# *Modelling loanword success – a sociolinguistic quantitative study of Māori loanwords in New Zealand English*

**Open Access**

Andreea Simona Calude*, Steven Miller and Mark Pagel

# Modelling loanword success – a sociolinguistic quantitative study of Māori loanwords in New Zealand English

**Abstract:** Loanword use has dominated the literature on language contact and its salient nature continues to draw interest from linguists and non-linguists. Traditionally, loanwords were investigated by means of raw frequencies, which are at best uninformative and at worst misleading. Following a new wave of studies which look at loans from a quantitatively more informed standpoint, modelling "success" by taking into account frequency of the counterparts available in the language adopting the loanwords, we propose a similar model of loan-use and demonstrate its benefits in a case study of loanwords from Māori into (New Zealand) English. Our model contributes to previous work in this area by combining both the success measure mentioned above with a rich range of linguistic characteristics of the loanwords (such as loan length and word class), as well as a similarly detailed group of sociolinguistic characteristics of the speakers using them (gender, age and ethnicity of both, speakers and addresses). Our model is unique in bringing together of all these factors *at the same time*. The findings presented here illustrate the benefit of a quantitatively balanced approach to modelling loanword use. Furthermore, they illustrate the complex interaction between linguistic and sociolinguistic factors in such language contact scenarios.

**Keywords:** loanwords, multivariate modelling, New Zealand English, sociolinguistics, Maori

**\*Corresponding author: Andreea Simona Calude,** School of General and Applied Linguistics, University of Waikato, Knighton Rd Private Bag 3105, Hamilton, Waikato, New Zealand, E-mail: andreea@waikato.ac.nz

**Steven Miller,** Department of Statistics, University of Waikato, Hamilton, New Zealand

**Mark Pagel,** School of Biological Sciences, University of Reading, Reading, Berkshire, UK

# 1 Introduction

Even the most seemingly isolated places on earth are not completely alone, without contact to any other cultures or languages. Just as "no man is an island", as famously coined by Donne, one might also say that no language is a [linguistic] island. Contact between languages, or more precisely, contact between speakers of two or more languages (Milroy 1997: 311), has formed the basis of a vast amount of research.

An obvious consequence of language contact is the borrowing of lexical material between the languages involved, typically words, but sometimes also phrases. The borrowing process has been written about under various labels, such as "interference" (Weinreich 1953), "code copying" (Johanson 1993), and "transference" (Clyne 2003). We use the terms *loanwords* and *lexical borrowings* interchangeably here to denote words which originate from a given language (donor language) and which enter into and are productively used within a distinct language (host language). Efforts have been made to study the phenomenon from both a synchronic and diachronic perspective (see Treffers-Daller 2010 for an overview), and consequently, much has been learned about it. Of course, some questions remain unanswered; such as how to disentangle code-switching from borrowing, and at what point in the contact situation loanwords are most likely to enter a host language.

In this paper, we hope to increase understanding of the borrowing process by implementing a novel methodological approach to the study of loanwords which offers a possible solution to a long-standing problem. Studies of loanwords typically gauge borrowing rates by reporting raw frequencies of use of loanwords (Imm 2009; Kouega 2009; Furiassi 2011; among others). This is problematic because a loanword's use depends not only on the very act of being borrowed from one language into another, but also on a speaker's (or writer's) desire to use the concept that the word denotes. Given that word frequency of use is highly skewed (Zipf 1935 and many others), comparing raw frequencies of use of loanwords may be rather meaningless. Specifically, raw frequencies of use only convey *actual* use of a loan, without taking into account its *potential* use in terms of opportunities available to a speaker to use a word denoting the concept in question. Crucially, certain words and concepts can become more widely used because they might be relevant to certain topics of conversation, thereby prompting speakers to be more sensitive to such words. This issue is completely ignored by raw frequencies and as such, constitutes an important factor that needs to be addressed in the study of loanwords.

Here, we present a model of borrowability which captures a measure of the *potential* use of a loan in order to provide a more accurate picture of a given loanword's frequency. While this idea is not completely new, there are only a handful of studies which actually operationalise it, as will be discussed in Section

2. Additionally, our model is unique in that it takes into account at the same time *both* linguistic *and* socio-linguistic factors relevant to the borrowing process. We illustrate this with a case study of borrowings from the Austronesian language of Māori into (New Zealand) English. The model is used to test for significant predictors of *loanword success* by incorporating factors identified by previous studies as being relevant to the borrowing process, such as loanword length and phonological fit of loans inside a host language, but also social attributes of the speakers using the loans, namely, age, gender and ethnicity (in a large-scale data comprising more than 700 participants). We define **loanword success** as being the chance that a given loanword $W_1$ (originating from a donor language) has to be used within a host language compared to an existing alternative word $W_2$ (or $W_3$, $W_4$, … ), when controlling for the number of opportunities that speakers of the host language have to use $W_1$ and $W_2$ (or $W_3$, $W_4$, … ). Our approach brings together all these factors simultaneously in a statistical model which takes into account speaker variation by using a Generalized Linear Mixed-effects Model (GLMM).

# 2 Measuring loan success

It has been shown that while virtually any part of the language system can be borrowed, some categories are more likely to be borrowed than others (for example, nouns are borrowed more frequently, prepositions less so, and suffixes and inflections even less so). Investigations of borrowing constraints have ultimately led to a borrowability probability scale (Thompson and Kaufman 1988/1991). A study by Poplack et al. (1988) investigating English loans into French found that nouns are statistically more likely to be borrowed in comparison to other lexical categories, with verbs and adjectives following (1988: 94).

Poplack et al. (1988) also found that the loans which were more likely to thrive in French were those integrated into the French lexicon in some fashion, either nouns that were assigned a gender in accordance with French syntax rules, or verbs which received French inflection marking. This finding does raise the cause-and-effect question of which came first: their integration (causing them to be used more frequently), or their frequent use (causing them to be more readily integrated). An earlier experimental study by Poplack and Sankoff (1984) found the same was true for English borrowings into Puerto-Rican Spanish. In general, not all loanwords are equally "foreign" to a given host language. Some loanwords involve sounds or sound combinations which are not found in the host language, for example the English loan "job" as borrowed by German contains the non-native German sound [dʒ]. In such cases, the loan can (though it need not) become "integrated" in the recipient language by

undergoing phonological replacement, where the non-native sound is replaced with the nearest available native sound. Phonological integration is, of course, not the only way that loans can become integrated in the host language (see for example, details what we might call 'semantic integration' in Macalister 2008). Loans that become phonologically and morphologically integrated tend to have better relative success in the host language.

The likelihood of a loan becoming phonologically and morphologically integrated in the host language varies from loan to loan, and from speaker to speaker (in particular, with speaker's age and bilingual ability, see Poplack et al. 1988 and Thomason and Kaufman 1988/1991; although this can also vary across communities, see Poplack and Sankoff 1984; for an example where both adults and children behaved the same way with regards to loanword use, despite having different bilingual abilities, the adults were Spanish-dominant, whereas the children were English-dominant). As regards the loan itself, this likelihood of integration is affected by the age of borrowing. The longer a loanword is used, the more pressure there is on it to integrate, to the point where speakers of the host language are no longer aware of the "foreign" status of the loan (see effects of this nature described in Haspelmath and Tadmor's *Word Loanword Database*, where a word's likelihood of having been borrowed is coded on a scale of 1–5, rather than as a binary feature, cf. Haspelmath and Tadmor 2009).

As regards the speakers' role in the borrowing process, somewhat surprisingly, Poplack et al. (1988) discovered that an individual's proficiency in the host language did not play as important a role as might have previously been assumed. This is surprising because early accounts of lexical borrowing attributed this to speaker's lack of proficiency in a new language, which they reportedly handled by bringing in words from the language(s) which they had comparatively higher proficiency in. However, Poplack et al. (1988) discovered that community norms were much more important than language proficiency (1988: 97–98). In their study, English loans were used more frequently by inhabitants of neighbourhoods that had high contact with English, speakers who had low occupational status (as opposed to high occupational status), men (more commonly than women), and those that were younger (15–35 years old, at any rate no older than 40).

Following on from Poplack et al.'s (1988) earlier work, Van Hout and Muysken (1994) showed that for Spanish borrowings into Bolivian Quechua, high frequency of occurrence of the loans in Spanish (donor language) correlated with high chances of borrowability into Bolivian Quechua (recipient/host language). Inhibitors to borrowability were highly inflected forms in both the donor and recipient languages, loans that were paradigmatically organised (that is, forms whose semantics is systematically divided up in a linguistically-specific

manner, cf. 1994: 55), and loans that were tightly involved in the structuring of the clause (as opposed to discourse markers, interjections and other clause-peripheral material which can be easily integrated in the discourse).

A diachronic study of lexical borrowing in French by Chesley and Baayen (2010) provides yet another perspective of loanword success. The authors searched for loanwords from several languages (not just one), including English, Spanish, and German (among others) with the aim of predicting loan-word success by comparing two corpora, ten years apart. They found that loans which were still prevalent in the latter corpus were shorter, more polysemous, frequently used in more contexts (that is, they exhibited higher dispersal), and more likely to occur in culturally non-restricted contexts (Chesley and Baayen 2010: 1368).

Building on earlier qualitative studies of loanwords which take into account the use of potential counterparts which already exist in the source language (Poplack and Sankoff 1984; Humbley 2008; and Graedler and Kvaran 2010), recent work on English loans denoting human roles (e.g., *nanny* and *backpacker*) into Dutch, Zenner et al. (2012) compared the extent to which English loans compete against Dutch counterparts (*kinderjuffrouw* and *rugzakker* or *rugzak-toerist*, respectively). In order to investigate the influencing factors of success, they coded each loan for lectal features (the dialect of Dutch which the native equivalents belonged to), the type of newspapers they occurred in (local or regional), word related features (era of borrowing, and length of the loan), and conceptual features (whether or not the loan was a necessary or a luxury loan, the frequency of the concept in the recipient language, and the lexical field of the loan, e.g., media and IT, sports and leisure, making money, social life). Their findings suggest that loans which designated a new concept in Dutch were more successful, as were loans that had been around for longer, and loans that were phonologically shorter than their Dutch equivalents. Finally, loans from lexical fields more closely associated with the Anglo-American culture were more successful overall.

Comparative word length was also found to be a significant predictor of English borrowings in Spanish, according to Shin (2010). Shin's analysis of English lexical insertions within a corpus of bilingual Latinos in New York uncovered that "while the tendency to shorten words is apparent in all communicative settings, in situations of language contact word-shortening can take the form of borrowing comparatively shorter words from a donor language to replace comparatively longer words in a recipient language" (2010: 56).

