and the many 'challenges' that the country is currently facing.

Acknowledgements

We thank the University of Waikato Summer Research Fellowship. AC and HW thank the NZ Royal Society Marsden Fast-Start Fund. (Removed for anonymity purposes.)

References

Anthony L 2018. AntConc (Version 3.5.7) [Computer Software] Tokyo, Japan: Japan Waseda University. Accessible from: http://www.laurenceanthony.net/

- Anthony, L., & Hardaker, C. (2016). FireAnt (Version 1.0) [Computer Software] Tokyo, Japan: Japan Waseda University. Accessible from: http://www.laurenceanthony.net/
- Bauer, L. (1993). Manual to accompany the Wellington Corpus of Written New Zealand English. Department of Linguistics, Victoria University of Wellington, New Zealand.
- 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., Harper, S., Whaanga, H., & Miller, S. (forthcoming) Loanwords as Textual Collocates: Investigating Māori Loanwords related to Matariki in New Zealand English newspapers. *Asia-Pacific Language Variation Journal*.
- Calude, A., Pagel, M., & Miller, S. (2018). Modelling Borrowing Success a quantitative study of Maori loanwords. *Journal of Theoretical Linguistics and Corpus Linguistics*, 15, 2-38. https://doi.org/10.1515/cllt-2017-0010
- Coupland, N. (2007). *Style language variation and identity*. Cambridge: Cambridge University Press.
- Daellenbach, U., Davenport, S., & Ruckstuhl, K. (2017). Developing absorptive capacity for midstream science in open innovation contexts. *International Journal of Technology Transfer and Commercialisation*, *15*(4), 447–462. https://doi.org/10.1504/IJTTC.2017.089674
- Davies, C., & Maclagan, M. (2006). Māori Words Read all about it: testing the presence of 13 Māori words in four New Zealand newspapers from 1997 to 2004. *Te Reo*, 49, 73–99.
- Daly, N. (2017). Pākehā–Māori, European–Native: Ethnic Labeling in the Dorothy Neal White Collection. *New Review of Children's Literature and Librarianship*, 23(1), 1–12. https://doi.org/10.1080/13614541.2017.1280335
- 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 Facchinetti, R., Crystal,

- D., & Seidlhofer, B. (Eds.), From International to Local English and Back Again. (pp. 165–196). Bern, Switzerland: Peter Lang.
- Degani, M., Onykso, A. (2010). Hybrid Compounding in New Zealand English. *World Englishes*, 29(2), 209–233.
- Deverson, T. (1991). New Zealand English Lexis: the Māori Dimension. *English Today* 28, 18–25.
- Gabrielatos, C. (2018). Keyness Analysis: Nature, Metrics and Techniques. In Taylor, C., & Marchi, A. (Eds.), *Corpus Approaches to Discourse: A Critical Review*. (pp. 225–258.). Oxon, England: Routledge.
- Hardie, A. (2014). Log Ratio an Informal Introduction. Post on the website of the ERSC Centre for Corpus Approaches to Social Science CASS. Accessible from http://cass.lancs.as.uk/?p=1133 [Accessed 8 January 2019].
- Harlow, R. (2005). Covert Attitudes to Māori. *International Journal of the Sociology of Language 172*, 133–147.
- Hikuroa, D. (2017). Mātauranga Māori—the ūkaipō of knowledge in New Zealand. *Journal of the Royal Society of New Zealand*, 47(1), 5–10.
- Holmes, J., & Bell, A. (1992). On shear markers and sharing sheep: The merger of EAR and AIR diphthongs in New Zealand English. *Language Variation and Change*, 4(3), 251–271.
- Kennedy, G., & Yamazaki, S. (1999). The influence of Māori on the New Zealand English ——lexicon. In Kirk, J. (Ed.), *Corpora Galore: Analyses and Techniques in Describing English*. (pp. 33–44). Amsterdam, The Netherlands: Rodopi.
- Kingi, T.K., Durie, M., Elder, H., Tapsell, R., Lawrence, M., & Bennett, S. (2018). *Maea te Toi Ora: Māori Health Transformations*. Wellington, New Zealand: Huia.
- Levendis, K., & Calude, A. (forthcoming 2019). Flagging loanwords and what they can tell us a case study from New Zealand English. *Ampersand* (Special Issue on Loanwords and Identity).
- Levendis, K., & Calude, A. (2018, November). 'When you don't know what you know': Use of Māori loanwords in a diachronic corpus of New Zealand English. Paper presented at the New Zealand Linguistics Society Conference, Victoria University of Wellington, New Zealand.

