

# Perception and flagging of loanwords – A diachronic case-study of Māori loanwords in New Zealand English



Katharine Levendis, Andreea Calude\*

University of Waikato, New Zealand

## ARTICLE INFO

**Keywords:**  
New Zealand English  
Māori  
Loanwords  
Flagging  
Integration

## ABSTRACT

This paper combines a quantitative and qualitative analysis of a diachronic corpus of New Zealand newspapers built to analyse the use of Māori loanwords in New Zealand English. We report findings in relation to flagging (marking of loanwords as being foreign material in a given language) and show that it is (weakly) predicated by frequency-of-use and by semantic category of the loanword (core loans are flagged more than cultural ones), but not by listedness. Alongside this trend, we note that perceptions of writers using the words vary enormously in regard to which loans are integrated and familiar, matching neither listedness nor frequency-of-use patterns. This indicates that in NZE, loanword use remains strongly tied up with socio-political identity and language ideology, rather than rooted in linguistic factors (such as, bilingualism or filling in lexical gaps).

## 1. Loanword spotting

(1) Vodafone New Zealand consumer director Matt Williams said supporting Google to improve pronunciation in Aotearoa was an important way they could play a part in preserving the taonga (treasure) of te reo Māori.

(TW\_2017\_12\_SEPT\_BPT\_002)

If you have never been to New Zealand, you will likely not know the meaning of some of the words and phrases in the 2017 newspaper article excerpt quoted above, namely *Aotearoa*, *taonga*, and *te reo Māori*. The author of the article does provide some help for the word *taonga*, giving its English equivalent in brackets, *treasure*; but not for the other Māori words, *Aotearoa* and *te reo Māori*. Māori is the indigenous language spoken in New Zealand (Aotearoa in Māori), a language which is now in a stage of revitalisation following a period of severe endangerment.

The practice of translating or explaining borrowed words, or of demarcating their occurrence in some way from surrounding discourse, by italics or bold face font has been noted in passing by various researchers but it has to date – to our knowledge – not received full attention in its own right. This paper aims to address this gap by presenting data from a case-study of New Zealand English. As discussed later in the paper, flagging can occur in both speech and writing, but our paper only concerns the presence of flagging in written language.

While our focus here is New Zealand English, we note that New Zealand English is not the only language to exhibit the phenomenon, as can be seen in examples (2a-d).

- (2)<sup>1</sup>
- a. English loanwords explained in Swedish (Ref. [1]; p. 234, ex. 20)  
Vi har pratat med HY [Eng. company acronym] om en co-operation här nu. Lite Svengelska här nu. [laughs] SAMARBETE [emphasized].  
We have spoken to HY [Eng. company acronym] about a co-operation here now. Some Swenglish here now. [laughs] CO-OPERATION [emphasized].
  - b. French loanwords explained and marked by quotations in Quebecois French (Ref. [2]; p. 24, ex. 2)  
Bloc leader Bouchard called, in April, for a “virage” or sharp turn in sovereigntist strategy.
  - c. English loanwords explained in German (Ref. [3]; pp. 97–98, ex. 12)  
Moderator: (geht zu Sprecher A and gibt ihm die Hand) Nice try, netter Versuch. Ist er nicht, dar gibt's doch nicht.  
Host: (goes toward speaker A to shake his hand) Nice try, nice try. It has not. That's impossible.
  - d. English loanwords italicized in a French text (Ref. [4]; p. 282, ex. 1)

\* Corresponding author.

E-mail address: [andreea@waikato.ac.nz](mailto:andreea@waikato.ac.nz) (A. Calude).

<sup>1</sup> The formatting in these examples is faithful to the original text in order to allow for detection of textual marking of the loanword as distinct from the surrounding discourse.

“Bien”, il dit, “tenez, m'en vas [sic] faire un *bargain* avec vous.”  
 “Well”, he says, “here, I'm gonna make a bargain with you.”

## 2. Flagging, integration, frequency of use, listedness

There is currently no consistent means for referring to the practice of marking the foreign origin of a word imported from another language, or a donor language (DL). Various terms are employed to describe the practice, such as marking [5], flagging [1,2,6,7], glossing [8–10], or simply describing in words the loanword use, i.e., ‘*they appear in italics*’ [4]. Moreover, the same term can be used in different ways. For instance, for Grant-Russell and Beaudet [2] flagging does not include translations or explanations but is specifically used to denote typographical markings in a written text. In code-switching studies, flagging refers to hesitations, self-corrections and other meta-linguistic commentary and is seen as a contrast to a ‘smooth codeswitch’ where the change from one language to another occurs without explicit marking [6,7,11,12]. These issues indicate that the phenomenon is both widespread, and in need of standardization. The succinct literature review that follows aims to show that flagging may have interesting consequences for the study of loanwords and that a more sophisticated approach to this phenomenon is needed.

Although loanword flagging is only mentioned in passing by studies on loanwords (e.g. Refs. [1,2,5–7]), including those on Māori loanwords [8–10], the phenomenon proves an important one, as it is used as supplemental evidence to support assessments of entrenchment (or integration) and frequency-of-use.

Sharp considers the use of loanword flagging in her corpus of Swedish business language to be ‘indicative of the unintegrated status of the English items used’ ([1]; p. 235). The flagging in this spoken data comes in the form of voice modification, particularly in regard to young speakers who adopt English accents and exaggerated changes in pronunciation when using English loanwords to signal ‘both the shift of language and also a shift of role in discourse, i.e., the words used are not to be considered his/her own: the speaker has adopted another persona and is merely playing, and thus does not want to be taken seriously’ ([1]; p. 235–236). It is clear that for Sharp, **integration** refers to the micro-linguistic environment rather than integration in the language as a whole.

In contrast, Refs. [7,11] investigate flagging in order to dig deeper to a macro-linguistic perspective of integration, looking at the status of the borrowed words in the language as a whole. Ref. [7] analyses spoken French containing English loanwords and uses flagging (for her, hesitations, pauses, self-corrections) to test the extent to which various lone English borrowings are integrated inside French. Because flagging of this type is not unique to language contact scenarios, as pauses and hesitations occur equally in the presence of discourse containing *native* French words alone (if such a thing exists), the practice can only be measured comparatively. Ref. [7]’s data showed that flagging was not used with English loanwords more frequently than in unmixed portions of discourse, and on this basis, she concluded that ‘lone English-origin nouns were treated as legitimate members of the lexicon of the language in which they are embedded’ ([7]; p. 225). Similarly, Ref. [11] found that flagging was never used with English loanwords in Jèrriais (an endangered Jersey French dialect spoken in the Channel Islands) and was only used in the context of code-switching to mark language switch boundaries. The language contact situation investigated here is particularly interesting as the receiver language (RL) is an endangered language, so the use of English words in Jèrriais constitutes an asymmetrical lexical movement ([11]; p. 3). This contact situation is relevant to the case-study of Māori loanwords in New Zealand English investigated in this paper because it also constitutes a case of asymmetrical lexical movement; though the direction of movement is reversed for us. In our case, the loanwords are going from the endangered minority language into a majority one.