Winter-Froemel, Onysko and Calude (2012) investigated English loans in German and found that word length, age of borrowing, and lexical field played a significant role in the relative success of the loans analysed.

Finally, we will see below that Onysko and Calude (2014) found that social characteristics of the speakers using the loans analysed were also statistically relevant in a small case-study of three Māori loanwords (*Māori, Kiwi* and *Pākehā*) in New Zealand English.

These studies point to two important observations. First, both social factors as well as linguistic factors come into play in the study of loanword success, and ideally, models of loan-use ought to take into account both types of factors. It is only by controlling for all these factors that we can get a clearer picture of what is influencing the success of the loans and of the complex interplay between linguistic and social factors (for example, the factors which influence loan success in a given recipient language and within a certain community might be different across the different sub-groups of that community). Secondly, the exact factors that may be relevant in influencing loanword success tend to vary with each language contact situation. For now, more case studies are needed to provide a more complete set of factors which apply (more or less) across the board (though perhaps in different ways, depending on the languages involved). Our study addresses both of these issues by presenting a model of lexical borrowing that involves social as well as linguistic factors (at the same time), and by contributing to the current body of case studies of loanword success.

# 3 The background of contact situation

## 3.1 Linguistic context of New Zealand

The indigenous Māori language (Te Reo Māori) was once a dominant and widespread language of New Zealand (or Aotearoa "The Land of the Long White Cloud", according to its Māori name). The arrival of the Europeans culminating in the signing of the *Treaty of Waitangi* in 1840 would have, in time, a strong effect on the livelihood of the Māori language. Over time, its status gradually declined, losing numbers of speakers and prestige, and suffering greatly as a result of two major changes within New Zealand society: the replacement of Māori with English as the language of the classroom, and the urbanisation of the Māori population.

By the 1970s, concerns for Te Reo Māori became more apparent which instigated revitalisation efforts (cf. Māori Language Commission). These concerns culminated in a change of legislation. In 1987, the *Māori Language Act* was passed recognizing Te Reo Māori as an official language of New Zealand/ Aotearoa and forging a powerful front to revitalise Māori language and culture.

Māori immersion preschools (*kōhanga reo*) and Māori immersion schools (*kura kaupapa*) were set up to encourage the learning of Māori in early childhood. This programme proved highly successful. Radio stations broadcasting in Māori took to the airways from 1987, with the first one broadcasting from Wellington, and the first Māori language TV station came into being in 2004, with a second following in 2008. This period has become known in New Zealand as the *Māori Renaissance* period (see Benton 1991 for an in-depth account of the general situation of Māori and Te Reo Māori in New Zealand).

The revitalisation efforts have paved the way for considerable improvement of the situation of Te Reo Māori (the Māori language), and the language seems to be recovering from its initial loss of vitality (though the battle is by no means won). The latest census from 2013 found that 257,500 (55 %) of Māori adults could hold a conversation about everyday things in Māori (cf. *Te Puni Kōkiri* and *Statistics New Zealand*). These figures exhibit an increase from the 153,500 (42 %) of Māori adults reported the 2001 Census. Although racial tensions are still reported, the "Survey of Attitudes" toward the Māori language, undertaken by *Te Puni Kōkiri* in 2006 found that Te Reo Māori enjoys positive attitudes and high status in both Māori society and a great majority of non-Māori New Zealanders.

The close contact between the Māori speaking and English speaking populations has resulted in changes in both languages. English pronunciation, grammar and vocabulary have influenced Māori, and conversely, many borrowings from Māori have made their way into English vocabulary (e.g., the current online version of the Oxford English Dictionary 3[1] lists 287 entries of Maori origin).

## 3.2 The study of Māori loans in New Zealand English

Unsurprisingly, the flow of Māori loans into New Zealand English has not gone unnoticed. According to Deverson, "the most unmistakably New Zealand part of New Zealand English is its Māori element" (1991: 19). Macalister sums up the borrowing situation in recent times as follows: "it is likely that the Māori presence in New Zealand English will continue to grow in future, and that this presence will continue to define the distinctiveness of New Zealand English lexicon" (2006a: 21). Below, we provide a brief history of the linguistic interest in the use of words of Māori origin in New Zealand English.

The first wave of borrowings from Māori during the late 18th and early 19th century consisted primarily of words that describe environmental and natural

---

1 Source: http://www.oed.com/ (accessed 26 January 2017).

terms (such as, *pāua* "abalone", *kūmara* "sweet potato", *rimu* "red pine"), Māori place names (*Tauranga, Rotorua, Whanganui*), and some indigenous culture terms (*marae* "Māori meeting house", *hāngi* "earth oven to cook food with steam and heat from heated stones").

This first wave of borrowings gave way to further "waves" (Macalister 2006a), bringing in more Māori words such as *waka* "canoe", *taonga* " treasure", *iwi* "tribe", *kuia* "female elder", *koha* "gift/donation", *hui* "meeting" and so on. This time, the words had to do with the organisation of Māori society and culture, and important historical events. Deverson points out an important difference between the first wave of Māori loans and subsequent ones: "while colonial borrowing from Maori was Pakeha-driven,[2] motivated by the European's need to come to terms with a strange world, the recent revival or new wave of borrowing is by contrast Maori-driven, initiated in large part by Maori speakers and writers themselves" (1991: 20). Some of the motivations identified as driving the use of Māori loans today have to do with filling semantic gaps in existing vocabulary of New Zealand English, economy of expression, expression of identity and display of empathy, clarity of meaning, and language play (Macalister 2007a). However, neither the motivations for Maori loan use, nor the categories of loans identified are based on statistical modelling of any kind.

Following a large body of work, Macalister (2001; 2006a; 2006b; 2007a; 2007b; 2008; 2009) argues that the number of Māori tokens continues to rise steadily, with about six out of every thousand words uttered being of Māori origin (this figure was also corroborated by Kennedy and Yamazaki 1999). In his 2008 paper, he notes that "Māori word presence in New Zealand English has been increasing for almost 40 years, reflecting social and cultural changes since around 1970" (Macalister 2008: 76). While proper nouns still constitute the vast majority of loans, the range of borrowed types has also increased, as has the number of uses per type (Macalister 2006a, b). Interestingly, current literature is moot on the adoption of any recent loanwords achieving high frequencies (Durkin 2014: 394).

The pervasiveness of the loans was noted in both spoken and written New Zealand English (Kennedy and Yamazaki 1999), in newspaper media (Macalister 2006a; 2006b; Davies and Maclagen 2006; Degani and Onysko 2010; Degani 2010), in children's books (Daly 2007), and in the work of many prominent New Zealand novelists (Keri Hulme, Witi Ihimaera, Alan Duff, Patricia Grace). Investigating the productivity of loans, Degani and Onysko (2010) found that many well-established loans do indeed "enter into productive processes of word

---

**2** Pākehā is a term referring to European or white New Zealanders (the use of the macron was not standardized until later and Deverson does not use it in his text).

formation" (2010: 231). The use of Māori loans in TV media was studied by De Bres (2006) and unlike other language mediums, de Bres found that Māori loans were used only to a "limited extent in the mainstream television news" and "in highly restricted areas", "almost solely in Māori-related news items" (2006: 32). This restricted usage was also reported by Degani (2010) with respect to three Māori loans, namely *aroha* "love", *mana* "power/respect", and *marae* "meeting house" in three New Zealand newspapers. However, De Bres (2006) did find that most loans which made an appearance in her data pertained to Māori culture, which was also identified as the main source of growth in Māori borrowings by Macalister (2006a) and by Davies and Maclagen (2006).

Looking at speaker effects on the basis of raw frequency counts, it was found that Māori use Māori loans more frequently than European New Zealanders, both in spoken and written New Zealand English (Kennedy and Yamazaki 1999), and in TV news reports (De Bres 2006); and that females use loans more frequently than males, this being particularly true of European New Zealanders gender comparisons (Kennedy and Yamazaki 1999; De Bres 2006). Unfortunately, these studies did not go beyond raw frequency counts or test the two factors (gender and ethnicity) at the same time, so we do not know whether controlling for one factor cancels the effect of the other one – a problem which we address in our model.

The only study which does consider Māori borrowings from an onomasiological concept-based approach by looking beyond raw frequencies is Onysko and Calude (2014). In that study, the use of three loanwords, namely the words *Māori* (native or indigenous), *Kiwi* (New Zealander) and *Pākehā* (European New Zealander) was found to be intimately linked to speaker ethnicity (though not to age or gender). The data used there constitutes a subset of the corpus analysed here and the current paper documents the larger model which comprises a larger set of loanwords. Thus the current paper expands on the pilot analysis reported on previously by expanding both the number of features investigated (to include linguistic characteristics of the loans) and the number of loans scrutinized.

# 4 Methods and data

Hence this study is borne out of the desire to investigate the success of Māori loans in New Zealand English by taking into account the wealth of knowledge gained from previous work, namely the fact that the use of loans might be sensitive to ethnicity, gender and age effects of the speakers involved. Similarly, we wanted to take into account linguistic factors which were found to be relevant to loanword success in studies of other language contact scenarios.

The data analysed here comes from the Wellington Corpus of Spoken New Zealand English (henceforth WSC), see Holmes et al. (1998) for a guide. The corpus contains one million words of various types of speech (spontaneous conversations, radio talkback, weather forecasts, judge summations, teacher monologues), with nearly half of it encompassing spontaneous conversation. One advantage of this corpus is that it provides full information of the participants who took part in the data collection (including their age, self-reported ethnicity, languages spoken, and profession). One limitation of the corpus is its age, being now almost twenty years old.

Given the size of the data, we used a combination of manual checking and Python programmes[3] to extract all the Māori loans used by each of the 843 participants contributing the one million words of speech. Overall, there were 129 Māori participants and 674 NZ European participants and 40 of other ethnicities. The participants of other ethnicities only used the Māori loans in two instances over the 42,256 words analysed. It is noteworthy that the speakers who primarily considered themselves to be of ethnicities other than White/European New Zealanders or Māori stayed virtually clear of all Māori loans. We excluded these speakers from the analysis and focused on the Māori and White/European New Zealanders speakers for the remainder of the analysis (which consisted of 950,718 words in total). The research team which put the WSC corpus together applied strict eligibility criteria for its participants. Speakers could only contribute spoken samples to the corpus if they fulfilled the following three requirements: (1) they had to have lived in New Zealand before the age of 10, (2) they had to have spent no more than 10 years overseas or less than half their lifetime (whichever was greater), and (3) speakers were not included if they had travelled outside New Zealand during the 12 months period prior to data collection (from the WSC guide, Holmes et al. 1998).

