USING TWITTER IN AN INDIGENOUS LANGUAGE

An analysis of te reo Māori tweets

*Te Taka Keegan**

Paora Mato†

Stacey Ruru‡

Abstract

Language revitalization theory suggests that one way to improve the health of a language is to increase the number of domains where the language is used. Social network platforms provide a variety of domains where indigenous-language communities are able to communicate in their own languages. Although the capability exists, is social networking being used by indigenous-language communities? This paper reports on one particular social networking platform, Twitter, by using two separate methodologies. First, Twitter statistics collated from the Indigenous Tweets website are analysed. The data show that languages such as Basque, Haitian Creole, Welsh, Irish Gaelic, Frisian and Kapampangan do have a presence in the "Twittersphere". Further analysis for te reo Māori (the Māori language) shows that tweets in te reo Māori are rising and peak when certain events occur. The second methodology involved gathering empirical data by tweeting in te reo Māori. This served two purposes: it allowed an ancillary check on the validity of the Indigenous Tweets data and it allowed the opportunity to determine if the number of indigenous-language tweets could be influenced by the actions of one tweeter.

* Senior Lecturer, Computer Science Department, University of Waikato, Hamilton, New Zealand. Email: tetaka@waikato.ac.nz

† PhD candidate, Computer Science Department, University of Waikato, Hamilton, New Zealand.

‡ Summer Internship, Computer Science Department, University of Waikato, Hamilton, New Zealand.

Keywords

Indigenous languages and technology

The advent and escalation of social media has forever changed the look, method and frequency of daily communication and social interaction. Technology-driven social change has been the result, and also the cause, of ongoing new media and the pervading availability of social media networking. The saturation of such media and the need to use the major languages in order to maximize the effect of global market capture emulates some of the effects of language shift in general, as indigenous people opt, for a variety of reasons, to use a language that is perceived to be more dominant, more prestigious and of more use to them. Indigenous-language speakers often use a major language, such as English, at the very least to enable the review of their broadcasts by a wider audience. As the shift to use a majority language continues, the smaller languages are being pushed nearer to the brink of extinction. With the advent of mass communication the number of endangered languages, under threat of extinction, has escalated—especially as minority languages are pushed further into the media background. This is partly due to the resulting propagation of major languages in areas where until now only minor languages were spoken (Katzner, 1995, pp. ix–x). Linguists predict that about half of the 6,000 to 7,000 languages currently spoken are likely to be lost within the next few generations as a result of this language shift (Austin & Sallabank, 2011; Harrison, 2007; Janson, 2012). Most at risk of being lost are the many indigenous languages. (Note: While the term indigenous can be used in a variety of situations, indigenous languages in this paper refer to those minority languages that are indigenous to a particular locale and are endangered,

and indigenous people refers to the speakers of those indigenous languages.)

Social networking platforms such as Facebook and Twitter have become a way for indigenous people all around the world to connect with others from the same language group regardless of geographic distances. A recent linguistic survey (Ungerleider, 2011) has identified 500 languages being used on Twitter, including Gamilaraay, an Australian indigenous language with three remaining speakers. It would seem reasonable that the use of this medium to develop an online language community might have some benefit for strategies aimed at promoting the daily use of an indigenous language and thereby securing some ongoing health for that language.

The continued health of an indigenous language is underpinned by the use of the language in as many forms as practicable (Fishman, 1991, p. 161; Fishman, 2001, p. 73). The development of a vibrant online community, electronically sharing information, ideas and conversation using their own indigenous language, supports the notion of increased language use in a variety of environments.

This paper adds to the literature that investigates language use in social media by indigenous groups. While it has been suggested that simply increasing the amount and range of media content in a language may not assist with its survival (Cormack, 2007, p. 58), and that new media provides new opportunities for dominant languages to be even more dominant (Jones, Cunliffe, & Honeycutt, 2013, p. 654; United Nations Educational, Scientific and Cultural Organization [UNESCO], 2003, p. 11), at least a visible presence by an indigenous language in a new media can suggest the language is relevant to younger speakers (Eisenlohr, 2004, p. 32).

Two methods are used to provide initial data on the presence of an indigenous language in Twitter: a statistical analysis of Twitter data generated by the Indigenous Tweets website, and empirical evidence gathered by a researcher who tweeted in te reo Māori (the Māori language). The work described here extends on earlier analysis that was undertaken by the authors (Mato & Keegan, 2013). This research is significant as it shows that indigenous languages can, and are, using the Twitter platform to extend the environments in which their languages are being used.

Twitter

Twitter is a microblogging website (https://twitter.com) that was founded in 2006 by a small group that included current Twitter CEO Jack Dorsey. The purpose of the website was to create a forum where users could quickly and easily share information. Microblogging means the messages that members send to each other are quite short. In Twitter these messages are known as "tweets" and are restricted in length to 140 characters or less (including hyperlinks and URLs).