- Macalister, J. (2008). Tracking changes in familiarity with borrowings from te reo Māori. *Te Reo 51*, 75–97.
- Macalister, J. (2007). Weka or woodhen? Nativization through lexical choice in New Zealand English. *World Englishes* 26(4), 492–506.
- Macalister, J. (2006). The Maori presence in the New Zealand English lexicon, 1850–2000: Evidence from a corpus-based study. *English World-Wide* 27, 1–24.
- Maclagan, M., & Gordon, E. (1996). Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change* 8(1), 125–147.
- Māori Language Commission, (2018). *Te Taura Whiri I te Reo Māori*. http://www.tetaurawhiri.govt.nz/ [accessed January 2019].
- Mead, H.M. (2012). "Understanding Mātauranga Māori Conversations on Mātauranga Māori". In Bean, D., Black, T., Collings, W., & Nuku, W. (Eds). *Conversations on Mātauranga Māori*. (pp. 9–14). Wellington, New Zealand: New Zealand Qualifications Authority.
- Mercier, O.R. (2013). Indigenous knowledge and science. A new representation of the interface between indigenous and eurocentric ways of knowing. *He Pukenga Korero*, 8(2).
- Ministry of Research Science and Technology. (2007). Vision Mātauranga: Unlocking the innovation potential of Māori knowledge, resources and people. Wellington, New Zealand: New Zealand Ministry of Research Science and Technology (MoRST).
- McEnery, T. & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge, England: Cambridge University Press.
- R Core Team (2014). R: A language and environment for statistical computing. R

 Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/
- Royal, T.A.C. (2009). *Mātauranga Māori*. Porirua, Wellington: Mauriora-ki-te-Ao/Living Universe Ltd.
- Royal, T.A.C. (2012). Politics and knowledge: Kaupapa Māori and mātauranga Maori. *New Zealand Journal of Educational Studies*, *47*(2), 30–37.

- RNZ (Radio New Zealand) (2017). *Why RNZ is dedicated to the use of te reo*. Published 1 December 2017, https://www.radionz.co.nz/news/on-the-inside/345170/why-rnz-is-dedicated-to-the-use-of-te-reo [accessed January 2019].
- Schneider, E. (2003). The Dynamics of New Englishes: from Identity Construction to Dialect Birth. *Language* 79(2), 233–281.
- Smith, L.T. (2012). *Decolonising methodologies: Research and Indigenous peoples*. London, UK: Zed Books.
- Smith, L.T., Maxwell, T.K., Puke, H., & Temara, P. (2016). Indigenous knowledge, methodology and mayhem: What is the role of methodology in producing , ussio. indigenous insights? A discussion from Mātauranga Maori. Knowledge Cultures, 4(3), 131–156.

Table 1. The 20 most frequent Māori loanwords in the entire NSCC corpus and their associated raw frequencies of occurrence

Rank frequency	Māori loanword	English equivalent	Raw frequency
1	MĀORI/ MAORI	(of) indigenous (origin)	5318 (with macron) +414 (without macron) =5732
2	MĀTAURANGA	traditional knowledge	1640
3	IWI	tribe	1348
4	OTAGO ¹	Otago	904
5	HAPŪ	sub-tribe	547
6	WAIKATO	Waikato	409
7	WHENUA	land	382
8	WHĀNAU	extended family	340
9	KAIKŌURA	Kaikoura	336
10	KAITIAKITANGA	guardianship	335
11	NGĀTI	prefix to tribal name	334
12	KAI	food	324
14	KĀHUI	group	321
14	AOTEAROA	New Zealand	312
15	TAMARIKI	children	304
16	TIKANGA	custom	259
17	TANGAROA	God of the sea and fish	253
18	KAITIAKI	guardian	225
19	WHAI	to have, to pursue	214
20	ORA	health, to be alive	204