In our analysis, we excluded portions of code-switching, but allowed some loan phrases (7 in total, such as *kia ora*), which were deemed to form coherent and recurrent multi-word units. Distinguishing between code-switching and loans is a notoriously impossible task. Here, we take the view that non-recurring expressions which involve more than one lexical item count as code-switching (the participants in the corpus used either single loanwords or a small set of recurrent lexical bundles, such as *kia ora*, or else whole sentences, making the decision pretty straight forward to implement). Polysemous loans were manually disambiguated for context (*kia ora* meaning "hello" versus "*kia ora*" meaning "thank you" versus *kia ora* "goodbye"). We excluded all loans for which there

---

**3** We are greatly indebted to Paul D. James whose programming code was invaluable for this work.

was no obvious English counterpart that could come sufficiently close in matching its meaning (for instance, *mihimihi, whakamāori, pōwhiri*). We included English equivalents for our loans (and therefore, only loans with such equivalents) in order to control for the fact that certain loans are used more frequently simply because the meaning they encode arises more frequently in the data. As discussed in the introduction, not all topics or concepts come up in discourse equally frequently, so to make a raw frequency comparison is to ignore opportunity of use. For example, the Māori loanword "iwi" (roughly equivalent to the English word "tribe") occurs only 17 times in the WSC corpus, whereas the loanword "āe" (meaning "yes") occurs 49 times. However, the significance of these frequencies changes markedly when expressed as a ratio of the total potential use within the corpus: "āe" was used in 1% of instances in which it could have been used, while "iwi" was used in 59% of applicable instances. Hence, "iwi" is not only the more successful loanword of these two, but its use is also comparatively more prevalent than the English equivalent. Thus we included in our data all loans for which we could identify such a reasonable English counterpart (in some cases, we also included proper nouns, e.g., *Aotearoa* /New Zealand, *Tautoko* /Levin). All loans were identified manually and the contexts in which they occurred were similarly checked manually. In total, we identified 117 distinct loans (or loan phrases) which were used 1876 times in total throughout the one million words (but only 1810 of the loan uses could be attributed to participants whose information was available, the remaining 66 uses came from speakers who were recorded incidentally but whose permission and details were not recorded in the corpus guide).

Once the loan set was established and coded for its meaning in context, we matched it with an English equivalent. As discussed above, these "equivalents" are not understood as perfect synonyms, but only as semantic anchors to measure the actual use of a loan versus the opportunity of using it (realised opportunities versus total opportunity available to a speaker). It is not a straightforward matter to decide the best English equivalent of each Māori loan. We consulted dictionaries and native Māori speakers wherever possible to check our decisions (see Appendix A for the full list of loans and their matched English counterparts). This gave us a loan/equivalent pairing, e.g., *whare* /house, which could be coded for the characteristics of interest, that is, social characteristics of the speakers using the loans and linguistic characteristics of the loans. In some few cases, a given loan had a number of equally strongly associated English counterparts (e.g., *tamariki* and "children" or "kids") and both were included and treated as separate loan/equivalent pairings (*tamariki* /children and *tamariki* /kids). Our count measures include token (rather than type) counts mainly because we are interested to capture English counterparts for the Māori

loanwords as a weighting of potential versus realised use, and Māori loanwords occur in the singular form only in our data (with only one exception, namely the loanword "tamariki" meaning "children/kids").

Python code was used to extract for each participant, the total number of uses of each Māori loan and the total number of uses of each of the loan's English counterpart(s), and a figure of the total number of words uttered for each interaction (some speakers were involved in more than one interaction).

We controlled for differential use of words by participants and between words by including the participant and meaning as random effects within the model (Baayen 2008). We modelled the number of times a Māori meaning was used relative to its English counterpart as a Binomial random variable (McCullagh and Nelder 1989). Alternative treatments, such as taking the log of the loan word frequencies (plus one, to cope with zeros) and assuming a Normal distribution, or employing a rate-model where the frequencies are treated as Poisson-distributed and adjusted for the total number of words used (see for example, Poplack and Sankoff 1984, Van Hout and Muysken 1994; and Zenner et al. 2012), might have derived similar point estimates of the fixed effects, but statistical inference, such as confidence intervals or measures of significance for these effects, will be more appropriate under the Binomial model we use, particularly for extreme cases where the loan word is rarely or almost exclusively used.

Each use of a loan and each use of an English loan equivalent was attributed to a speaker. Below is a list of the various features and variables coded for each loan /equivalent pairing observed for each speaker recorded in the WSC corpus, starting with the sociolinguistic variables coded.[4]

Three sociolinguistic variables were coded for each speaker, as documented in the corpus guide.

**A. GENDER OF SPEAKERS** (categorical factor, binary feature). Information about the gender of the speakers included in our corpus has been recorded from the corpus manual (Holmes et al. 1998), which provides relevant data for each speaker from the questionnaires conducted during data collection (i.e., male or female).

**B. AGE OF SPEAKERS** (categorical factor, binary). During data collection, speakers were asked to tick a box giving the approximate age group in which they belonged at the time of the recording. These data give rise to a total of 16 age-groups (16–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, 75–79, 80–84, 85–89, 90 +). We did not want to include each age group separately in the analysis for two reasons: (1) we do not have

---

**4** Please refer to Table 1 and Table 2 for the number of observations per level for each predictor.

reason to believe that any one group might behave significantly differently to another, and (2) the number of participants is unevenly spread across the different age groups (there are many more speakers in the 20–24 and 25–29 age group than there are in say 90 + age group). However, there are important practical reasons to divide the age-group variable into two major categories as follows. As discussed in Section 3, the 1970s marked an important turning point in the history of the Māori language. So in our data, speakers around 46 years of age and older (32 % of our Māori participants and 34 % of our White/European NZ participants) would have been schooled before and up to the 1970 milestone. Conversely, speakers who were younger than 46 years old at the time of the data collection (68 % of our Māori participants and 66 % of our White/European NZ participants) would have been part of the new generation of speakers that were schooled and came-of-age in a climate where Te Reo Māori was on longer officially discriminated against. Given the historical facts of the New Zealand language contact situation, we wanted to compare these two age groups.

**C. ETHNICITY OF SPEAKERS (and ADDRESSEES)** (categorical factor, binary for speakers, and three-way distinction for addressees). Previous work has identified speaker ethnicity as being an important factor in loan-use. We hypothesize that Māori and White/European NZ speakers may have different motivations for using the various loans and might use these differently to each other. In order to test for these, we separated our data by speaker's ethnicity and built two separate statistical models, one for Māori participants and another for White/European NZ participants. The second factor pertaining to ethnicity was to do with the audience or addressees involved. Roughly half of our data consists in spontaneous, unplanned conversations which took place in the participants' own homes. In accordance with Audience Design theory (Bell 1984; Coupland 2007), we anticipate that speakers may construct and adapt their discourse (at least) in part, by tending to the kinds of hearers that are present at the time of the interaction. As regards loan-use, we hypothesise that speakers may be influenced by the ethnicity of their hearers (different groups may tend to this feature differently, that is, Māori and White/European NZ participants may behave differently towards their hearers' ethnicity, hence separate models were built here also). The hearer(s)' ethnicity was coded by means of a three-way distinction: Māori only addressee(s), White/European NZ only addressee(s), or a group of mixed Māori and White/European NZ addressees.

The second part of our coding involved the coding of seven linguistic features for each loan, as described below.

**1. COMPARATIVE OBSTRUENTS** (numerical variable). Obstruent sounds have been identified to be more difficult to pronounce than sonorant sounds (Goldberg et al. 2007; Miozzo 2003). We coded the obstruents feature by taking

into account the relative number of obstruents in the loans compared to the New Zealand English equivalents. Ignoring word boundaries, we coded the total number of obstruents in the loan and subtracted the total number of obstruents in the New Zealand English equivalent, to obtain an overall relative difference in ease of pronunciation. The values for this factor range from −3 (a loan with 3 more obstruents than its New Zealand English equivalent) to +6 (a loan with 6 fewer obstruents than its New Zealand English equivalent).[5]

**2. PHONOLOGICAL FIT INSIDE ENGLISH** (binary variable). We wanted to test whether loans whose sound patterns did not conform to English phonology or phonetics (for example, a word beginning with a velar nasal is allowed in Māori but not in English) will have a tougher time being imported into English.

**3. COMPARATIVE LENGTH (DIFFERENCE IN SYLLABLES)** (numerical variable). In accordance with previous studies looking at loanword success, we coded the difference in number of syllables between each loan and their closest English equivalent(s), ignoring word boundaries (both in Māori loans and their English equivalents). Each loan thus received a count between −4 (signalling that the English equivalent was 4 syllables shorter than the Māori loans) and +5 (for loans which were five syllables shorter than their English counterparts).

**4. POLYSEMOUS USE** (binary variable). While Māori words are generally highly polysemous, we hypothesised that loans imported into English with more than one meaning might become more successful in their host language due to their multiple uses. Each loan was coded for whether it was found to be used monosemously or polysemously in the WSC corpus. For example, the word *pāua* is listed in the online Māori dictionary (http://maoridictionary.co.nz/, accessed 20 April 2017) as having three meanings: (1) abalone, (2) spinner/fishing lure and (3) hoof. However, in our corpus the loanword *pāua* was only used in the first meaning so it is coded as non-polysemous.

**5. LEXICALIZATION** (binary variable). Some loanwords imported into New Zealand English do not have lexicalized equivalents in the host language, and require multiple words (i.e., an entire phrase) to express the equivalent concept intended, such as, *Aotearoa* for *New Zealand* (two words, non-lexicalized). So in the case of "non-lexicalized" counterparts, an actual phrase was the only means to express a similar concept in the host language. Compounds figure as non-lexicalised items for us (though the decision is, of course, arbitrary given that the cline between *phrase → compound → word* is a continuum rather than a strict

---

5 We would like to thank one of the anonymous referees for suggesting the relative obstruent difference (previously we only considered the total number of obstruent sounds in the individual loans, and relative difference in length, but not relative difference in number of obstruent sounds).

delineation). While we did not want to exclude these types of loans on the grounds of not having equivalent counterparts, we also wanted to keep track of how they performed in the model, in case lexicalization is relevant to relative success. Certain loans themselves encompass multiple words, such as *kia ora*, and these were automatically classed as lexicalized. Note that whether or not the loans themselves consisted of a multi-word unit was not relevant to this category, as this was already taken into account by the comparative length factor.[6]

**6. CATEGORY**[7] (binary variable). We wanted to split our loans into two major syntactic categories, namely content and function words. It is well documented that function words are less readily borrowed compared to content words, and although we are not specifically interested to test this here, we wanted to include it in the model due to its salience in the borrowing literature (cf. Tadmor 2009: 59).