Registered users can read and post tweets, but unregistered users can only read tweets. Addressed messages are public tweets that begin with the @ symbol followed by a username, for example, @dorsey, to ensure that the public tweet is also posted into the Twitter home page belonging to the user registered as "dorsey". Members are also able to ensure that tweets from particular users are posted to their own folders or page—this is known as following. Privacy protocols can be set to make registered users' tweets only visible to the members who have been accepted as followers, although this is a relatively rare occurrence in Twitter. Private tweets can occur when a user privately tweets another user—this message will not be shown on either user's home page.

Tweets are also able to be forwarded without amendment. This is known as retweeting and these tweets are easily identified by a retweeted message banner that includes the username of the person who is retweeting. Hashtags (#) are used to categorize messages, trace topic threads and denote specific events. For instance, including #olympicgold in a message would relate it to all other tweets with the same hashtag. Twitter also has some useful search functions. Members are able to search for specific phrases, ideas or keywords to see what others are saying about the topic.

The popularity of Twitter has increased remarkably since its inception. At the start of 2014, Twitter had 645,750,000 active users that were sending 58 million tweets per day (Statistic Brain, 2014). The Twitter website has been listed in the top 10 most visited websites (Alexa, 2014) on the Internet.

One feature of the Twitter platform is that it is language-agnostic; messages can be sent in any language scripted in the Unicode Standard (see Unicode.org). That means there is nothing to stop someone broadcasting messages using any language they choose. One source suggests only 34% of tweets are in English (Fox, 2013) though a recent article in the *New York Times* claims this figure is closer to 51% (Seshagiri, 2014). In any case, the agnostic nature of the Twitter platform means it is available as a medium for use by indigenous languages.

The Indigenous Tweets website

Indigenous-language tweeting is, quite simply, using the Twitter platform to post messages that are written in an indigenous language. One particular website has collated statistics to determine the extent of indigenous-language tweeting.

The Indigenous Tweets website (http://indigenoustweets.com/) collates, groups and provides summary statistics for tweets of selected languages. The site was developed by Professor Kevin Scannell of the Department

Language	Users	Tweets ▾	Top User	Tweets	First Tweet
Euskara	17048	6840486	berria	62358	eastigarraga
Kreyòl Ayisyen	14267	4630231	amour109	72983	tichrist
Cymraeg	14244	3573458	newyddcymraeg	74315	meigwilym
Kapampangan	1379	2075603	keeyttguevarra	21759	desperada
Gaeilge	7426	781164	Tuigim	42767	imeall
Frysk	2655	713471	omropfytsban	72887	eetweetje
Setswana	314	683781	sesutho	40595	WameDre
Asturianu	770	365439	iyangc	23042	Pingarates
Hausa	1331	349524	bbchausa	29047	mojaam
Yorùbá	2239	260048	weirdmcofficial	6250	kojere
Soomaaliga	558	190888	Weedhsan	16898	HaPpYMaXaMeD
Ikinyarwanda	289	153502	TweetRwanda	26761	kwitonda
Igbo	1179	140135	uyaka1	8960	Igbopeople
Gàidhlig	662	127867	sconewt	27303	ruighean
Boarisch	5	122838	bayernuhr	120071	steffkellerband
Māori	347	102268	maonewt	29810	DigitalMaori
Gaelg	28	89215	glvold	40479	greinneyder
Aragonés	127	83786	Purnas	14487	jesmar
Chamoru	9	62437	chanewtest	25699	jenccamacho
Tamasheq	3	57363	tmjnew	27484	tmjbbl
مصری	25	55155	Miadhu_dv	12643	muraasil
Malagasy	252	47771	Ainjah	6197	elsifaka
Lingua Latina	65	44472	NemoOudeis	7608	henduu14
Winaray	33	44153	ilovetacloban	7221	piamaemendz
Cántabru	14	42758	DieguSG	8416	r_macho
Chicheŵa	144	41529	skinnybrizzo	8773	caseynesbit

INDIGENOUS TWEETS.COM

Blog

Indigenous Blogs

Kevin Scannell

Follow @IndigenousTweet

FIGURE 1. The Indigenous Tweets website home page.

of Mathematics and Computer Science at the University of Saint Louis, Missouri, USA. Scannell (2011) states the aim of this website is to "help build online language communities through Twitter", hoping that the site will enable speakers of indigenous languages to connect in their own languages amidst "a vast sea of English, French, Spanish and other global languages that dominate Twitter" (para. 3). A screenshot of the home web page is displayed in Figure 1.

The Indigenous Tweets website was launched in March 2011 and initially catalogued tweets from 35 of the indigenous languages identified by the Crúbadán Project (Scannell, 2007). The website compares tweets with a database of corpora for almost 500 indigenous languages. This database was developed and is updated by the Crúbadán Project. By May 2011 the number of indigenous languages identified had more than doubled, with tweeter profiles for over 80 indigenous languages displayed on the website. User profiles are currently (March 2014) listed for 157 indigenous minority languages.