¹ The native Māori word had the form "Ōtākou", which was later changed to "Otago" to fit with Anglicized spelling and phonology. Despite this change, we note the Māori origin of the word and count it as a loanword here.

Table 2. A comparison between web and Twitter components in the NSCC with respect to the 20 most frequent Māori loanwords.

Rank frequency	Māori loanword WEB	English equivalent	Māori loanword TWITTER	English equivalent
1	MĀORI/ MAORI	(of) indigenous (origin)	MĀORI/ MAORI	(of) indigenous (origin)
2	MĀTAURANGA	traditional knowledge	#MOTURESEARCH	hashtag
3	IWI	tribe	OTAGO	Otago
4	OTAGO	Otago	MĀTAURANGA	
5	HAPŪ	sub-tribe	WAIKATO	Waikato
6	WAIKATO	Waikato	#HLKORERO	hashtag
7	WHENUA	land	KŌRERO	talk/ a talk
8	WHĀNAU	extended family	KAURI	
9	KAIKŌURA	Kaikoura	TAHI	one
10	NGĀTI	tribe	KAIKŌURA	Kaikoura
11	KAITIAKITANGA	guardianship	#BIOHUI	hashtag
12	KAI	food	AOTEAROA	New Zealand
13	KĀHUI	group	IWI	tribe
14	TAMARIKI	children	ORA	health
15	AOTEAROA	New Zealand	MIHI	introduction
16	TIKANGA	custom	KIWIS	New Zealanders, flightless bird
17	TANGAROA	God of the sea and fish	NGAITAHU	the tribe of Ngāi Tahu
18	KAITIAKI	guardian	#KAURIRESCUE	
19	WHAI	to have, to pursue	KAITIAKITANGA	guardianship
20	TIPU	growth	KIWI	native bird, New Zealander

Table 3. The 50 most frequently-occurring loanwords in the National Science Challenge Corpus, by semantic class (excluding hashtags, and the grammatical words *ka*, *o* and *te*)

PROPER NOUNS	FLORA & FAUNA	MATERIAL CULTURE	SOCIAL CULTURE
MĀORI	KAURI	WHENUA	MĀTAURANGA
OTAGO	KIWIFRUIT	KAI	IWI
WAIKATO		MOANA	HĀPU
AOTEAROA		WĀNANGA ¹	WHĀNAU
KAIKŌURA		MARAE	KAITIAKITANGA
TANGAROA		WAI	KĀHUI
WAIKATO		TAI	TAMARIKI/TAMAITI
TAURANGA		MOTU	TIKANGA
TARANAKI			KAITIAKI
TAHU	0,		WHAI
RAUKAWA			TIPU
IHU			PAINGA
MOTU			REA
			TANGATA
			MAURI
			KAUPAPA
			WĀNANGA
			MANA
			HUI
			TAONGA
			TAIAO
			MANAAKI
			TAU
			HAUORA
			ТАНІ
			KĀINGA
			ORANGA
			AO
TOTALS	TOTALS	TOTALS	TOTALS
TYPES 13	TYPES 2	TYPES 8	TYPES 28

 $^{^{1}}$ Sometimes $w\bar{a}nanga$ refers to a specific institution, and sometimes to a particular type of activity, so it can be classed in two different ways (as material or social culture), depending on context.