**7. CULTURAL/CORE**[8] (binary variable). Following Myers-Scotton (2002: 41), we coded the loans as either pertaining to "cultural" or "core" aspects of vocabulary, for example, *kaimatua* ('elder') and *iwi* ('tribe') were coded as "cultural", whereas *whare* ('house') and *wahine* ('woman') were coded as "core". We follow Myers-Scotton in favour of the more traditional distinction between 'luxury' and 'needed' loanwords because we agree with Haspelmath and Tadmor (2009: 46–49) and Onysko and Winter-Foremel (2011: 1551–1553) that these latter terms appear to be tainted by judgments regarding the borrowing process, and also that strictly speaking, loanwords are never fully required (languages can always draw on other internal resources to express a given concept), nor purely superfluous (even when close counterpart equivalents exist, the loan will still serve a particular communicative function which led to its being imported into the host language in the first place). The deciding principle used in categorizing the loanwords as "cultural" or "core" rested on whether or not the loanword in question was used to identify an object, event, or custom that is distinctively associated with the Māori tradition and perspective in some way. For example, *kaumatua* 'elder', *iwi* 'tribe', and *karakia* 'prayer' were all coded as "cultural"

---

**6** Contrary to initial appearances, comparative length and lexicalization measure different and unrelated things. For example, a given loan might have a lexicalized counterpart but at the same time, still be shorter than its English equivalent, in number of syllables (such as, *korua* and *old man*), or it may have a non-lexicalized counterpart and still be longer than its English equivalent (such as, *takahanga* and *tramping trip*).

**7** One referee suggested coding the place names words as a separate category from the non-place names, with the view that they "behaved" rather differently. While we are in theoretical agreement with the referee about this, the model did not find any statistical significance between the two types (potentially due to the restricted number of place names loans included).

**8** We thank one of the anonymous referees for pointing us into this direction.

because, they carry specific and distinctive meanings within a Māori perspective, whereas *kao* 'no', *mahi* 'work' and *maunga* 'mountain' do not carry such special association with the Māori world-view.

One last feature we considered including here was the frequency of use of the loans in Māori. Findings reported in Van Hout and Muysken (1994) suggest that the frequency with which a loan might be used in the donor language may have a bearing on how well it does in the host language. While we tried to include this feature in our model initially, the only Māori corpus available to date is the Broadcasting Corpus collected put together by Mary Boyce (Boyce 2006), which seemed restricted (one million words of broadcasting language). Many of our loans were not found in that corpus and we felt that the data obtained from it were not truly representative of the use of these loans in Māori as a whole. For this reason, we abandoned coding this factor.

# 5 Results

There were 129 unique Māori speakers in the data, who contributed speech to a combined total of 179 records (most speakers were recorded once only, but some occur in the data multiple times, including one speaker who spoke in six separate recordings). Speakers used between 1 and 34 of the identified loan/equivalent pairs in each record. There were 675 unique White/European NZ speakers who contributed speech towards 972 records. These speakers used between 1 and 23 of the identified loan/equivalent pairs in each record.

## 5.1 Modelling loan-use – general remarks

Because repeated measurements were made on some speakers, and loan/ equivalent pairs were used by more than one speaker, intercepts for the speakers and loan/equivalent pairs were treated as random effects. This assumes that the speakers and meanings observed are random draws from their respective populations, and we are only interested in modelling the variability of those populations, rather than estimating the probabilities for each of the speakers and each of the meanings we have observed in the data. Our interest lay in modelling how sociolinguistic characteristics of the speakers and linguistic characteristics of the loans might affect the probability of using a given Māori loanword. Our model treats these as fixed effects.

We model the data as being drawn from Binomial distributions, where the number of trials is taken to be the total number of times the Māori loanword or its English counterpart is expressed (i.e., the number of times the meaning is evoked), and the number of events is taken to be the number of times the loanword is used in those cases. We employ the logistic link function to estimate how each of the fixed effects affect the odds of using a Māori loanword.

We started with models involving all fixed effect variables, and performed variable selection to produce simpler models. There is ongoing debate about the correct approach to judge the statistical significance of effects in GLMMs (Bolker et al. 2008). We eliminated variables that had low *F*-statistics to produce sufficiently parsimonious models. This resulted in models where the fixed effects were generally significant at, or close to the 5 % level, under a debatable assumption concerning the appropriate degrees of freedom. In addition, a likelihood ratio test of the original model including all the fixed effect variables and the reduced model involving a subset of these variables was undertaken. In all cases, the likelihood ratio test was not significant at the 5 % level of significance, suggesting the reduced model was not significantly inferior to the more complicated model. In addition, the model comparison criteria AIC and BIC also favoured the reduced models.

Estimated effect sizes and associated confidence intervals were derived by simulating 1000 sets of data from the reduced models, including the random effects, using the parametric bootstrap procedure employed by the bootMer function in the R boot package. The model was fit to each of these data sets, and we take the median estimated effect value for each parameter as the point estimate of the effect size, and the 2.5$^{th}$ and 97.5$^{th}$ percentiles to be the lower and upper bounds respectively of the 95 % confidence interval for the parameter value.

For categorical variables, the coefficients are reported as the estimated log-odds of using a Māori loanword rather than the English equivalent for the stated category relative to the odds for the unstated category. Thus a negative coefficient implies the odds of using a loanword are lower for the stated category relative to the baseline category, while a positive coefficient implies the odds of using a loanword are greater for the stated category relative to the baseline category. The larger the absolute value of the estimated coefficients, the greater the relative difference in odds. Note that on the log scale, equal odds of using a loanword by the dichotomous categories of a factor is represented as a coefficient value of 0. For a continuous variable, the coefficient represents the estimated average change in relative odds of using the Māori loanword as the variable increases by one unit. This change is additive in the log-scale. Thus a negative coefficient represents the odds of using a loanword decreasing as the

variable increases, while a positive coefficient represents the odds of using a loanword increasing as the variable increases.

## 5.2 Results 1. Māori speakers

Counting every instance of a loan use for Māori speakers (using the list of loans in Appendix A), we found 1,393 loan tokens (and 13,128 equivalent non-loan tokens). From the 115 meanings which came up in the corpus data for this speaker group, 97 loan types were used. In other words, even for Māori speakers, some (few) of the meanings used were never expressed by means of a loanword. Some Māori speakers used both the loanword and New Zealand English counterpart for the same meaning (this happened for 32 of the loan types, in other words, for about a quarter of the meanings investigated). Finally, we calculated the average of each loanword for Māori speakers (per 100,000 words of running speech), and these are given in Appendix B (see the first column).

Table 1 summarises our results for the Māori speakers (please refer to Appendix C.1 for estimated errors).

The odds-ratio measure given in the table above provides information of the proportional influence of each predictor that was found to be significant (as given by the confidence interval associated with it, confirming the

**Table 1:** Model results for Māori speakers.

| Effect | Median estimated odds-ratio from 1000 bootstrap replicates | 95 % confidence interval for the estimated odds-ratio from 1000 bootstrap replicates |
|---|---|---|
| Gender – Male (66 female speakers, 65 male speakers) | 0.479 | (0.275, 0.828) |
| Lexicalisation – Non-lexicalised (89 lexicalised loans, 26 non-lexicalised loans) | 6.835 | (1.479, 40.247) |
| Category – Function Word (106 function loans, 9 content loans) | 0.067 | (0.006, 0.536) |
| Cultural or Core – Core (50 cultural loans, 65 core loans) | 0.074 | (0.019, 0.229) |
| Excess number of syllables in the English word (right-skewed distribution centred on −1) | 2.086 | (1.402, 3.208) |

Null deviance: 1744.1 on 2200 degrees of freedom. Residual deviance: 1678.0 on 2195 degrees of freedom.

exclusion of 1 in that interval), so that an odds ratio of 1 would correlate with the same probability of using a loanword for each category. For example, the gender category has an odds-ratio of 0.479 and the category specified in the table as the baseline is 'male', so that means, in our model, the odds of males using a loanword is 0.479 times lower than females (using a loanword). Table 1 shows that Māori speakers were more likely to use a loanword when the New Zealand English equivalent is a phrase or compound (i.e., when it is non-lexicalised) and that loanwords whose English equivalents have comparatively more syllables were favoured. However, they were less likely to use loanwords that are function words compared to content words, and less likely to use core words compared to cultural words. Among the Māori speakers recorded in the corpus, female speakers were more likely to use loanwords compared to male speakers. It is relevant to note here that the variable of lexicalization has a large confidence interval (1.479, 40.247), which probably stems out of the reduced number of non-lexicalised loanwords (89 lexicalized and 26 non-lexicalised meanings).

## 5.3 Results 2. White/European NZ speakers

White/European NZ speakers used 417 loan tokens in total (50,029 equivalent non-loan tokens). They used 47 loan types of the 113 meanings which came up in the corpus data for these speakers (in other words, for just over half of the meanings investigated, White/European NZ speakers – as a group – did not use the available loan). Finally, some speakers made use of both the NZ English equivalent and the Māori loanword (i.e., we find variability within individual speakers as regards their lexical choices for the meanings used). There were 21 meanings for which this happened (roughly a fifth of the total meanings that arose in the corpus). As before, we calculated the average of each loanword for White/European NZ speakers (per 100,000 words of running speech), and these are given in Appendix B (see the second column).

Table 2 summarises our results for the White/European NZ speakers (please refer to Appendix C.2 for estimated errors).