The matching of tweets to a particular language is a two-stage process. The first stage uses statistics of character sequences, specifically three-character sequences (known as 3-grams or tri-grams), to determine if tweets are similar in nature to the minority languages listed in the database. Then a word-based classifier is used to determine probabilities of which particular language the tweet most likely matches. The classifier uses a model that considers many characteristics, including number of words in each language, word sequences and word frequencies

in each language (K. Scannell, personal communication, 6 March 2014). While testing the accuracy of the Indigenous Tweets classifier was beyond the scope of this paper, the Māori tweets that were posted by our researcher were all correctly classified as Māori tweets.

Given the prodigious amount of tweets that are sent each day (over 50 million) it would be impractical to scan every single tweet to determine if the tweet was using an indigenous language. Consequently, tweets are randomly scanned. However, once a tweet has been identified as an indigenous-language tweet, the user's profile is added to a list of indigenous-language tweeters and their followers' profiles are targeted for scanning. This profile list is regularly scanned, as are their followers. Therefore, the catalogue of indigenous-language tweets is collated more by the interaction of indigenous-language tweeters than the scanning of all tweets. This method has proven to be remarkably effective in this research. Our researcher began using Twitter in the Māori language (as part of this research), and within two days her profile appeared on the Indigenous Tweets website.

Indigenous tweets statistics

The Indigenous Tweets website suggests that tweeting in indigenous languages is occurring for at least 157 indigenous languages. This figure, initially at least, might appear quite promising—a foothold for a small percentage of indigenous languages. However, closer examination shows the data are less encouraging for some of the languages, as they reveal relatively low numbers of indigenous-language tweeters.

During the initial analysis, two indicators were considered:

1. the number of people tweeting in an indigenous language, and
2. the number of indigenous-language tweets being sent.

The breakdown for the number of people that are tweeting per indigenous language is shown in Figure 2. The figures, from the Indigenous Tweets website, suggest a large number of languages have only a small number of people tweeting in that language. There are 29 languages that only have one person tweeting in that language, and a further 29 languages that only have two people tweeting in that language. Ninety-six languages have 10 or less people
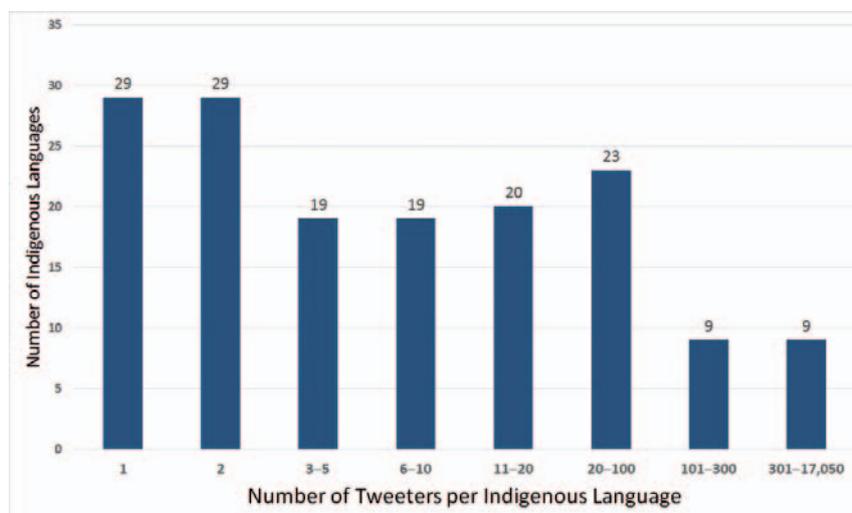
FIGURE 2. Number of people tweeting per indigenous language (data retrieved 10 March 2014 from Indigenous Tweets website).
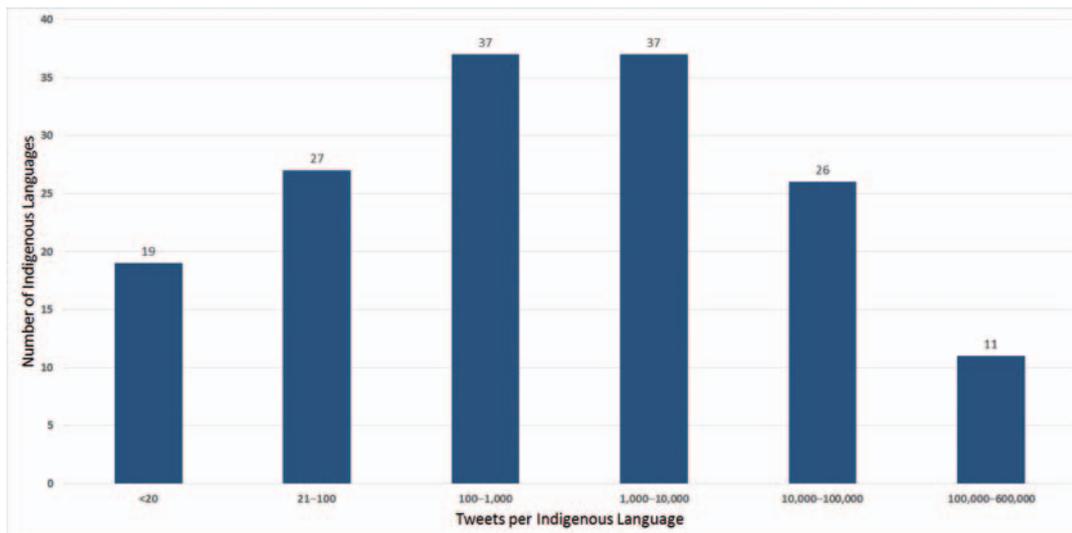
FIGURE 3.   Tweets (and retweets) per indigenous language (data retrieved 10 March 2014 from Indigenous Tweets website).