The studies by Refs. [7,11] suggest that discourse flagging might be

helpful in distinguishing between borrowing and code-switching – one of the holy grail questions of language contact linguistics (see ongoing debate in Refs. [13–15], and summary discussions in Refs. [12,16] among others).

In other studies, flagging helps to determine the status of loanwords. For [2], the presence of lexical flagging in written Quebecois French texts indicate the author’s awareness of the origin of the words used (in this case, English), but also its separate status and degree of intentionality. The *separate status* of a loanword refers to the explicit effort made to increase the force of its impact. Ref. [2] explains that ‘widely established borrowings are not necessarily consciously motivated. The more a borrowing is perceived as standard usage, the more likely it is to be unmarked and the less intentionality it conveys’ ([2]; p. 26). In agreement with observations made by Ref. [17], they find an inverse relationship between **frequency-of-use** and loanword flagging: the more frequent a borrowing, the more accepted, and thus less likely to be flagged ([17]; p. 26).

Acceptance is linked to another parameter that is becoming widely adopted as an indicator of loanword status, namely, listedness. Muskyen [12] defines **listedness** as ‘the degree to which a particular element or structure is part of a memorised list which has gained acceptance within a particular speech community’ ([12]; p. 71). How one might access this ‘memorised list’ is not entirely clear, but it is tempting to assume that listedness is a function of frequency; the more frequent a loanword, the more likely it is to be listed. Yet a study by Stammers and Deuchar [16] shows the situation is more complex.

They operationalise the concept of listedness by recourse to dictionaries: words which occur in the dictionary count as listed, i.e., they are accepted by the community as part of a given language or as ‘established borrowings’ ([16]; p. 631). Conversely, loanwords which are not found in the dictionary are unlisted. Their study of English verbs borrowed into Welsh finds that frequency and listedness constitute distinct parameters, which can (and should) be disentangled from one another. Some of the English verbs in their data underwent soft mutation, that is, phonological adaptation to Welsh phonotactic constraints (a measure of integration). The patterns uncovered reveal that as expected, frequency is a predictor of mutation rates, with more frequent loans being more likely to mutate, but also that listedness is indeed a significant predictor above and beyond frequency, such that, loans which occurred with the same frequency behaved differently depending on whether or not they were listed: the listed ones were more likely to mutate than the unlisted ones ([16]; p. 641).

The short review above indicates that flagging contributes to a number of the key questions in the area of loanwords, including (a) how to detect whether a loanword is firmly entrenched in the lexicon of an RL, and (b) what lies at the boundary between borrowing and code-switching? While these questions are beyond the scope of the current paper, in our view, their link to flagging calls for a more sophisticated understanding of the phenomenon. To this end, our work presents a quantitative and qualitative analysis of flagging in New Zealand English and puts forward two main arguments:

- 1) flagging is not only merely a linguistic reflex, it is also a socio-linguistic lens through which speakers are able to exhibit linguistically their perceptions and attitudes towards a particular language and culture (by means of their lexical choices in using, for instance, novel loanwords),
- 2) flagging is mediated by speaker perception and speaker attitude and is therefore not necessarily an accurate barometer for the degree of acceptance or the degree of integration of a given loanword in a RL, especially in language contact scenarios of asymmetrical lexical movement.

## 3. New Zealand English: a different contact scenario

New Zealand English is spoken in one of the most remote places on

earth, the island group of New Zealand (comprising the North Island, the South Island and Stewart Island). It is an emerging variety of English which shows the effects of language contact with the Austronesian language Māori. *Te Reo Māori* (the Māori language) is spoken by the indigenous people of New Zealand, who first arrived from Polynesia between approximately 1210AD and 1385AD [18]. Discounting a brief encounter with Abel Tasman in 1642, the first official contact between Māori and Europeans took place in 1769, with the arrival of Captain James Cook [18].

According to Macalister [19], it is ‘widely accepted’ that Māori words are ‘the most distinctive feature of New Zealand English’ ([19]; p. 1). Moreover, Māori has also donated a large number of lexical items to the English language worldwide, making it the fifteenth highest contributor of lexis to English ([20]; p. 25).

New Zealand English presents a fascinating case-study for loanwords in (at least) three respects. First, the use of Māori loanwords has been intensely studied over the past thirty years (see for instance Ref. [21]) which allows us to build a detailed picture of its development through time.

Secondly, the sociolinguistic scenario is unique as loanwords are moving from a minority language into a majority one (see Ref. [16]; p. 634, Table 1 for typical contact types, exhibiting no situation like the one discussed here). Arguably, there is always some imbalance of power and prestige between a source language (SL) and receiving language (RL). But generally speaking, most studies consider scenarios in which the transfer of loanwords takes place from a higher status language into a lower status language. There are to date only a few studies which investigate a clear asymmetrical loanword movement, namely by Ref. [11] (from English into Jèrriais) and [16] (from English into Welsh), but the direction of movement is inverse to our case study, because the loanwords in our case-study are imported from a non-dominant language (Māori) to a more dominant language (New Zealand English). As regards agentivity, the contact situation is largely RL-dominant (cf. Ref. [22]) and one of (strictly-speaking) *borrowing* rather than *imposition* (in the sense of Ref. [23]).

Thirdly, the Māori loanwords borrowed in NZE do not follow expected patterns with regard to frequency of occurrence in specific genres. While [24] predicts more borrowings in speech than in writing due to a decrease in language formality with a likely increase in language interference (cited in Ref. [3]; p. 19), this is not what we find in NZE. Comparisons by Ref. [24] suggest that (at least in the 1990s), there was no real difference in the loanword frequencies in spoken and written NZE. Since then, a wealth of studies of newspaper language have also documented ever-increasing loanword usage in formal, highly edited, written NZE [8,9,19,25–28]. As one anonymous referee suggests, the function of loanwords in written, newspaper language extends much beyond purely referential content and into stylistic functions, and perhaps even to stance marking [29]. Loanword use in this genre cannot be seen as a careless slip or rash choice because of the high level of scrutiny that newspaper language undergoes. Furthermore, newspaper language represents a monolingual environment, in which language interference cannot be invoked as a possible factor.