As with Māori speakers, White/European NZ speakers were more likely to use a loanword if it replaced a phrase rather than a single word as its English equivalent, although precisely how much more likely is difficult to say from this data (see the large confidence interval range, 88 lexicalized loans and 25 non-lexicalised loans). Also similarly to Māori speakers, the fewer syllables a loanword had compared to its English equivalent, the more likely the loanword was to be used. Again, content words were preferred to function words, and

**Table 2:** Model results for White/European NZ speakers.

| Effect | Median estimated odds-ratio from 1000 bootstrap replicates | 95 % confidence interval for the estimated odds-ratio from 1000 bootstrap replicates |
|---|---|---|
| Polysemous use – Yes (8 'yes' loans, 105 'no' loans) | 0.003 | (0.001, 0.391) |
| Lexicalisation – non-lexicalised (88 lexicalised, 25 non-lexicalised) | 179.727 | (5.837, 29,050.878) |
| Category – function word (9 function loans, 104 content loans) | 0.003 | (0.000, 0.279) |
| Excess number of syllables in the English word (right-skewed distribution, centred on −1) | 5.483 | (2.064, 14.367) |
| Cultural or core – core (48 cultural loans, 65 core loans) | 0.013 | (0.001, 0.170) |

Null deviance: 687.0 on 7715 degrees of freedom. Residual deviance: 636.4 on 7710 degrees of freedom.

cultural words were preferred to core words. However, unlike Māori speakers, there were no significant gender differences for White/European New Zealanders detected, and loans with polysemous meanings were significantly less favoured than monosemous loans (in other words, White/European New Zealanders favoured loans with single meanings).

## 5.4 Results 3. Ethnicity of the addressee(s) in conversation data

Because the WSC corpus contains conversational data among the spoken interactions recorded, in which the ethnicity of the addressee(s) is known, we wanted to test any possible effects of the ethnicity of the interlocutors addressed. As this information was only available for the conversational data, we only used that subset of the corpus for this analysis. This subset includes 500 interactions (roughly half of the original data). As a starting point, we considered the model with the variables identified as being significant from the full data set for each of the Māori and White/European New Zealander groups, respectively, and added the addressee ethnicity variable. We then performed variable selection on the new variables, and confirmed that the reduced model after variable selection was both insignificantly different from the model with all new

interactions, and significantly better than the model without any mention of the ethnicity of the audience. The results are given in Table 3 for Māori speakers and graphically depicted in Figure 1.

**Table 3:** Model results for conversational data for Māori (variables involved in significant interactions only).

| Effect | Median estimated odds-ratio from 1000 bootstrap replicates | 95 % confidence interval for the estimated odds-ratio from 1000 bootstrap replicates |
|---|---|---|
| | | Ethnicity of Audience |
| – Mixed | 0.088 | (0.019, 0.328) |
| – White/European NZ only | 0.402 | (0.063, 3.181) NS |
| Excess number of syllables in the English word (when addressing Maori audiences) | 4.394 | (2.330, 9.502) |
| Interaction between Ethnicity of Audience and Excess number of syllables in the English word | | |
| – Mixed | 0.400 | (0.260, 0.647) |
| – White/European NZ only | 1.402 | (0.520, 7.791) NS |
| Cultural or Core – Core (when addressing Maori audiences) | 0.011 | (0.001,0.078) |
| Interaction between Ethnicity of Audience and whether a word is Cultural or Core | | |
| – Mixed | 8.16 | (2.691, 31.826) |
| – White/European NZ only | 0.027 | (0.002, 0.186) |

Residual deviance with no effect of Ethnicity of Audience: 799.9 on 1045 degrees of freedom. Residual deviance allowing for an effect of Ethnicity of Audience: 761.6 on 1039 degrees of freedom.

For the Māori speaker group, a significant interaction was detected between the ethnicity of the audience and the excess number of syllables in the English counterpart word, suggesting that the effect of the brevity of the Māori word on the probability the word was preferred depends on the ethnicity of the audience being addressed. As indicated by the figures in Table 3, as the relative length of the English word in syllables increases, the probability of using the Māori loanword rather than its English counterpart increases regardless of the ethnicity of the audience, but it increases significantly faster if the audience is exclusively Māori or White/European NZ compared to if the audience is a mix of both ethnicities. This effect, taken with differences in the inherent probability of using the Māori loanword depending on audience ethnicity, is shown in Figure 1.
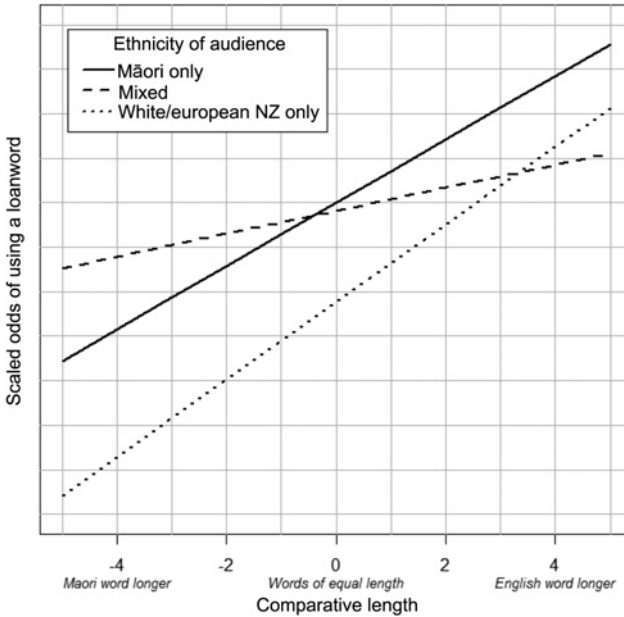
**Figure 1:** Effect of the difference in syllables between word pairs on the log-odds of using a Māori loanword by Māori speakers, depending on the ethnicity of the audience (x-axis: comparative length is calculated as number of syllables in the comparative English word minus number of syllables in the Māori loanword, y-axis: logged-odds of using a loanword over an English equivalent).

For loans which are 5 syllables longer than their English counterparts (i.e., a comparative length of –5), the Māori loanword is more likely to be used when speaking to mixed audiences compared to Māori only audiences, and more likely to be used when speaking to Māori audiences compared to White/European NZ audiences. However, for English words which are 5 syllables longer than their Māori equivalents (i.e., a comparative length of 5), the Māori loanword is most likely to be used when addressing a Māori audience. For word pairs that are equal in length, the Māori word is more likely to be used when addressing a Māori or mixed ethnicity audience than a White/European NZ audience.

For Māori speakers, there was also an interaction between the ethnicity of the addressee(s) and whether the loanword was a cultural or core word. While cultural words were consistently much more likely to be used than core words, regardless of the audience, the difference was greatest when addressing a White/European NZ audience, less so when addressing a Māori audience, and less again when addressing an audience of a mixture of ethnicities.

Once we trimmed the data to only consider conversation (where we had information about the ethnicity of the audience member(s)), we only had 89 loan tokens. This resulted in a huge loss of power, making it very difficult for the model to converge and to find a sensible answer regarding the dependent variables investigated.

Looking at the remaining variables identified as significantly affecting the probability of using a Māori loanword among White/European New Zealanders, and including the ethnicity of the addressee and its interactions, there were no significant interactions between the ethnicity of the audience the speaker was addressing, and any of the variables previously identified. There was a hint of a suggestion that the ethnicity of the audience being addressed might influence the odds of a White/European New Zealander using a Māori loan word, but the uncertainty in the data could not allow us to determine this effect as significant.

We include below a summary of the main findings from the statistical models discussed in this section.

Summary of Main Findings:

1.  Both Māori and White/European NZ speakers are less likely to use a Māori loanword if that word is lexicalized in English (in other words, if English has an actual word for it), or if the word is a function word rather than a content word.

2.  Although both Māori and White/European NZ speakers are more likely to use a Māori loanword the longer the English counterpart word is in terms of number of syllables in relation to the loanword, there is a suggestion that Māori speakers modify this behaviour depending on their audience. Although the direction of the effect remains the same (and is significant) regardless of audience, it makes less of a difference when addressing a mixed audience than when addressing an audience made up of exclusively Māori participants.

3.  Both Māori and White/European NZ speakers are less likely to use a Māori loan if that loan expresses a core rather than cultural meaning.

4.  For Māori speakers only, gender has an effect, where males are on average less likely to use a Māori loanword than females.

5.  For White/European New Zealanders only, the probability of using a Māori loanword decreases if that word is polysemous.

6.  For Māori speakers, the probability of using a loanword can be different depending on the ethnicity of the audience, but this difference will be moderated by the comparative length of the word being used, as noted previously.

# 6 Discussion

The summary given above brings a number of insights to the findings identified in the literature concerning the New Zealand English contact situation, which we feel also make a wider contribution to the study of loanwords in general. One of the most important of these is the finding that speakers adjust their loanword use according to the audience they are addressing. Specifically, in our case, the ethnicity of the audience has a bearing on the number of loanwords used, with a higher number of loanwords in situations where the audience is exclusively Māori, or exclusively Pākehā (White/European NZ), with a mixed audience drawing the smallest number of loanwords. The effect is observed for Māori speakers, but not for Pākehā speakers. This could be because the ethnicity of the audience is relevant for Māori but not for Pākehā, or it could be that it may be relevant for both ethnic groups, but the lower rates of loanwords by Pākehā speakers lead to a loss of predictive power in the model. It would be interesting to look at this factor in a larger data sample to check whether the patterns stay the same.

At any rate, our data indicates that for Māori, the use of loanwords is – at least in part (note the interaction effects) – motivated by the expression of solidarity with the Māori perspective and the desire of aligning their identity within a Māori background, exemplifying Bell's (1984) and Coupland's (2007) notions of Audience Design. While it is not surprising that Māori might use more loanwords with a Māori only audience, it does come as a surprise that they use more loanwords with a Pākehā only group compared to a mixed ethnicity group. One explanation might be that the use of loanwords involves both *convergence* to the audience (using more loans with a Māori group, in other words, a group which shares the speaker's own ethnic affiliation), as well as an element of *divergence* from the audience (using more loans with an exclusively non-Māori group, that is, a group which is ethnically distinct from the speaker's ethnic group).

The model also shows that the effect of the ethnicity of the audience for Māori speakers is mediated by linguistic properties of the loanwords, namely, it is mediated by comparative length. The shorter the loanword compared to its English equivalent, the more likely it is to be used by Māori speakers when addressing a Māori only audience or a Pākehā only audience. The desire for economy of expression seems to be a significant predictor of loanword success for both Māori and Pākehā speakers in general (regardless of ethnicity of the audience), but its effect is further amplified in the context of Māori speakers addressing a Māori only audience.

A second insight from our study is that for the New Zealand English context, *both* social factors (such as the ethnicity of the audience members, or the gender

of speakers), and certain linguistic properties of the loans themselves have a statistically significant bearing on loanword success. As shown by others in different language contact situations, loanword success is not just a social matter; it is also a linguistic matter. But our work goes further to show that in fact, controlling for both types of factors in the same model is important because they each bring different elements into play as regards loanword use, highlighting not just how different parts of a given community use loanwords, but also their potentially distinct motivations. For instance, as noted above, while economy of expression is an important facilitator of loanword use for both Māori and Pākehā speakers, loans which have this linguistic characteristic are significantly more successful when used in situations involving a Māori only audience.