tweeting, 43 have between 10 and 100 tweeters and 18 languages have more than 100 people tweeting in their particular language. Almost 74% of the selected 157 languages have less than 20 tweeters and nearly 89% have less than 100 users tweeting in the language.

The recorded number of tweets that are being sent in an indigenous language is displayed in Figure 3 (these figures include retweets). It is evident that 46 indigenous languages (19 + 27) have had less than 100 tweets posted in their language. A further 37 languages have had between 100 and 1,000 tweets posted, 37 languages have had between 1,000 and 10,000 tweets, and 37 languages fall into the last two categories of having more than 10,000 tweets posted in their language.

Table 1 and Table 2 rank indigenous tweets by the top 20 languages. Table 1 ranks by the total number of registered users. Table 2 ranks by the total number of tweets. The language name displayed in the first column is the indigenous-language name for the language, while the fourth column displays the English-language version of that name.

The same languages feature in the top six on both tables. This would suggest healthy online language communities for the languages of Basque, Haitian Creole, Welsh, Irish Gaelic,

Frisian and Kapampangan. All of these languages have over 1,000 users, suggesting these indigenous-language communities are actively communicating on Twitter.

It should be noted that Haitian Creole is a special case on the Indigenous Tweets website. Due to constraints that include a limitation on the number of queries per day and the large number of people in the Haitian Twitter community, Haitian Creole is the only language on the Indigenous Tweets website for which Twitter wasn't completely trawled for specific language tweets (K. Scannell, personal communication, 22 November 2013). The numbers listed for this language should be regarded as indicative only—they are likely to be higher—and should not be used in comparisons with other languages.

Indigenous tweets top user statistics

Analysing the statistics of top users is important because it can indicate whether a language community has many users tweeting in an indigenous language or it has few users who are tweeting often in an indigenous language. One might expect that many users tweeting often in their own language would reflect a language community fully engaging in this form of social media.

TABLE 1. Number of people tweeting per indigenous language (data retrieved 10 March 2014 from Indigenous Tweets website).

Language	Tweeters	Rank	English name
Euskara	17,047	1	Basque
Kreyòl Ayisyen	14,267	2	Haitian Creole
Cymraeg	14,235	3	Welsh
Gaeilge	7,409	4	Irish Gaelic
Frysk	2,049	5	Frisian (Netherlands)
Capampañgan	1,379	6	Kapampangan (Philippines)
Asturianu	767	7	Asturian (Spain)
Gàidhlig	658	8	Scottish Gaelic
Soomaaliga	558	9	Somali (Africa)
Māori	345	10	Māori (New Zealand)
Brezhoneg	333	11	Breton (France)
Setswana	314	12	Setswana (Africa)
Hausa	302	13	Hausa (Africa)
ייִדיש	262	14	Yiddish
Malagasy	252	15	Malagasy (Madagascar)
Chicheŵa	143	16	Chichewa (Africa)
Davvisámegiella	134	17	Northern Sami
Aragonés	105	18	Aragonese (Spain)
Occitan	99	19	Occitan (France)
Kernewek	98	20	Cornish

TABLE 2. Number of tweets per indigenous language (data retrieved 10 March 2014 from Indigenous Tweets website).