#### 4. Data: Māori Language Week corpus

This study reports findings from the analysis of a corpus of newspaper articles which we collected by constraining the topic of discourse

**Table 1**  
Types and tokens in the corpus.

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

specifically to Māori Language Week.<sup>2</sup> The *Māori Language Week Corpus* (henceforth the *MLW Corpus*) is to date the only corpus of Māori loanwords in NZE to consist of data on a single topic. We have built the *MLW Corpus* by sourcing *New Zealand Media and Entertainment* subsidiary newspaper articles from the newspaper database *Newztext*. Only articles with either of the search terms ‘Māori Language Week’ or the Māori equivalent ‘te Wiki o te Reo Māori’ were included. After removing duplicates, the *MLW Corpus* consisted of 290 newspaper articles, spanning the ten-year period of 2008–2017, and totalling 108,925 words and 10,535 types.<sup>3</sup>

Māori loanwords were manually extracted and carefully analysed, as there is currently no software which allows the automatic identification of Māori loanwords.<sup>4</sup> Individual loanwords were classified as being lexical units rather than unique structural words (see Ref. [19]; p. 10). For example, *tangata whenua* (people of the land) and *ka pai* (good) were both considered to be one lexical unit. With the exception of the types *Māori*, *Kiwi*, *Pākehā* and *Matariki*, all non-proper noun loanwords were excluded from loanword frequencies<sup>5</sup>; as were any instances of loanwords which occurred in the search phrases used to construct the corpus, to avoid conflation of the data. Occurrences of code-switching were also discarded from the corpus. The loanword and code-switching debate notwithstanding, we operationalised code-switches as being any lexical material which communicated more than one single concept (in total, we eliminated 1,202 words occurring in 52 articles). Following these exclusions, a total of 3,795 loanword tokens distributed across 186 types were found (see Table 1).

## 5. Results

### 5.1. Flagging typology

We use *flagging* as an overarching term for all explicit loanword demarcation types, including textual (italics, bold face text), oral (pauses, hesitations), body actions (metaphorical quotes made with fingers in the air), or discourse-level (self-correction, explanations, translations). As already mentioned, our paper concerns flagging in written language because we are dealing with newspaper language. Table 2 summarises our proposed typology of flagging practice. We distinguish between *loanword flagging* and *code-switch flagging* on a functional basis; loanword flagging relates to the explicit marking of a loanword, whereas code-switch flagging relates to the explicit marking of a language switch. As discussed in section 2, studies show that specific types of flagging are associated with either loanword or code-switch flagging, for instance, pauses and hesitations or bold font function as code-switch flagging.

#### 2

Table 3 provides some examples, including certain examples which presented difficulties in classification between typographic flagging and flagging at the discourse level.

Purely typographically flagged loanwords tended to be marked for the purpose of either signalling sentence focus (as in ex. 3 below) or signalling direct speech (see ex. 4), rather than for signalling the language of origin.

- (3) There's a push to have ‘āpōpō’ become part of all weather forecasts

<sup>2</sup> *Māori Language Week* has been held annually in New Zealand since 1975 (originating in 1972 as *Māori Language Day*) and aims to celebrate, promote and revitalise the Māori language.

<sup>3</sup> Word and type frequencies were calculated using LancsBox [30].

<sup>4</sup> Writing a script to check online dictionaries does not work well because of the abundance of proper nouns borrowed from Māori which need manual checking.

<sup>5</sup> These high frequency proper noun loanwords were retained in order to better analyse their use.

**Table 2**  
Typology of flagging types.

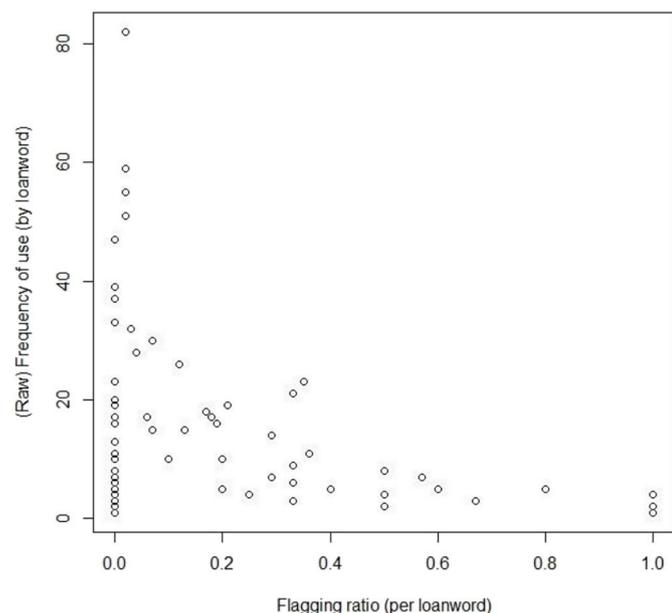
Language Medium	Phonological	Discourse (level)	Typographic	Body Language
<b>Spoken</b>	pauses	self-correction		quotes in the air facial expressions
	accent change	explanation		
	tone hesitations	translation		
<b>Written</b>		commas	bold font	
		self-correction	italic font	
		explanations	quotation marks	
		translation	brackets	
			dashes	

**Table 3**  
Examples of flagging by type.

Type	Example
Typographical flagged loanword	a. Most New Zealanders would acknowledge the Māori language is a <i>taonga</i> that should be protected. b. Most New Zealanders would acknowledge the Māori language is a “taonga” that should be protected.
Flagged loanwords at the discourse level	Most New Zealanders would acknowledge the Māori language is a taonga, a treasure, that should be protected.
Flagged loanwords both at the discourse level and typographically	a. Most New Zealanders would acknowledge the Māori language is a taonga (a treasure) that should be protected. b. Most New Zealanders would acknowledge the Māori language is a treasure – a taonga – that should be protected.
Unflagged loanword	a. Most New Zealanders would acknowledge the Māori language is a taonga that should be protected. b. People are saying nothing but “taonga”

**Table 4**  
Summary of frequencies of flagging in all loanword types.

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



**Fig. 1.** Relationship between ratio of flagging and frequency of use (two outlier loanwords were removed – *Māori* and *te reo* – because of their high frequencies).

(MLW\_2014\_26\_JUL\_NZHS\_002)

(4) Like Jack Tame saying ‘mōrena’ on the Breakfast show

(MLW\_2017\_15\_SEPT\_TDP\_002)

After the removal of data in which flagging was unclear (16 tokens, 11 types), there remained 3,779 loanword tokens (175 types) to analyse. Of these, there were 177 tokens flagged in some way, distributed across 111 types; that is, roughly < 5% of all Māori loanwords in the corpus. Thus, these flagged loanwords only occurred in just over a quarter of all corpus articles (78 of 290).

One final observation is that discourse-level flagging has two possible directions – flagging Māori loanwords with an English description or translation (ex. 5), or flagging native English words by giving a Māori loanword equivalent (ex. 6). The former type is the expected

default as it is used for imported, but not widely used and/or well-known loanwords. The latter type is used in cases where a speaker/writer is actively bringing an incoming, new loanword into a receiver language.