Another example relates to whether or not a loanword is polysemous: for Pākehā, but not Māori speakers, polysemy in a loan decreases its relative success, in other words, Pākehā speakers prefer monosemous loanwords. This finding is itself interesting given that Chesley and Baayen (2010) found exactly the opposite to be true of loanwords being borrowed into French. This suggests that while a word's semantic versatility might be seen as an attractive characteristic in one language contact scenario, it might be seen as a negative characteristic in another. It could be that for Pākehā speakers, multiple meanings lead to uncertainty about a given loanword's actual meaning and they fear using it incorrectly, which ultimately ends in avoidance of the loanword altogether. This explanation rings true given that polysemy does not function as such an inhibitor for Māori speakers, who are more likely to be confident in their use of the (loan)words (in our corpus data, it so happens that all the speakers who reported having Māori ethnicity, also reported having some fluency in Māori – this is by no means the norm for the current New Zealand society).

Testing the factors which drive loanword success together also allows us to note that the speaker's gender has different implications for the use of loanwords depending on the ethnic group analysed. In accordance with previous literature (Macalister's work in general but also Kennedy and Yamazaki 1999), Māori use loanwords more than Pākehā, and females use more loanwords than males (Kennedy and Yamazaki 1999). However, gender effects were only observed for the Māori group, and not in the Pākehā group. In light of the positive attitude towards Te Reo Māori in New Zealand society and towards the borrowing of Māori loans into New Zealand English, if Macalister (2006a) and Davies and Maclagan (2006) are indeed correct in their predictions of ongoing changes in New Zealand English to accommodate an increase in Māori loanwords, then it would be precisely females that we would expect to be leading the way in this innovation (see discussion in Meyerhoff 2006: 220) – in our case it is Māori females driving it. In his 2008 study, Macalister suggests that the gap

between Māori and non-Māori, and male and female with regards to size of vocabulary of Māori words is decreasing. This could be interpreted as suggesting that more people are becoming more familiar with and more confident in using Māori words. While the corpus data analysed here still identifies such a gap between genders for Māori speakers, and between ethnicities, a more recent body of data is required to investigate exactly how the overall increase in loan-use discussed by Macalister and others is distributed across the New Zealand population today.

Our results appear directly orthogonal to a recent study of English loan-words into Dutch. Also using multivariate modelling, Zenner, Speelman and Geeraerts found the exact opposite in their data: females used comparatively fewer loanwords than males (2015: 341). However, the two studies differ in (at least) one crucial aspect: loanword meanings. The Dutch data contains a high amount of expletives ("*shit*" and "*fuck*" were the most frequent loanwords, 2015: 337). Comparing the two studies highlights an important aspect of loanword use: while social factors are most certainly relevant in the use of loanwords (and they can indeed vary with each language contact situation), one cannot overlook the types of loanwords being used and their functions in discourse.

As an exercise in model building, we merged the data from both ethnic groups to see what would happen. The results from this combined dataset were pulled in the direction of the Māori group's results and the effects found for the Pākehā group disappear (see Table 4 below). This seems contrary to our initial expectations given the small number of Māori participants (129[9]) compared to the Pākehā group (674), but in reality, it shows the power and importance of the statistical model. Although we have more observations from the Pākehā group, these speakers do not use nearly as many loans as the Māori group does, which is why their results become drowned out by the behaviour of the Māori group when combined. However, at the same time, the table shows some of the pitfalls of confounding factors in the model: certain patterns are obscured (such as the effects of polysemy for Pākehā speakers) by the merging the two sets.

While social factors relevant to loanwords use are likely to vary and behave differently with each language contact context, it is useful to consider the linguistic properties which hold their weight above and beyond the social factors investigated and to compare these with those noted in other language contact situations.

Linguistically-speaking, three major drivers of success were identified in the New Zealand data, namely loanword meaning (core versus cultural), economy of expression (measured by lexicalization and comparative loanword length), and

---

**9** Two Māori speakers did not use loanwords.

**Table 4:** Model combining all the data (both the Māori and White/European NZ speakers).

| Effect | Median estimated odds-ratio from 1000 bootstrap replicates | 95 % confidence interval for the estimated odds-ratio from 1000 bootstrap replicates |
|---|---|---|
| Lexicalisation – non-lexicalised | 19.242 | (3.576, 84.513) |
| Category – function word | 0.050 | (0.005, 0.355) |
| Cultural or core – core | 0.053 | (0.015, 0.183) |
| Excess number of syllables in the English word | 2.066 | (1.398, 3.056) |
| Number of words in the loan | 0.221 | (0.042, 0.936) |

word category (content words were preferred to function words). We found that loanwords which encode cultural meanings (rather than core meanings) were propelled to greater relative success. This finding corroborates some of the explanations proposed by Macalister particularly in his 2007a article regarding the desire for using words with a "high degree of cultural specificity" whose cultural reference can nevertheless be extended beyond the Māori context per se (p. 503). It is interesting to note that in our data, while all Māori speakers reported some fluency in Māori, none of the Pākehā speakers reported any familiarity with the Māori language. Despite this discrepancy in bilingual abilities, both groups seemed to be drawn to the loanwords which encoded culturally specific terms. We do not know whether the core/cultural distinction applies more widely to (all?) other language contact scenarios, but certainly, in our case study of New Zealand English, this appears to be highly relevant. One problem with coding the core/cultural distinction is that it remains a subjective measure and we feel that more objective criteria could help improve its reliability. One way forward is to look for entrenchment measures, as proposed by Zenner et al. (2014) in their study of English loanwords into Dutch, but this remains beyond the scope of the current paper.

Economy of expression was a second predictor of loanword success for both ethnic groups so that words which were comparatively shorter than their host language equivalents achieved greater success, as did non-lexicalized loans compared to lexicalised ones. The strive towards economy of expression has been observed in several linguistic levels (most famously noted by Zipf 1935), and specifically proposed by Macalister for Māori loanwords in New Zealand English (2007a) as well as noted in other language contact situations (Shin 2010;

Chesley and Baayen 2010; Zenner et al. 2012; and Winter-Froemel et al. 2012). The push for economy in the use of loanwords in many different language contact scenarios suggests that future studies of loanword success should control for this factor.

Finally, we also found that function words were less favoured than content words (by all groups of speakers). This is by no means newsworthy, see early work by Thompason and Kaufman (1988/1991) or discussion in Tadmor (2009), but we nevertheless found it surprising that this factor remained statistically significant in the model even after controlling for the core/cultural opposition in meaning (invariably, function words are always core meanings and never cultural meanings).

While we hope our study makes some meaningful contributions to the general body of work on the use of loanwords, it is of course not without its pitfalls. Statistical models are only imperfect approximations of any situation investigated and our own model could be improved by adding further factors which may be relevant to loan-use, such as frequency of use of the Māori words in a Māori corpus (which might be used to calculate entrenchment measures of the sort suggested by Zenner et al. 2014), and psycholinguistic cues which may prime speakers to use a loan that has been previously used by another speaker. Given the importance of the audience and addresses' ethnicity in our data, measuring priming effects (that is, who uses a loanword first) would be beneficial in future work. Other variables in the addressees' social profile could also be included here, such as age or gender (our study only investigates the effect of addressee ethnicity). Also, given the limited number of loanwords used by Pākehā speakers, a larger body of data, comprising more loanwords, would be fruitful in clarifying the extent and role of various loanword success predictors for that group.

The corpus data analysed could also be complimented by experimental data which could test whether speakers do indeed "capitalize" on the opportunity of using certain loanwords in a given context (specific culturally relevant contexts might be set up to elicit the use of loanwords, and the effect of different types of interviewers could be probed). Controlling the topic of conversation would have its own drawbacks of course, because the data would no longer be completely spontaneous and unplanned.

A different future direction would be to look towards the language of the media and compare the use of loanwords in various media outlets (such as newspapers or radio shows) with the spontaneous data analysed here. Measures of entrenchment of the loanwords in the media (for example similar to those in Zenner et al. 2013) would provide another potential predictor in the relative success of loanwords in conversational data.

Summary of implications for future studies of loanword use:

1.  Loanword success is most fruitfully captured by a complex combination of social and linguistic factors (and perhaps even interactions between these) – both sets of factors could affect the use of loanwords in different ways so ideally, they should both be investigated together, in the same data set.

2.  Audience effects involving both convergence and divergence are relevant to loanword use and whenever possible, these should be taken into account. This requires that the data analysed come from conversations, which also provides added advantages, such as the potential for investigating more natural, less careful, less planned speech. Of course, it does come with the obvious disadvantage in time and effort costs involved in recording and transcribing it.

3.  Word category (function versus content) and economy of expression (whether measured in relative length or lexicalization) appear to be important factors in loanword success more universally and should be taken into account in any future studies of loanword use.

# 7 Conclusion

Our study proposes a balanced and detailed model for investigating loanword use with the aim of capturing relative loanword success by incorporating a measure of *actual* loan-use measure against *potential* loan-use. In line with previous studies, we show that both sociolinguistic and linguistic factors are relevant to investigations of loan-use, but our study highlights the importance of investigating both sets of factors within the same data set. Our case study of indigenous Māori words borrowed into New Zealand English demonstrates the benefits of controlling for multiple factors which influence loan-use at the same time and shows that loanword use can be linked to Audience Design theories. We hope that our model will serve as a starting point for future loanword studies and that it can be further developed and improved in order to help shed light on the complex factors that are involved in the process of adopting words from language into another.

# References

Baayen, Harald. 2008. *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bell, Allan. 1984. Language style as audience design. *Language in Society* 13. 145–204.

Benton, Richard. 1991. *The Māori language: Dying or reviving?* Honolulu: East West right. (Reprinted by New Zealand Council for Educational Research in 1997).

Bolker, B.M., M.E. Brooks, C.J. Clark, S.W. Geange, J.R. Poulsen, M.H.H. Stevens & J-S.S. White. 2008. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends in Ecology and Evolution* 24. 127–135.

Boyce, Mary. 2006. *Māori Broadcasting Corpus: A corpus of modern spoken Māori*. Unpublished PhD thesis available in the library at Victoria University of Wellington.

Chesley, P. & H. Baayen. 2010. Predicting new words from newer words: Lexical borrowings in French. *Linguistics* 48. 1343–1374.

Clyne, M. 2003. *Dynamics of language contact*. Cambridge: Cambridge University Press.

Coupland, Nicholas. 2007. *Style language variation and identity*. Cambridge: Cambridge University Press.