Language	Tweets	Rank	English name
Euskara	5,931,646	1	Basque
Kreyòl Ayisyen	4,294,836	2	Haitian Creole
Cymraeg	3,058,972	3	Welsh
Kapampangan	1,906,256	4	Kapampangan (Philippines)
Gaeilge	640,496	5	Irish Gaelic
Frysk	622,095	6	Frisian (Netherlands)
Setswana	617,553	7	Setswana (Africa)
Asturianu	316,658	8	Asturian (Spain)
Hausa	263,249	9	Hausa (Africa)
Soomaaliga	166,363	10	Somali (Africa)
Gàidhlig	109,935	11	Scottish Gaelic
Māori	93,283	12	Māori (New Zealand)
Gaelg	87,772	13	Manx Gaelic
Aragonés	67,383	14	Aragonese (Spain)
Chamoru	62,155	15	Chamoru (Mariana Islands)
Tamasheq	57,363	16	Tamasheq (Timbuktu)
ייִדיש	52,966	17	Yiddish
Malagasy	43,381	18	Malagasy (Madagascar)
Winaray	42,183	19	Winaray (Philippines)
Chicheŵa	37,765	20	Chichewa (Africa)

TABLE 3. Top tweeters per indigenous language (data retrieved 10 March 2014 from Indigenous Tweets website).

Language	Users	Tweets	Top user	Tweets	%
Basque	17,047	5,931,646	berria	52,909	0.9
Kapampangan	1,379	1,906,256	itsmeshellie	19,469	1.0
Haitian Creole	14,267	4,294,836	amour109	62,341	1.5
Welsh	14,235	3,058,972	newyddcymraeg	66,511	2.2
Irish Gaelic	7,409	640,496	aonghusoha	33,277	5.2
Setswana	314	617,553	sesutho	33,539	5.4
Asturian	767	316,658	iyangc	19,682	6.2
Emiliano-Romagnolo	53	30,746	ingiro_concerti	2,419	7.9
Northern Sami	134	22,881	tuejjlaz	1,955	8.5
Hausa	302	263,249	bbchausa	25,390	9.6
Somali	558	166,363	Weedhsan	16,433	9.9
Frisian	2,049	622,095	omropfytsban	66,700	10.7
Malagasy	252	43,381	Ainjah	5,951	13.7
Southern Sami	17	1,073	guektiengieline	150	14.0
Hawaiian	28	1,966	malamaikaaina	309	15.7
Cornish	98	17,688	RodTLyon	2,786	15.8
Waray-Waray	33	42,183	ilovetacloban	6,934	16.4
Yoruba	64	27,171	alakoweyoruba	4,928	18.1
Latin	56	36,921	NemoOudeis	6,698	18.1
Breton	333	27,585	ebrezhoneg	5,201	18.9

There are four indigenous languages that have tweeted over three million times: Basque, Kapampangan, Haitian Creole and Welsh. In these languages the top tweeter is responsible for less than 5% of the total tweets for that language (see Table 3). This seems to reflect indigenous-language communities that are engaging with the Twitter platform and indicates the statistics are not being skewed by a small number of fervent users.

In contrast, a single tweeter in Māori is responsible for 33% of the Māori-language tweets and a single user in Scottish Gaelic is responsible for 29% of their total language tweets (these figures are not shown in Table 3). This suggests the Twitter statistics for these indigenous-language communities are being heavily influenced by the actions of one or a few users and could also indicate the media is not being used extensively by the language community.

Māori-language tweeting

The Indigenous Tweets website provides statistics for each of the indigenous languages that it has profiled. Earlier research (Mato & Keegan, 2013) indicated that the top three tweeters (shown in Table 6)—maonewt, maoold and maobbl—post religious-based tweets that appear to be Bible passages. The tweets contain Māori-language translations of passages from the New Testament (maonewt), the Old Testament (maoold) and the Bible in general (maobbl). On closer investigation the three users appear to be the same person/organization originating from California, USA. It appears that a simple script has been written to periodically tweet bible scriptures, hence the robotic nature of the tweets. The total number of Māori-language tweets from these three profiles is 63,820, which account for 68.4% of the total Māori-language tweets. Given the

significant number of these tweets and that they are not the result of personal postings, the data relating to these profiles are not included in the following analysis of Māori-language tweeting. This has been done to ensure that the further in-depth analysis is not skewed by a few profiles (in this case three) and that the outcomes are reflective of postings by actual people.

With the robotic religious tweets removed, the number of Māori-language tweets sent per month for the years 2011–2013 is shown in Table 4. Tweets have been steadily increasing since 2011. Tweeting increased from 2011 to 2012 by 294% (1,566–4,611) and from 2012 to 2013 by 282% (4,611–12,999). In 2013 an average of 1,083 tweets (and retweets) in te reo Māori were posted every month, equating to 36.5 tweets per day.

The number of tweets has many peaks and troughs throughout the year, as indicated by Table 4. A clearer representation of these peaks and troughs has been provided by collating the number of tweets by week and plotting the outcomes—see Figure 4.