(5) the Māori language is a *taonga* (treasure)

[Māori to English]

(6) the Māori language is a treasure (*taonga*)

[English to Māori]

Our manually coded data shows by and large the expected direction pattern (only 15 from the 177 total flagged tokens were flagged from English to Māori, with a majority of 162 tokens being flagged from Māori to English).

5.2. Flagging, frequency, listedness and core/cultural loanwords

Flagging is not a binary feature of loanword use. In our data, loanwords were relatively evenly distributed across the three possible categories: loanwords which were *always* flagged, loanwords which were flagged in *some articles but not in others*, and loanwords which were *never flagged* (see Table 4, and Fig. 1 for some specific examples of each type). For loanwords which are flagged, they are flagged roughly 20% of the time.

There is also an interesting relationship between ratio of flagging of a loanword and its frequency in the corpus, see Fig. 1. The plot shows that for very frequency loanwords tend to be flagged less often, and conversely, infrequently occurring loanwords are more likely to be flagged. We return to frequency of use in what follows.

However, there was greater variation in token frequency. Consider the continuum of how often a particular loanword is flagged provided in Fig. 1. It is noteworthy to find high frequency loanwords, such as *Māori* (1,649 tokens), *te reo* (language, 947 tokens) and *iwi* (tribe, 82 tokens) flagged at all. This use of flagging questions the degree of integration that these loanwords have in the receiver language. Flagged loanwords also included words listed in the *New Zealand Oxford*

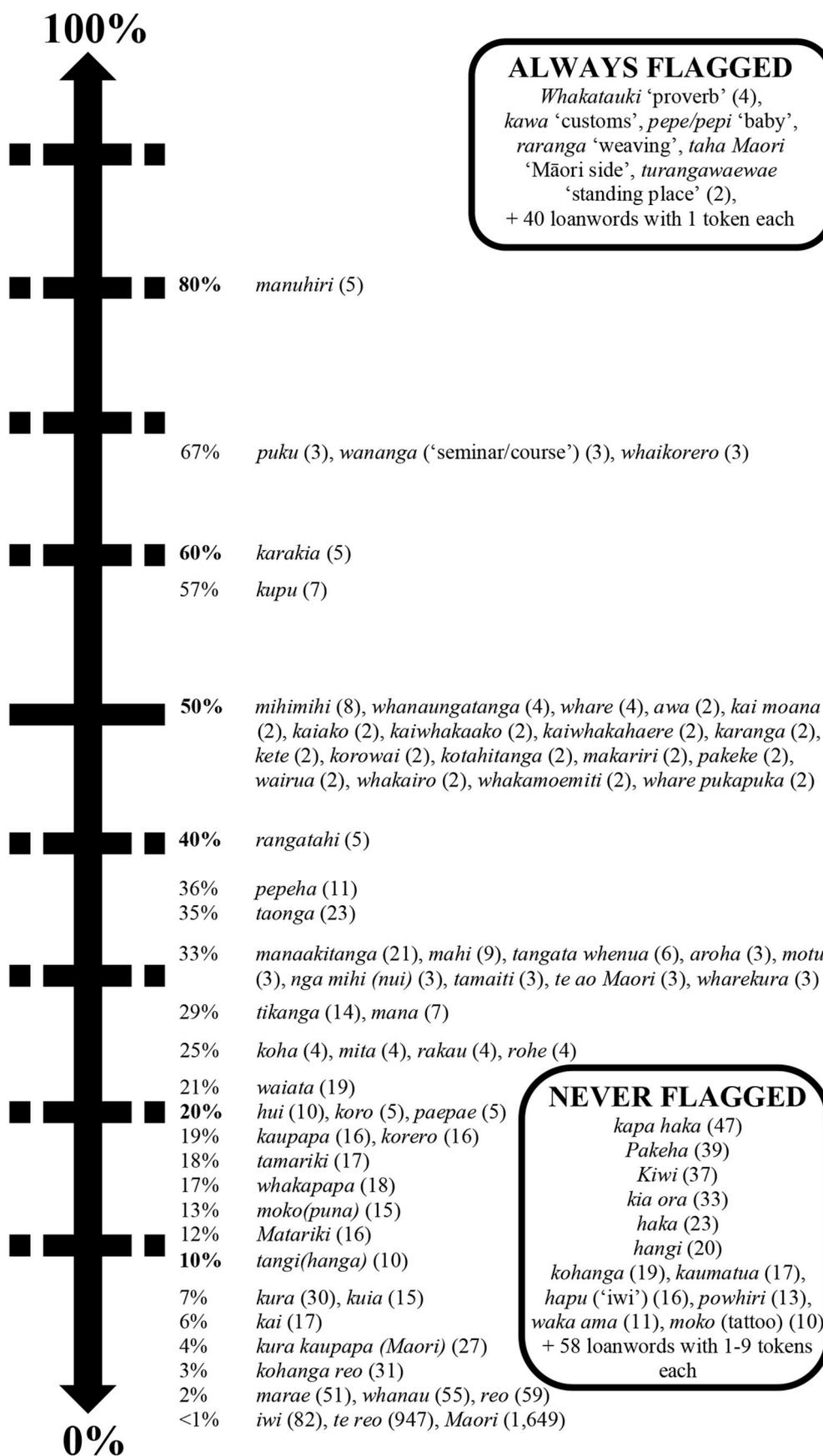


Fig. 2. Percentages of flagged and unflagged loanwords (the number in brackets represents the raw frequency of occurrence of flagged instances, the percentage represents the ratio of flagged versus unflagged instances).