Daly, Nicola. 2007. Kūkupa, Koro and Kai: The use of Māori vocabulary items in New Zealand English children's picture books. *New Zealand English Journal* 21. 20–33.

Davies, C. & M. Maclagan. 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.

De Bres, Julia. 2006. Māori lexical items in the mainstream television news in New Zealand. *New Zealand English Journal* 20. 17–34.

Degani, M. & A. Onysko. 2010. Hybrid compounding in New Zealand English. *World Englishes* 29. 209–233.

Degani, Marta. 2010. The Pakeha myth of one New Zealand /Aotearoa: An exploration in the use of Maori loanwords in New Zealand English. In R. Facchinetti, David Crystal & Barbara Seidlhofer (eds.), *From international to local English – and back again*, 165–196. Frankfurt am Main: Peter Lang.

Deverson, Tony. 1991. New Zealand English lexis: The Māori dimension. *English Today* 28. 18–25.

Durkin, Philip. 2014. *Borrowed words: A history of loanwords in English*. Oxford: Oxford University Press.

Furiassi, Cristiano. 2011. False Italianisms in English dictionaries and corpora. In A. Koll-Stobbe & S. Knospe (eds.), *Language contact around the globe*, 42–72. Oxford: Peter Lang.

Goldberg, A.M., J. Choline, J.W. Bertz, B. Rapp & M. Miozzo. 2007. Evidence for morpho-phonological processes in spoken production. *Brain and Language* 103. 162–163.

Graedler, Anne-Line & Gudrun. Kvaran. 2010. Foreign influence on the written language in the Nordic language communities. *International Journal of the Sociology of Language* 204. 31–42.

Haspelmath, Martin & Uri Tadmor. 2009. *Loanwords in the world's languages. A comparative handbook*. Berlin: Mouton de Gruyter.

Holmes, J., B. Vine & B.G. Johnson. 1998. *Guide to the wellington corpus of spoken New Zealand English*. Wellington, New Zealand, School of Linguistics and Applied Language Studies: Victoria University of Wellington.

Humbley, John. 2008. How to determine the success of French language policy on Anglicisms – some methodological considerations. In R. Fischer & H. Pulaczewska (eds.), *Anglicisms in Europe. Linguistic diversity in a global context*, 85–105. Cambridge: Cambridge University Press.

Imm, Tan Siew. 2009. Lexical Borrowing from Chinese Languages in Malaysian English. *World Englishes* 28(4). 451–484.

Johanson, L. 1993. Code-copying in immigrant Turkish. In G. Extra & L. Verhoeven (eds.), *Immigrant languages in Europe*, 197–221. Bristol US, Adelaide Australia: Multilingual Matters.

Kennedy, G. & S. Yamazaki. 1999. The Influence of Māori on the New Zealand English lexicon. In J. Kirk (ed.), *Corpora galore: Analyses and techniques in describing English*, 33–44. Amsterdam: Rodopi.

Kouega, Jean Paul. 2009. Campus English: Lexical variations in Cameroon. *International Journal of the Sociology of Language* 199. 89–101.

Macalister, John. 2001. Introducing a New Zealand newspaper corpus. *New Zealand English Journal* 15. 35–41.

Macalister, John. 2006a. The Maori presence in the New Zealand English lexicon, 1850 –2000: Evidence from a corpus-based study. *English World-Wide* 27. 1–24.

Macalister, John. 2006b. The Maori lexical presence in New Zealand English: Constructing a corpus for diachronic change. *Corpora* 1. 85–98.

Macalister, John. 2007a. Weka or woodhen? Nativization through lexical choice in New Zealand English. *World Englishes* 26. 492–506.

Macalister, John. 2007b. Revisiting Weka and Waiata: Familiarity with Maori words among older speakers of New Zealand English. *New Zealand English Journal* 21. 34–43.

Macalister, John. 2008. Tracking changes in familiarity with borrowings from Te Reo Māori. *Te Reo* 51. 75–97.

Macalister, John. 2009. Investigating the changing use of Te Reo. *NZ Words* 13. 3–4.

Māori Language Commission. 2015. Te Taura Whiri I Te Reo Māori. http://www.tetaurawhiri.govt.nz/ (accessed 17 September 2015).

McCullagh, P. & J.A. Nelder. 1989. *Generalized linear models* 37. New york: Springer US/ Chapman & Hall (CRC press).

Meyerhoff, Miriam. 2006. *Introducing sociolinguistics*. London: Routledge.

Milroy, J. 1997. Internal versus external motivations for linguistic change. *Multilingua* 16. 311–323.

Miozzo, M. 2003. On the processing of regular and irregular forms of verbs and nouns: Evidence from neuropsychology. *Cognition* 87. 101–127.

Myers-Scotton, C. 2002. *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

Onysko, Alexander & Andreea Calude. 2014. Comparing the usage of Māori loans in spoken and written New Zealand English: A case study of Māori, Pākeha, and Kiwi. In Eline Zenner & Gitte Kristiansen (eds.), *New perspectives on lexical borrowing*, 45–72. Berlin, New York: De Gruyter.

Onysko, Alexander & Winter-Froemel. Esme. 2011. Necessary loans–luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43(6). 1550–1567.

Poplack, S., D. Sankoff & C. Miller. 1988. The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics* 26. 47–104.

Poplack, Shana & David Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22(1). 99–135.

Shin, Naomi. 2010. Efficiency in lexical borrowing in New York Spanish. *International Journal of the Sociology of Language* 203. 45–59.

Statistics New Zealand, *Profile of New Zealander Responses Quick Stats About culture and identity: 2013 Census*. http://www.stats.govt.nz (accessed November 2015).

Tadmor, Uri. 2009. Loanwords in the world's languages: Findings and results. In Martin Haspelmath & Uri Tadmor (eds.), *Loanwords in the world's languages. A comparative handbook*, 55–75. Berlin: Mouton de Gruyter.

Te Puni Kōkiri – Ministry of Māori Development. *Te Reo Māori statistics*. http://www.tpk.govt.nz/en/whakamahia/te-reo-maori/. (accessed November 2015).

Thompason, S. & T. Kaufman. 1988/1991. *Language contact, creolization and genetic linguistics*. Los Angeles: University of California Press.

Treffers-Daller, Jeanine. 2010. Borrowing. In M. Fried, J. Östman & J. Verschueren (eds.), *Variation and change: Pragmatic perspectives*, 17–35. Amsterdam: John Benjamins.

Van Hout, R. & P. Muysken. 1994. Modelling lexical borrowability. *Language, Variation and Change* 6. 39–62.

Weinreich, U. 1953. *Languages in contact*. The Hague: Mouton de Gruyter.

Winter-Froemel, Esme, Onysko Alexander & Calude Andreea. 2012. Why some non-catachrestic borrowings are more successful than others: A case study of English loans in German. In A. Koll-Stobbe & S. Knospe (eds.), *Language contact in times of globalization*, 119–144. Frankfurt am Main: Lang.

Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2012. Cognitive Sociolinguistics meets loanword research: Measuring variation in the success of Anglicisms in Dutch. *Cognitive Linguistics* 23(4). 749–792.

Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2013. What makes a catchphrase catchy? Possible determinants in the borrowability of English catchphrases in Dutch. In Eline Zenner & Gitte Kristiansen (eds.), *New perspectives on lexical borrowing*, 41–64. Berlin, New York: De Gruyter.

Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2014. Core vocabulary, borrowability and entrenchment. *Diachronica* 31. 74–105.

Zenner, Eline, Dirk Speelman & Dirk Geeraerts. 2015. A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism* 19(3). 333–346.

Zipf, George K. 1935. *The psychobiology of language*. Oxford: Houghton-Mifflin.

# Appendix A.

Māori loans identified and their English counterparts (in alphabetical order) – as they appear in the corpus data (the WSC corpus does not make use of the usual macrons).

| | | |
|---|---|---|
| ae /yes | korero tuku iho /tribal | or Maori language |
| Ahuriri /Napier | history | rimu /red pine |
| ao /world | koro /old man | runanga /council |
| Aorangi /Fielding | koroua /old man | taha Maori /Maori |
| Aotearoa /NZ | kuia /old woman | perspective |
| aroha /love | kumara /sweet potato | takahanga /tramping trip |

(*continued*)

(*continued*)

| | | |
|---|---|---|
| atawhai /show kindness | kura kaupapa /Maori immersion school | takiwa /area |
| atua /demon | kuratini /polytechnic or polytech | Tamakimakaurau / Auckland |
| aue /heck | | tamariki /children or kids |
| hakari /feast | kutai /mussel | tane /husband |
| Hamoa /Samoa | mahi /work | tangata whenua /people of the land |
| hapu /kinship group | mako /shark | |
| haurangi /drunk | manakitauira /student allowance | tangi /funeral |
| hea /where | | taniwha /monster |
| Heretaunga /Hastings | manuka /teatree | tapu /sacred |
| hoe /paddling | Maori /indigenous | Taranaki /Mt Egmont |
| hoha/nuisance | Maori /native | taringa /ear |
| hohonu /deep | maoritanga/maoridom | taurima /carer |
| hui /meeting | matauranga /knowledge | tautoko /support |
| reo irirangi /radio | maunga /mountain | Tautoko /Levin |
| iwi /tribe | moana /sea | Tawhito /Ancient World |
| kaha /strength | mokai /slave | tikanga /custom |
| kai /food | moko /grandchild | tipuna /ancestor |
| kaiarahi /language leader | motoka /car | tirotiro /investigate |
| kaimoana /seafood | moumou taima /wasted time | urupa /cemetery |
| kaitiaki /trustee | | wahine /woman |
| kao /no | ne /eh | waiata /song |
| kaore /none | pa /fortified village | wairua /spirit |
| kapu /cup | Pakeha /European New Zealander | wananga /Maori tertiary institution |
| karakia /prayer | | |
| kaumatua /elder | paua/abalone | whaikorero /orate |
| kaupapa /philosophy | pouaka whakaata / television or TV | whakaiti /belittle |
| kawhe /coffee | | whakapapa /ancestry |
| kea/mountain parrot | pohutukawa /New Zealand Christmas tree | whakarongo /listen |
| kei te pai /well done | | whanau / family |
| ki /to | porangi /insane | whare /house |
| kia ora /goodbye or hello or thank you | pukeko /swamp hen | wharekura /school |
| | putea /fund | wharerokiroki /women's refuge |
| Kiwi /New Zealander | rangatira /chief | |
| koe and koutou /you | rangatiratanga / sovereignty | wharewananga/university or uni or varsity |
| kohanga /preschool | | |
| kohanga reo /Maori immersion preschool | Rarotonga /Mt Smart | whatarangi /platform |
| | reira /therefore | |
| komiti /committee | reo /language | |
| korero /talk | | |