Some spikes in Figure 4 are clearly evident. The spikes that appear in July 2012 and July

TABLE 4. Tweets and retweets in te reo Māori per month 2011–2013 (data retrieved 10 March 2014 from Indigenous Tweets website).

| Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Total
42	76	169	128	123	208	123	129	98	121	221	128	1,566
Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	
155	201	220	251	328	297	856	376	354	740	509	324	4,611
Jan-13	Feb-13	Mar-13	Apr-13	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13	
582	814	578	781	1,032	1,884	1,941	811	851	1,521	1,426	778	12,999
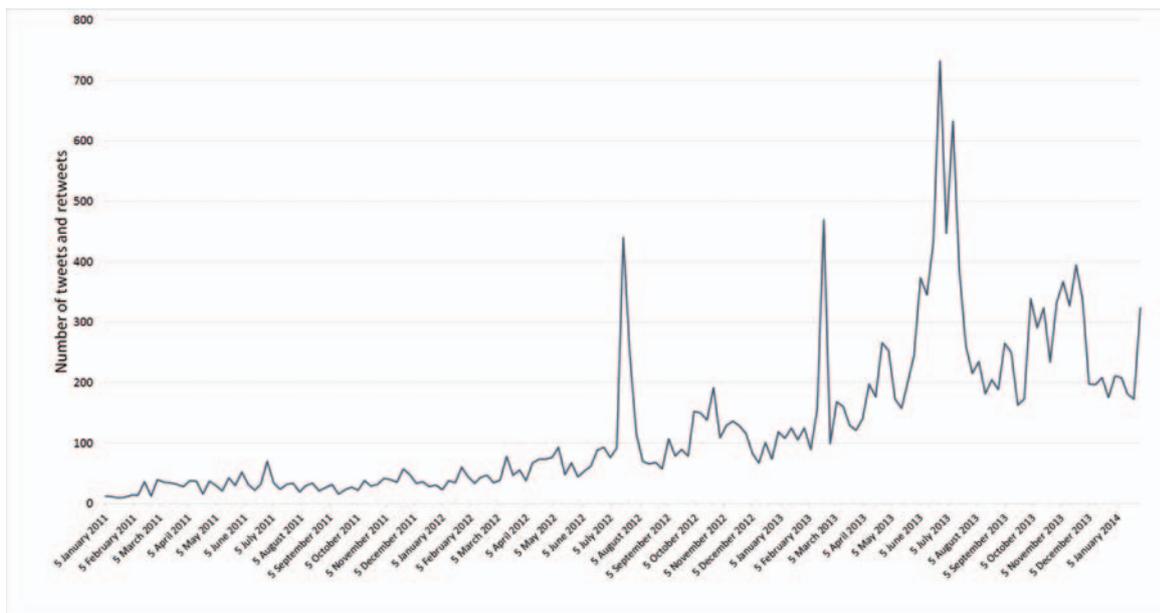
FIGURE 4. Tweets and retweets in te reo Māori per week 2011–2013 (data retrieved 10 March 2014 from Indigenous Tweets website).

2013 correspond to Māori Language Week. Māori Language Week is a government-sponsored initiative that encourages all New Zealanders to promote the use of te reo Māori. A large spike in the last week of February 2013 corresponds to the 2013 Te Matatini National Kapa Haka Festival (see www.tematatini. co.nz). Clearly these two events are significant and impact on when users choose to tweet in te reo Māori.

How many are tweeting in te reo Māori?

Given that tweeting is happening in te reo Māori, how many actual people are tweeting in te reo Māori? The number of unique tweeters has been captured for each week of the period January 2011 to December 2013—see Table 5. These numbers have then been averaged to show the mean number of users per week for each month—see Table 5. The data show that, on average, the number of people that are tweeting in te reo Māori each week has almost doubled each year for 2011–2012 and 2012–2013.

The number of unique users posting Māori-language tweets has also been graphed and is displayed in Figure 5 for each week of the period January 2011 to December 2013.

Peaks are clearly visible (see Figure 5) in July 2011, July 2012 and July 2013—the

TABLE 5.   Average number of unique tweeters in te reo Māori per week per month 2011–2013 (data retrieved 10 March 2014 from Indigenous Tweets website).