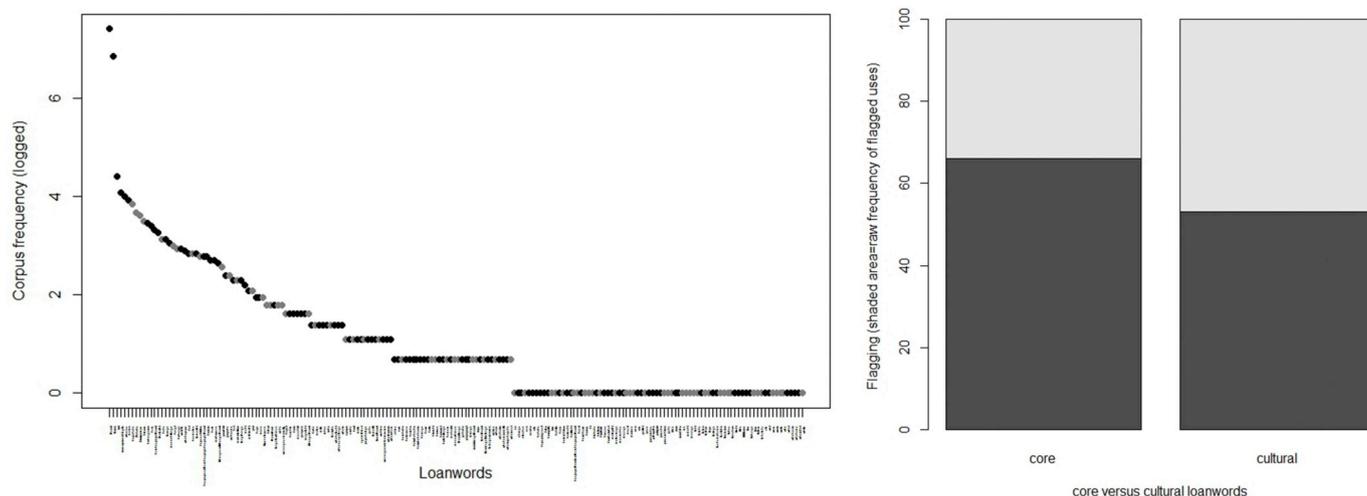


Fig. 3. Panel a. Relationship between flagging and frequency-of-use, (black dots = flagged loanwords; grey dots = unflagged loanwords). Panel b. Relationship between flagging and core/cultural distinction.

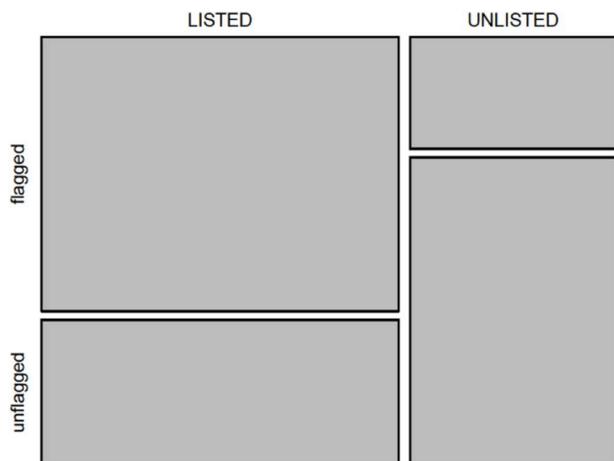


Fig. 4. Flagging and listedness of loanword types mentioned explicitly as known by various authors.

Dictionary (NZOD [31], e.g., *whakatauki* (proverb, saying) and *tūrangawae* (standing place). Conversely, there were unlisted words which were not flagged in our corpus, such as *mōrena* (good morning) and *taringa* (ear). But frequency of use is not likely to only drive flagging rates down (on the rationale that frequency is a proxy for integration in the language), it may – at the same time – also drive flagging rates up because with each use of a loanword, there is a chance for it to be flagged (a loanword cannot be flagged unless it is used in the first place). So frequency is also a control variable here. These observations make the relationship between flagging, frequency-of-use, and listedness in NZE all the more puzzling.

We investigated this relationship by building Generalised Linear Models<sup>6</sup> with a binomial distribution testing whether frequency-of-use and listedness could be significant predictors of flagging. It is important to mention that for this analysis, we treated the flagging parameter as a characteristic of *types* rather than *tokens* (that is, each loanword was classified as either flagged or not flagged, and a single flagged use sufficed to classify it as “flagged” regardless of how many unflagged other uses of it were in the corpus). We also added a new parameter to this list which was found to be relevant to the Māori loanwords in NZE in previous research (see Ref. [34]), namely the core/cultural

distinction (following [35]). The 186 loanword types were each coded for four factors: flagging (yes/no), listedness in the NZOD (yes/no), core/cultural,<sup>7</sup> frequency-of-use (logged). The initial model was trimmed to only include significant factors, which resulted in two (borderline significant) predictors, namely frequency-of-use (Chisq = 244.75, df = 183, p = 0.075), and the core/cultural distinction (Chisq = 247.92, df = 184, p = 0.087).<sup>8</sup> There was no significant difference between the null model and the reduced model (Chisq = 244.75, df = 1, p = 0.526). While as expected, frequently used loanwords tended to be unflagged (Fig. 2a), core loanwords were flagged more frequently than cultural loanwords (Fig. 2b).

In other words, loanword flagging is weakly predicted by frequency effects, but also by whether a loanword *can* be translated or explained (i.e., being a ‘core’ word with some alternate counterpart in RL), or alternatively, by whether the word is deemed (especially) important to be clarified, given its potential wide use. One possible confounding factor in our analysis is the missing variable of author – as flagging may be intimately tied with author (and editor) perception (see section 5.3), the model should also include authors (as a random variable). However, this would only be possible for a token analysis (and not a type analysis), which in our case would present serious data sparsity problems. A second problem is also skewness (the data is highly skewed towards the two most frequently occurring loanwords, *Māori* and *te reo*).

### 5.3. Flagging and perception

There were a number of loanwords in the corpus which authors explicitly mentioned as being widely understood and used by New Zealanders. These provided the opportunity for an analysis which considers speaker’s perception and awareness in relation to language use, akin to folk linguistic analysis [36]. By comparing the frequencies of both listedness and flagging in these mentioned loanword types, it was surprising to find that the majority of flagged loanwords were in fact listed; while unlisted loanwords were more likely to appear unflagged than flagged (Fig. 3).

This runs counter to the expectation that listed (and therefore established loanwords) would be less likely to be flagged, if at all. It speaks to the fact that flagging is not functioning as an indicator of the

<sup>7</sup> As alternative, we also coded our data for catachrestic/non-catachrestic distinctions, following [33], but this factor did not turn up to be significant in our model.

<sup>8</sup> We tested for an interaction between frequency-of-use and core/cultural distinction but found none (Chisq = 244.64, df = 182, p = 0.740).

<sup>6</sup> All modelling was done using R [see Ref. [32]].