Table 4. Loanwords which are significantly more frequent in the NSCC compared to the Wellington Corpus of Written NZE. The table gives the Māori loanwords in the top 500 keywords words, in the order of keyness, excluding place-names and function words (all macrons were removed from the corpora for this analysis)

Loanword	Wellington Written Corpus (Holmes et al 1998) score			
	Rank	Raw Freq (NSCC)	Keyness score	Log Ratio
MĀORI	11	5591	2521.07	2.1783
MĀTAURANGA	22	1659	1833.35	8.7109
IWI	36	1334	1429.54	7.1739
HAPU	134	579	604.08	6.4552
WHĀNAU	223	380	407.06	7.1696
KĀHUI	229	356	401.62	9.0755
KAITIAKITANGA	253	326	367.77	8.9485
TAMARIKI	269	303	341.82	8.8429
TIKANGA	322	258	291.05	8.611
WHENUA	329	377	284	3.5732
KAITIAKI	375	221	249.31	8.3876
KAI	391	319	237.59	3.5248
PAINGA	408	201	226.75	8.2508
REA	409	201	226.75	8.2508
TIPU	422	203	218.05	7.2651
WHAI	439	214	208.23	5.3412
NGATI	453	381	199.4	2.4729
KAUPAPA	474	169	190.65	8.0006
MOANA	484	189	187.35	5.577
MAURI	488	193	185.36	5.1922

Table 5A. Loanwords which are significantly more frequent in the NSCC compared to the

Matariki Corpus (all loanwords, excluding place-names and function words)

Loanword		Matariki Corpus (Calude et al in press)				
	Rank	Raw Freq (NSCC)	Keyness score	Log Ratio		
MĀTAURANGA	43	1659	165.62	4.005		
HAPŪ	281	579	45.93	2.9718		
IWI	297	1334	44.04	1.4104		
KAITIAKITANGA	306	326	42.74	5.465		
TIPU	476	203	26.62	4.7816		
KĀHUI	514	356	24.78	2.592		

Table 5B. Loanwords which are significantly more frequent in the NSCC compared the *Māori*

Language Week Corpus.

Loanword	Māori Language Week Corpus (Calude and Levendis in prep)				
	Rank Raw Freq Keyness score Log Ratio (NSCC)				
MĀTAURANGA	20	1659	158.73	7.3442	
KAITIAKITANGA	284	326	31.18	4.9969	

Figure 1A. Network graphs of the most frequency collocates for the loanword Māori in the eleven National Science Challenge websites (MI score, span or 5 <> 5, threshold=6).