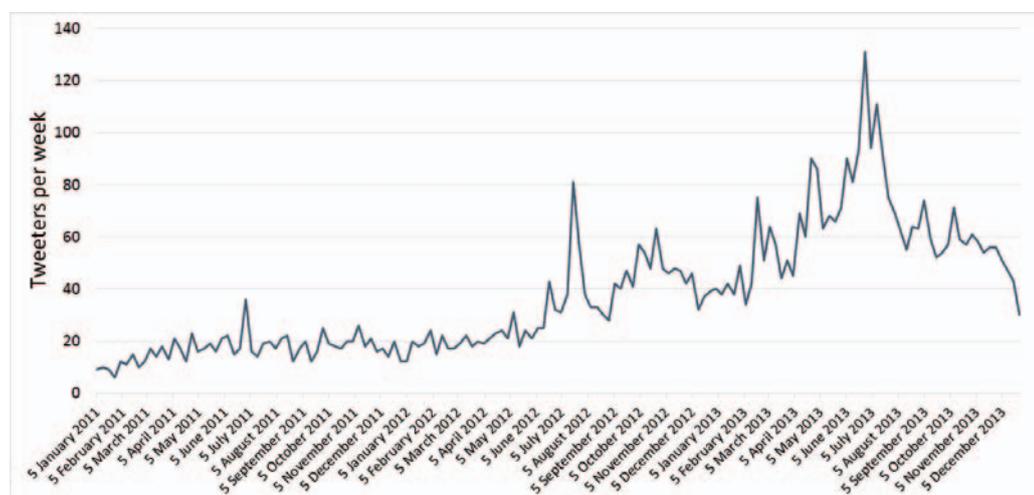| Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Total
9	12	12	18	17	22	17	18	18	19	20	16	**16**
Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	
17	19	20	22	23	31	52	32	43	54	16	39	**33**
Jan-13	Feb-13	Mar-13	Apr-13	May-13	Jun-13	Jul-13	Aug-13	Sep-13	Oct-13	Nov-13	Dec-13	
41	51	54	66	71	99	88	61	60	61	56	43	**63**

FIGURE 5.   Unique tweeters in te reo Māori per week 2011–2013 (data retrieved 10 March 2014 from Indigenous Tweets website).

weeks that the Māori language is promoted during Māori Language Week. However, the peaking of unique users in February 2013 (see Figure 5) is not as significant as the peaking of Māori-language tweets in February 2013 (see Figure 4). This suggests that the event that occurred at that time (Te Matatini) did not lead to a significant number of new people tweeting in te reo Māori. Rather, it suggests that people posted considerably more tweets in te reo Māori during that time. One would suspect that a comparative analysis of the tweets during the Māori Language Weeks and the tweets during Te Matatini would reveal different content based on their contexts; the former favouring more introductory language tweets and the latter containing more performance appraisal tweets.

Māori tweeting geography

Twitter functionality enables determination of where people are tweeting from (Scannell, 2013). Some tweets are geo-location activated. That is, when a tweet is sent from a mobile device a longitude/latitude reading is captured in certain instances and recorded with the tweets. Other users provide locations with their Twitter profiles; longitude/latitude readings can be derived from this information. Our gratitude is extended to Professor Kevin Scannell of the University of Saint Louis, who has derived and processed these data to create geographical profiles of Māori-language tweeting.

The locations from where people are tweeting in te reo Māori are displayed in Figure 6. The larger the circle, the higher the number of tweets from that area. It is evident that the highest number of tweets are sent from the larger centres: Auckland, Wellington and Christchurch. There is also a significant number of Māori-language tweets being sent from some other less populous areas such as Nelson, Picton, Palmerston North, Tauranga and some locations in Northland (north of Auckland). Many of these locations are areas that do not have high concentrations of Māori-language

FIGURE 6.   Origins of Māori-language tweets up to 2013 (graph supplied by Professor Kevin Scannell of the University of Saint Louis, March 2013).

speakers—further research is needed to determine why these locations are appearing high in these statistics and to determine the relationships between isolation and use of Twitter.

Twitter conversations are displayed in Figure 7. The lines represent the two locations

FIGURE 7.   Māori-language Twitter conversations up to 2013 (graph supplied by Professor Kevin Scannell of the University of Saint Louis, February 2013).

of the users who are conversing via Twitter. Similar features are apparent between Figure 7 and Figure 6; that is, while most conversations are happening between tweeters in the major centres of Auckland, Wellington and Christchurch, they are also happening between other areas such as Nelson, Picton, Palmerston North, Tauranga and locations in Northland.

Likely scenarios for high levels of tweets in generally less populated areas could be attributed to the presence of tertiary institutions (Palmerston North and Tauranga), ferry terminals and ports (Picton, Nelson and Tauranga), and tourist attractions (Northland), possibly reflecting higher numbers of "transient" tweeters.

Tweeters of te reo Māori

The data for the top 15 tweeters of te reo Māori are displayed in Table 6.

A large percentage of tweets (68.1%) posted in te reo Māori were posted by what appears to be three Twitter robotic scripts (as described earlier). These users/robots were tweeting religious texts. Two of the user accounts have been suspended (maonewt, maoold) and the third hasn't posted for almost a year.

Many of the users (7 in 15) who tweet in te reo Māori, as displayed in Table 6, appear to be acting as individuals. These users are temihinga, necromaori, MaoriStream, CmmdrZed, MsMocktavia, Wayne_Abraham and RatuTibble. Users such as temihinga and necromaori have a high percentage of Māori-language tweets (72.8% and 71.7% respectively) while some of the other users such as MsMocktavia and MaoriStream have low percentages of tweets in te reo Māori (2.2% and 3.3% respectively).