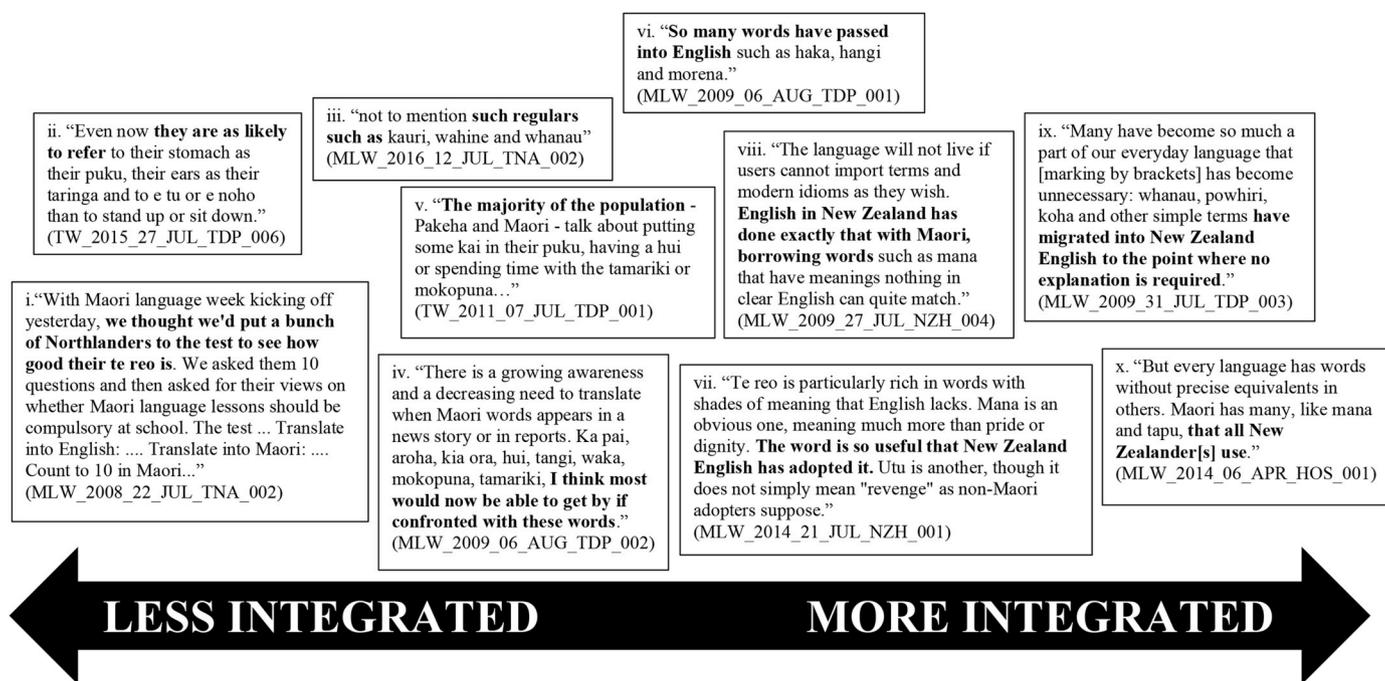


Fig. 5. Authors' perceived continuum of loanword integration (bolding is our own for ease in comprehension).

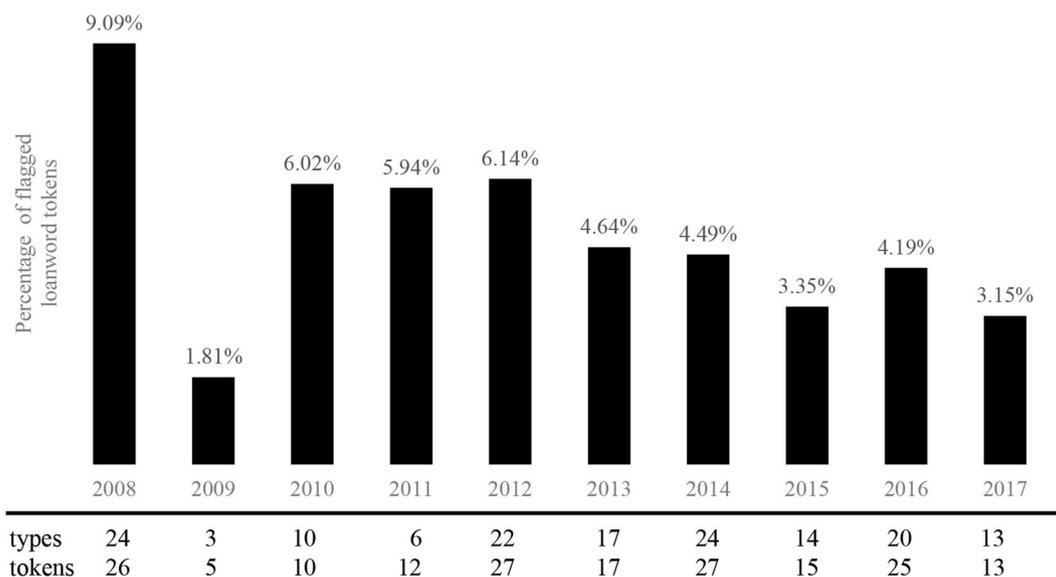


Fig. 6. Flagged loanwords by year: the bar graph provides percentages of flagged loanword tokens and the table under the bars gives the raw numbers of types and tokens.

integration of Māori loanwords as they pertain to NZE. Rather, it seems likely that author perceptions influence the likelihood of flagging of loanwords. The subjectivity of these perceptions can be seen in Fig. 4, where explicitly mentioned loanwords are ranked according to how integrated they are perceived to be in NZE by the newspaper authors (in their own metalanguage), and placed on a continuum of author perceived integration. Take for instance, the writer who states “not to mention such regulars as kauri, wahine and whanau” (MLW, 1 July 2016, in TNA). For this author, the Māori words are “regulars” in English thereby implying not just general familiarity of the New Zealand public with these words, but also frequent use of these loanwords. This is explicitly spelled out by another writer: “Māori has many, like mana and tapu, that all New Zealander[s] use” (MLW, 31 July 2009, in TDP). Borrowing Māori words into English is also not just something that happens, but something that happens for good reason: “the [loan]word

is so useful that New Zealand English has adopted it” (MLW, 21 July 2014, in NZH), and “every language has words without precise equivalents in others. Māori has many, ...” (MLW, 6 April 2014, in HOS). This perspective appeals to logical reasoning in a bid to validate the use of Māori loanwords, namely by means of the lexical richness that they bring with them. The very fact that their use might require validation betrays their slightly controversial status – not everyone is convinced by their use and some readers will be suspicious of the language ideology that these loanwords represent; a fact that newspapers writers are highly sensitive to.

Interestingly, Fig. 4 also shows that authors seem to provide rather different examples of the kinds of Māori loanwords that they feel the NZ population is familiar with. While writers agree that many (or all) New Zealanders used Māori loanwords, they are not in much agreement regarding which actual loanwords are used.

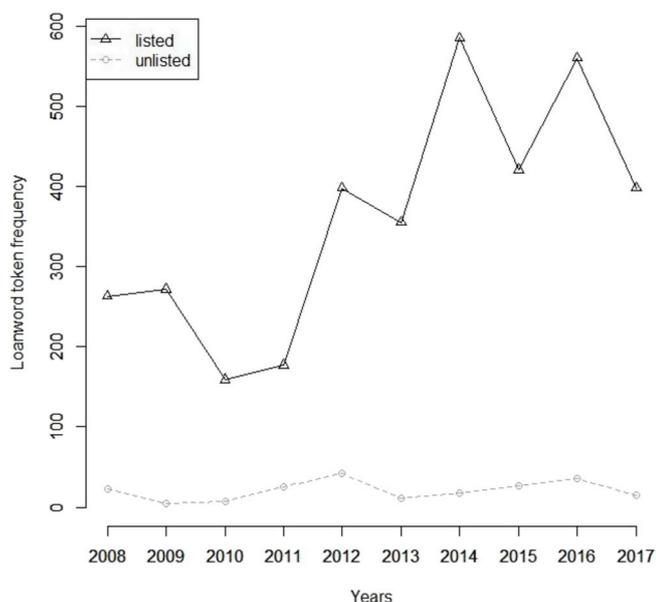


Fig. 7. Listed (solid line) and unlisted (dashed line) loanword token frequencies per year.