# Appendix B.

The average use of various loanwords, per loan, per group (Māori and Pākehā).

| Meaning | Māori use of loanword | Māori use of non-loanword | Pākehā use of loanword | Pākehā use of non-loanword |
|---|---|---|---|---|
| ae/yes | 37 | 472 | 5 | 2610 |
| Ahuriri/Napier | 0 | 5 | 2 | 9 |
| ao/world | 1 | 73 | 4 | 211 |
| Aorangi/Fielding | 0 | 1 | 1 | 3 |
| Aotearoa/NZ | 9 | 123 | 1 | 631 |
| aroha/love | 10 | 28 | 0 | 119 |
| atawhai/show kindness | 1 | 0 | 2 | 0 |
| atua/demon | 1 | 2 | 0 | 1 |
| aue/heck | 2 | 15 | 0 | 7 |
| hakari/feast | 0 | 0 | 2 | 0 |
| Hamoa/Samoa | 2 | 4 | 0 | 7 |
| hapu/kinship group | 7 | 2 | 0 | 0 |
| haurangi/drunk | 1 | 2 | 0 | 37 |
| hea/where | 8 | 274 | 0 | 1107 |
| Heretaunga/Hastings | 1 | 0 | 0 | 5 |
| hoe/paddling | 2 | 6 | 1 | 0 |
| hoha/nuisance | 4 | 1 | 0 | 9 |
| hohonu/deep | 1 | 8 | 0 | 61 |
| hui(s)/meeting(s) | 27 | 44 | 5 | 123 |
| reo irirangi/radio | 0 | 23 | 0 | 104 |
| iwi/tribe | 17 | 4 | 0 | 6 |
| kaha/strength | 1 | 4 | 0 | 14 |
| kai/food | 7 | 21 | 0 | 98 |
| kaiarahi/language leader | 2 | 0 | 0 | 0 |
| kaimoana/seafood | 2 | 1 | 0 | 0 |
| kaitiaki/trustee | 5 | 14 | 0 | 7 |
| kao/no | 1 | 848 | 0 | 4571 |
| kaore/none | 1 | 10 | 0 | 58 |
| kapu/cup | 0 | 32 | 0 | 81 |
| karakia/prayer | 3 | 1 | 0 | 7 |
| kaumatua/elder | 10 | 2 | 1 | 2 |
| kaupapa/philosophy | 12 | 5 | 0 | 12 |
| kawhe/coffee | 0 | 12 | 0 | 57 |
| kea/mountainparrot | 1 | 0 | 1 | 0 |
| kei te pai/well done | 4 | 1 | 0 | 9 |
| ki/to | 5 | 4420 | 1 | 17,238 |
| kia_ora/goodbye | 2 | 2 | 0 | 9 |

*(continued)*

(*continued*)

| Meaning | Māori use of loanword | Māori use of non-loanword | Pākehā use of loanword | Pākehā use of non-loanword |
|---|---|---|---|---|
| kia_ora/hello | 15 | 16 | 1 | 111 |
| kia_ora/thank you | 2 | 36 | 0 | 184 |
| kiwi/New Zealander | 12 | 14 | 79 | 72 |
| koe and koutou/you | 5 | 4232 | 0 | 17,049 |
| kohanga /preschool | 12 | 1 | 1 | 2 |
| kohanga reo/maori immersion preschool | 18 | 0 | 7 | 0 |
| komiti/committee | 1 | 25 | 0 | 74 |
| korero/talk | 17 | 86 | 2 | 360 |
| korero tuku iho/tribal history | 0 | 0 | 3 | 3 |
| koro/old man | 2 | 12 | 0 | 7 |
| koroua/old man | 5 | 0 | 0 | 0 |
| kuia/old woman | 4 | 3 | 0 | 1 |
| kumara/sweet potato | 1 | 0 | 4 | 1 |
| kura kaupapa/maori immersion school | 17 | 0 | 6 | 0 |
| kuratini/polytech | 0 | 0 | 0 | 32 |
| kuratini/polytechnic | 1 | 7 | 0 | 6 |
| kutai/mussel | 1 | 0 | 0 | 9 |
| mahi/work | 3 | 205 | 0 | 768 |
| mako/shark | 1 | 20 | 0 | 0 |
| manakitauira/student allowance | 1 | 1 | 0 | 9 |
| manuka/tea tree | 0 | 0 | 11 | 0 |
| Maori/indigenous | 67 | 7 | 15 | 3 |
| Maori/native | 696 | 3 | 158 | 8 |
| Maoritanga/maoridom | 1 | 6 | 0 | 4 |
| matauranga/knowledge | 0 | 32 | 1 | 49 |
| maunga/mountain | 0 | 10 | 0 | 38 |
| moana/sea | 0 | 10 | 0 | 54 |
| mokai/slave | 0 | 4 | 1 | 5 |
| moko/grandchild | 12 | 2 | 3 | 2 |
| motoka/car | 1 | 84 | 0 | 370 |
| moumou taima/wasted time | 1 | 3 | 0 | 6 |
| ne/eh | 15 | 500 | 0 | 666 |
| pa/fortified village | 16 | 0 | 0 | 0 |
| Pakeha/Europen New Zealander | 55 | 0 | 36 | 0 |
| paua/abalone | 1 | 0 | 1 | 0 |

(*continued*)

(*continued*)

| Meaning | Māori use of loanword | Māori use of non-loanword | Pākehā use of loanword | Pākehā use of non-loanword |
|---|---|---|---|---|
| poaka whakaata/television | 0 | 68 | 0 | 29 |
| poaka whakaata/tv | 1 | 18 | 0 | 2 |
| pohutukawa/NZ Christmas tree | 0 | 0 | 1 | 3 |
| porangi/insane | 1 | 0 | 0 | 1 |
| pukeko/swamp hen | 0 | 0 | 3 | 0 |
| putea/fund | 4 | 5 | 0 | 13 |
| rangatira/chief | 2 | 8 | 0 | 16 |
| rangatiratanga/sovereignty | 14 | 2 | 0 | 1 |
| Rarotonga/MtSmart | 3 | 0 | 7 | 0 |
| reira/therefore | 1 | 10 | 1 | 38 |
| reo/language | 23 | 88 | 0 | 67 |
| reo/Maori language | 7 | 30 | 0 | 3 |
| rimu/red pine | 0 | 0 | 2 | 0 |
| runanga/council | 2 | 61 | 0 | 85 |
| taha Maori/Maori perspective | 3 | 3 | 0 | 0 |
| takahanga/tramping trip | 1 | 0 | 0 | 0 |
| takiwa/area | 0 | 79 | 1 | 250 |
| Tamakimakaurau/Auckland | 1 | 60 | 0 | 224 |
| tamariki/children | 4 | 102 | 0 | 186 |
| tamariki/kids | 3 | 99 | 0 | 205 |
| tane/husband | 1 | 14 | 0 | 50 |
| tangata/people of the land | 0 | 2 | 1 | 0 |
| tangi/funeral | 4 | 9 | 1 | 11 |
| taniwha/monster | 0 | 2 | 6 | 4 |
| tapu/sacred | 19 | 0 | 3 | 2 |
| Taranaki/Mt Egmont | 16 | 0 | 22 | 2 |
| taringa/ear | 0 | 10 | 1 | 57 |
| taurima/carer | 3 | 0 | 0 | 0 |
| tautoko/support | 0 | 46 | 0 | 114 |
| Tautoko/Levin | 1 | 2 | 0 | 9 |
| Tawhito/Ancient World | 0 | 0 | 1 | 0 |
| tikanga/custom | 17 | 0 | 0 | 4 |
| tipuna/ancestor | 3 | 2 | 0 | 2 |
| tirotiro/investigate | 2 | 3 | 0 | 9 |
| urupa/cemetery | 4 | 1 | 0 | 3 |

(*continued*)

(*continued*)

| Meaning | Māori use of loanword | Māori use of non-loanword | Pākehā use of loanword | Pākehā use of non-loanword |
|---|---|---|---|---|
| wahine/woman | 9 | 182 | 3 | 305 |
| waiata/song | 2 | 18 | 0 | 49 |
| wairua/spirit | 8 | 4 | 0 | 10 |
| wananga/Maori tertiaryinstitution | 1 | 0 | 2 | 0 |
| whaikorero/orate | 9 | 0 | 0 | 0 |
| whakaiti/belittle | 1 | 0 | 0 | 1 |
| whakapapa/ancestry | 4 | 0 | 2 | 6 |
| whakarongo/listen | 1 | 24 | 0 | 124 |
| whanau/family | 35 | 3 | 0 | 0 |
| whare/house | 7 | 95 | 3 | 402 |
| wharekura/school | 1 | 244 | 0 | 424 |
| wharerokiroki/women's refuge | 3 | 5 | 0 | 1 |
| wharewananga/uni | 0 | 1 | 0 | 7 |
| wharewananga/university | 0 | 33 | 1 | 130 |
| wharewananga/varsity | 0 | 11 | 0 | 70 |
| whatarangi/platform | 2 | 5 | 0 | 3 |

# Appendix C.

Fixed effects in the link scale from Generalised Linear Mixed Effects models.

**Table C.1:** Model results for Māori speakers (c.f. Table 1).

| | Estimate | Std. error | z value | Pr(>|z|) |
|---|---|---|---|---|
| Gender – Male | −0.7568 | 0.3352 | −2.258 | 0.023974 |
| Lexicalisation – Non-lexicalised | 1.9099 | 0.7907 | 2.416 | 0.015708 |
| Category – Function Word | −2.7321 | 1.0322 | −2.647 | 0.008125 |
| Cultural or Core – Core | −2.6188 | 0.6241 | −4.196 | 2.71e-05 |
| Excess number of syllables in the English word | 0.7287 | 0.1991 | 3.660 | 0.000252 |

**Table C.2:** Model results for White/European NZ speakers (c.f. Table 2).

|  | Estimate | Std. error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| Polysemous use – yes | −6.1559 | 2.9520 | −2.085 | 0.037043 |
| Lexicalisation – non-lexicalised | 5.0537 | 1.8531 | 2.727 | 0.006389 |
| Category – function word | −5.6819 | 2.5349 | −2.241 | 0.025000 |
| Excess number of syllables in the English word | 1.7605 | 0.4701 | 3.745 | 0.000181 |
| Cultural or core – core | −4.5439 | 1.4217 | −3.196 | 0.001393 |