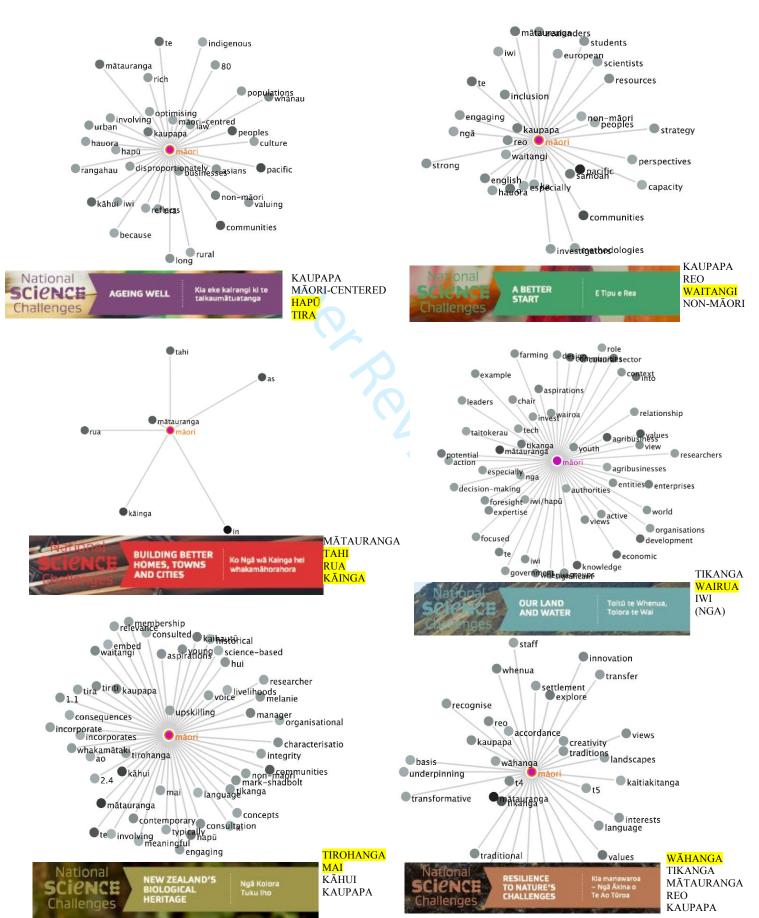
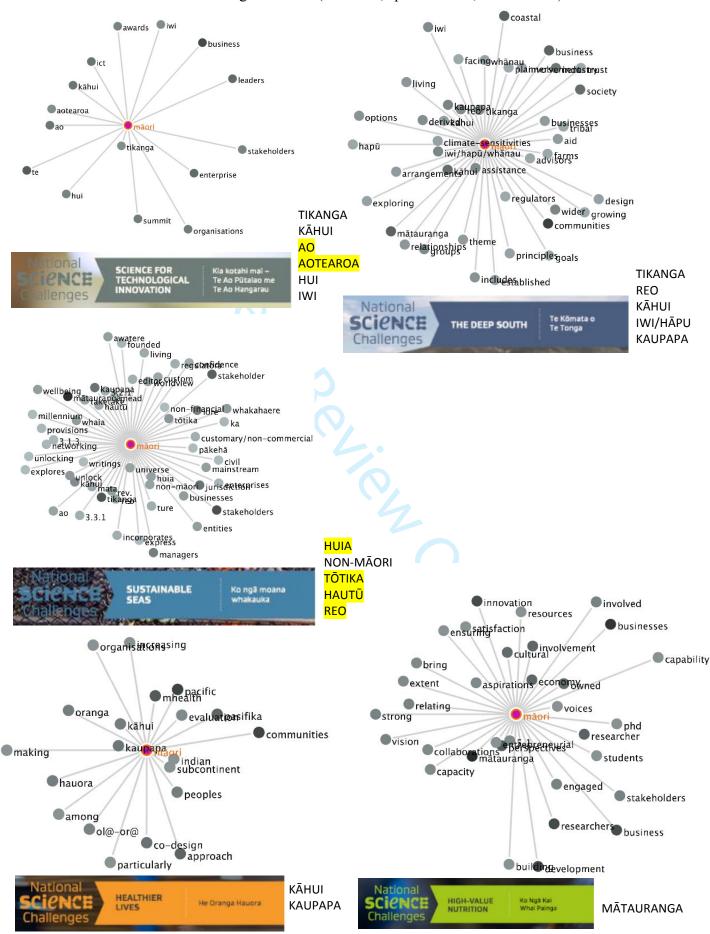


Figure 1B. Network graphs of the most frequency collocates for the loanword Māori in the eleven National Science Challenge websites (MI score, span or 5 >> 5, threshold=6).



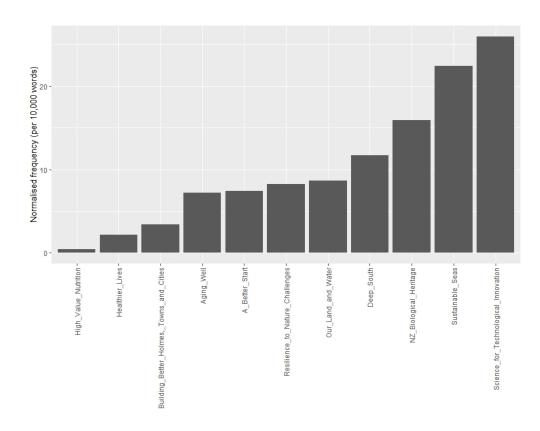


Figure 2. Frequency of the use of Mātauranga in each Challenge (normalized per 10,000 words) $306x247mm~(72 \times 72~DPI)$