There was no detectable correlation between author perception and loanword use over time. Moreover, regarding the specific examples of loanwords mentioned by the various authors as being largely familiar to New Zealanders, we could also not detect a diachronic increase (of these individual loanwords), which may be a result of the small data set (50 loanword tokens, 41 types across 10 articles).

#### 5.4. Flagging and diachrony

Next, we report diachronic patterns of loanword use in articles concerning *Māori Language Week*, as observed in the *MLW Corpus*. There was a statistically significant increase in the use of loanwords over the ten-year period (Kendall Tau test,  $\tau = 0.6$ ,  $p = 0.020$ ), with an overall average occurrence of 34.84 per 1,000 English words. This was 4.5 times greater than the most recent study of the same in newspapers: 7.7 per 1,000 words, done in 2000 [19]. The increase in loanword use is matched by a significant increase in unflagged loanwords over the ten-year period ( $\tau = 0.511$ ,  $p = 0.049098$ ). But when frequencies were normalised per 1,000 words (per year), flagged loanwords showed a statistically significant decrease in occurrence over the same period ( $\tau = -0.511$ ,  $p = 0.0491$ ), see Fig. 5.

The simultaneous significant increase in the use of loanwords, combined with the decrease of the frequency of flagged loanwords suggests that not only are more loanwords being used in NZE, they are also being treated as belonging to the New Zealand English lexicon; that is, more integrated.

These patterns go hand in hand with an increase in listed loanwords over this period ( $\tau = 0.539$ ,  $p = 0.038879$ ), see Fig. 6. Because we are observing these patterns all at once, it is impossible to tell which (if any) is the driver. Although more loanwords are being used and fewer are being flagged, it appears that the increase in loanwords coincides with an increase in listed loanwords. Thus, the decrease in flagged loanwords is likely the natural result of an increase in the reusing of certain already-known loanwords. Fig. 6 shows that few unlisted loanwords occur in newspapers, and despite the overall increase in loanwords, this increase does not involve unlisted loanwords (there was a slight diachronic increase, but this was not statistically significant,  $\tau = 0.289$ ,  $p = 0.28313$ ) (see Fig. 7).

Looking diachronically across the use of loanwords suggests a stabilization of the stock of loanwords being imported from Māori, with the same familiar (listed) words being used productively. However,

recall that previous research finds the use of loanwords to be intimately linked to certain topics of discourse and to certain speakers/writers. Yet these writers appear to treat loanwords more and more as though they were largely known by their entire readership (as indicated by the reduction of flagging), even though, this may not be the case. We interpret the assumption of widespread familiarity with loanwords to be informed by an ideological stance, rather than linguistic reality.

## 6. Concluding remarks

Our findings suggest that the flagging patterns analysed in the *MLW Corpus* are not predicated by listedness, nor are they entirely the product of frequency (some frequently occurring loanwords are flagged; conversely, some rarely used loanwords are not), yet there is a weakly significant relationship between flagging and frequency-of-use. There is also a relationship between flagging and semantic properties of the loanwords: core loans are more likely to be flagged than cultural ones. Added to this mix is the perception of authors – ultimately it is they (and possibly newspaper editors) who decide whether to flag a loanword or not. A closer look at these perceptions appears to indicate a mismatch between listedness and flagging. Our work suggests that flagging does not seem to be wholly a reflection of linguistic factors, as there is more going on than just use, familiarity, and integration. One aspect that we do not touch on here is individual preferences of authors, and the amount of variation which exists among these (i.e., are there “flaggers”, that is, writers who often flag loanwords, and “non-flaggers”, writers who tend not to flag any loanwords?).

We suggest that loanword use in NZE remains tied up with personal identity (not everyone uses the loanwords) and with language ideology (the belief that using loanwords is “useful” in English as expressed by some writers), but also the belief that say, using Māori bilingual place-names is a step towards the revitalisation of Māori as discussed in Ref. [37]). Echoing earlier proposals made by previous research (see Refs. [8,38] and others), we see flagging as an indicator of author perception and author’s intention regarding a particular loan. Authors tend to neither use loanwords to fill a linguistic gap in their knowledge (most are likely monolingual), nor in the receiver language (English); although, paradoxically, this is cited as a reason to borrow words from Māori in the first place. They use such words to signal a given political and social stance towards an ethnic group (Māori), or to actively encourage and promote the maintenance of the Māori language. They perceive (sometimes correctly and sometimes not) certain loanwords as being widely known and go on to treat them as such (they do not explain or translate them). This perception can be gleaned from the increasing diachronic trend of using more loanwords matched by a decreasing trend of flagging fewer of these. Some authors are quite clearly aiming to teach Māori words to an audience they perceive to be unfamiliar with the loans (see ex. 7). In this example, the writer flags the loanwords *kawa*, *pepeha* and *noho marae*, which are indeed neither listed, nor frequent (*marae* is both, but not the expression *noho marae*). They also flag the loan *reo*, which is both listed and highly frequent, and yet do not flag the loanwords *mahi* or *wananga* (which are both, neither frequent; nor listed). The decision of what to flag seems arbitrary here. All this points to the fact that loanword use is heavily influenced by sociolinguistic factors rather than by linguistic factors alone.

(7) ‘As the year went by, we tried hard to do the mahi. We learnt so much of the kawa [customs] and reo [language] and meaning behind the words,’ she said. ‘I loved the culture of the wananga, you could make mistakes and laugh at it, you can self-correct and just have a go. I loved the noho marae [marae stay], I got to say my pepeha [tribal belonging speech] and I was shaking.’

(TW\_2014\_23\_JUL\_TDP\_003)

Our data does not support work which directly links flagging to integration per se, and we think this is owing to the dynamics found in an asymmetrical loanword movement contact situation. Because Māori

is endangered and currently in the midst of a strong wave of revitalisation, loanword use in New Zealand is a highly sociolinguistically charged linguistic act. Our findings endorse Jones [11], who cautions that when threatened languages are involved, loanword use must be analysed by taking into account the wider sociolinguistic context, and in particular the speakers' attitude.

## Funding

This work was supported by the Marsden Fast Start Fund, Royal Society of New Zealand [16-UOW-015].

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.amper.2019.100056>.

## References

- [1] H. Sharp, Swedish-English language mixing, *World Englishes* 26 (2) (2007) 224–240.
- [2] P. Grant-Russell, C. Beaudet, Lexical borrowings from French in written Quebec English, *University of Pennsylvania Working Papers in Linguistics [Selected Papers from NWAV 27]*, vol. 6, 1999, pp. 17–33 2.
- [3] S. Fiedler, Phraseological borrowing from English into German: cultural and pragmatics implications, *J. Pragmat.* 113 (2017) 89–102.
- [4] S. Poplack, N. Dion, Myths and facts about loanword development, *Lang. Var. Change* 24 (2012) 279–315.
- [5] H. Kruger, *Postcolonial Polysystems: the Production and Reception of Translated Children's Literature in South Africa*, John Benjamins Publishing Company, Amsterdam, 2012.
- [6] S. Poplack, S. Wheeler, A. Westwood, Distinguishing language contact phenomena: evidence from Finnish–English bilingualism, *World Englishes* 8 (3) (1989) 389–406.
- [7] D. Turpin, 'Le français, c'est le last frontier': the status of English-origin nouns in Acadian French, *Int. J. Biling.* 2 (2) (1998) 221–233.
- [8] C. Davies, M. MacLagan, Māori words – read all about it: testing the presence of 13 Māori words in four New Zealand newspapers from 1999 to 2004, *Te Reo* 49 (2006) 73–99.
- [9] M. Degani, The Pakeha myth of one New Zealand/Aotearoa: an exploration in the use of Maori loanwords in New Zealand English, in: R. Facchinetti, D. Crystal, B. Seidlhofer (Eds.), *From International to Local English – and Back Again*, Peter Lang, Frankfurt am Main, 2010, pp. 165–196.
- [10] J. Macalister, Investigating the changing use of te reo, *NZ Words* 13 (2009) 3–4.
- [11] M.C. Jones, Some structural and social correlates of single word intrasentential code-switching in Jersey Norman French, *Fr. Lang. Stud.* 15 (2005) 1–23, <https://doi.org/10.1017/S0959269505001894>.
- [12] P. Muysken, *Bilingual Speech: A Typology of Code-Mixing*, Cambridge University Press, Cambridge, 2000.
- [13] C. Myers-Scotton, *Duelling Languages: Grammatical Structure in Codeswitching*, Clarendon Press, Oxford, 1993.
- [14] S. Poplack, D. Sankoff, Borrowing: the synchrony of integration, *Linguistics* 22 (1) (1984) 99–135.
- [15] J. Treffers-Daller, Evidence for insertional codemixing: mixed compounds and French nominal groups in Brussels Dutch, *Int. J. Biling.* 9 (3–4) (2005) 477–508.
- [16] J.R. Stammers, M. Deuchar, Testing the nonce borrowing hypothesis: counter-evidence from English-origin verbs in Welsh, *Lang. Cogn.* 15 (3) (2012) 630–643.
- [17] C. Myers-Scotton, *Codes and Sequences: Choosing Linguistic Varieties*, Oxford University Press, Oxford, 1998.
- [18] P. Moon, *Turning Points: Events that Changed the Course of New Zealand History*, New Holland Publishers Ltd, Auckland, 2013.
- [19] J. Macalister, The Maori presence in the New Zealand English lexicon, 1850–2000: evidence from a corpus-based study, *Engl. World-Wide* 27 (1) (2006) 1–24.
- [20] P. Durkin, *Borrowed Words: A History of Loanwords in English*, Oxford University Press, Oxford, 2014.
- [21] T. Deveson, New Zealand English lexis: the Māori dimension, *Engl. Today* 28 (1991) 18–25.
- [22] F. van Coetsem, *A General and Unified Theory of the Transmission Process in Language* Winter, (2000) Heidelberg.
- [23] F. van Coetsem, *Loan Phonology and the Two Transfer Types in Language Contact* Foris, Dordrecht, 1988.
- [24] S. Poplack, Variation theory and language contact, in: Dennis R. Preston (Ed.), *American Dialect Research*, John Benjamins, Philadelphia, 1993, pp. 252–286.
- [25] G. Kennedy, S. Yamazaki, S. The influence of Maori on the New Zealand English lexicon, in: J.M. Kirk (Ed.), *Corpora Galore: Analyses and Techniques in Describing English: Papers from the Nineteenth International Conference on English Language Research on Computerised Corpora (ICAME 1998) [Conference Proceedings]*, Editions Rodopi B.V., Amsterdam, 2000, pp. 33–44.
- [26] M. Degani, A. Onysko, Hybrid compounding in New Zealand English, *World Englishes* 29 (2) (2010) 209–233.
- [27] G. Kennedy, Lexical borrowing from maori in New Zealand English, in: B. Moore (Ed.), *Who's Centric Now? the Present State of Post-Colonial Englishes*, Oxford University Press, Canberra, 2001, pp. 59–81.
- [28] A. Onysko, A. Calude, Comparing the usage of Māori loans in spoken and written New Zealand English: a case study of Maori, Pakeha, and Kiwi, in: E. Zenner, G. Kristiansen (Eds.), *New Perspectives on Lexical Borrowing: Onomasiological, Methodological and Phraseological Innovations*, de Gruyter Mouton, Berlin, 2014, pp. 143–169.
- [29] A. Onysko, Anglicisms in German: Borrowing, Lexical Productivity, and Written Codeswitching, Walter de Gruyter, Amsterda, 2007.
- [30] V. Brezina, T. McEnery, S. Wattam, Collocations in context: a new perspective on collocation networks, *Int. J. Corpus Linguist.* 20 (2) (2015) 139–173.
- [31] T. Deveson, G. Kennedy, *The New Zealand Oxford Dictionary*, Oxford University Press, Oxford, 2005.
- [32] R. R Core Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013 <http://www.R-project.org>.
- [33] A. Onysko, E. Winter-Froemel, Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing, *J. Pragmat.* 43 (6) (2011) 1550–1567.
- [34] A.S. Calude, M. Pagel, S. Miller, Modelling borrowing success – a quantitative study of Māori loanwords in New Zealand English, *Corpus Linguist. Linguistic Theory* 15 (2) (2017) 1–28, <https://doi.org/10.1515/clt-2017-0010>.
- [35] C. Myers-Scotton, *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*, Oxford University Press, Oxford, 2002.
- [36] N.J. Albury, Defining Māori language revitalisation: a project in folk linguistics, *J. Sociolinguistics* 20 (3) (2016) 287–311.
- [37] N.J. Albury, L. Carter, A typology of arguments for and against bilingual place-naming in Aotearoa New Zealand, *J. Multiling. Multicult. Dev.* 38 (9) (2017) 831–842.
- [38] J. Macalister, Weka or woodhen? Nativization through lexical choice in New Zealand English, *World Englishes* 26 (4) (2007) 492–506.