The remaining users represent news/media organizations—waateanews, TeKarereTVNZ and TeKaeaOfficial—and users that appear to be tweeting for business purposes—HURIMOZ and MaoriPacificJob. HURIMOZ could warrant further investigation. This user is tweeting from Moorea, in French Polynesia, and appears to be based around a website offering translations in eight languages that include reo Tahiti (Tahitian language) and te reo Māori.

TABLE 6.   Māori-language tweeters (top 15 by total tweets) (data retrieved 10 March 2014 from Indigenous Tweets website).

Username	Māori tweets	Total tweets	% in Māori	Followers	Following	Most recent tweet
maonewt	29,810	32,020	93.1	21	0	16 May 2013
maoold	22,390	24,659	90.8	17	0	21 June 2012
maobbl	11,620	12,507	92.9	12	0	5 June 2013
HURIMOZ	3,878	124,733	3.1	1,086	1,846	17 March 2014
temihinga	1,382	1,898	72.8	428	642	17 March 2014
necromaori	1,112	1,551	71.7	318	807	17 March 2014
MaoriStream	921	27,925	3.3	771	1,230	16 March 2014
waateanews	816	9,596	8.5	1,657	970	17 March 2014
CmmdrZed	776	19,393	4	319	395	17 March 2014
MsMocktavia	774	35,660	2.2	512	130	17 March 2014
TeKarereTVNZ	717	3,256	22	1,746	384	16 March 2014
TeKaeaOfficial	668	2,710	24.7	1,122	523	17 March 2014
Wayne_Abraham	630	1,542	40.9	122	250	16 March 2014
MaoriPacificJob	562	13,491	4.2	1,476	1,462	17 March 2014
RatuTibble	545	3,428	15.9	251	281	15 March 2014

Given the similar syntax of the Tahitian and Māori languages, and the short nature of the tweets, it may be difficult, at times, to accurately define which language is being used in each tweet.

A more comprehensive analysis of the user profiles of those who tweet in te reo Māori is detailed in Mato and Keegan (2013).

Infrequent Māori-language tweeters

The user profiles discussed above give details for the 15 te reo Māori tweeters who post the highest number of tweets in te reo Māori. However, there are a further 330 users who have tweeted in te reo Māori. Many of these users have only tweeted a small number of tweets in Māori. Only 65 (of 345) Māori-language tweeters have posted more than 100 tweets in Māori, the median number of tweets being 18 (that is, half of the tweeters in te reo Māori posted 18 tweets or less), and a quarter (85) of the users have posted less than 5 tweets in te reo Māori.

Another characteristic of Māori-language tweeting is that many of the users are using the Māori language in a small percentage of their tweets. Almost 90% (298 of 345) of the users have posted less than 40% of their tweets in Māori. Of the 47 users who have tweeted in te reo Māori more than 40% of the time, less than half (22) have posted 10 tweets or more. This indicates that the actual community of Māori-language tweeters is much smaller than it first appears, with perhaps as few as only a dozen users interacting regularly in te reo Māori.

Influencing Māori-language tweeting

In order to determine if the incidence of tweeting that was occurring in te reo Māori could be actively encouraged from within the Twitter platform itself, a small study was undertaken. A Twitter account was created by an undergraduate research student while she was working on a summer research internship. The researcher used a number of strategies over an 11-week period to encourage the use of te reo Māori in tweets.

The researcher created a Twitter account and then sought out users who were tweeting in te reo Māori. These users were identified from the Indigenous Tweets website, using the Twitter search facility, and further analysed to determine whom they were following and who was following them. These users were then "followed" by the researcher so that she could interact with them when they posted in te reo Māori.

The researcher, over an 11-week period, from 17 November 2013 to 25 January 2014, employed strategies to generate te reo Māori tweets and to engage with tweeters of te reo Māori. These strategies were:

- Tweet questions to elicit responses.
- Tweet comments on topical issues.
- Encourage conversations by replying to tweets and retweeting Māori-language tweets.

The strategies were deployed over the 11 weeks, aside from a 2-week period when no Twitter activity by the researcher occurred (due to the Christmas holiday period). In general, all tweets and strategies were conducted using te reo Māori. There were three tweets posted bilingually (Māori and English) to see if this would elicit more responses, but it did not appear to make any difference. On occasion users would respond in English, and consequently some conversations were bilingual.

Māori-language tweeting experience

An initial question about tweeting in te reo Māori received no responses the first day so this was retweeted the following day, which led to some replies and some Twitter conversations. More questions were tweeted over the following week, which led to replies, some retweeting and some more conversations. Topics of interest

were then tweeted, which again led to more replies, retweets, favourites (a form of tagging) and conversations. Conversations regarding specific topics attracted responses from users who, perhaps not surprisingly, had an interest in those particular topics. The use of retweeting, by both the researcher and others, led to an increase in the number of te reo Māori tweets since the retweets showed up as a tweet on the home page of each retweeter. Replying to other users' tweets served to generate conversation because tweets tended to invite replies, which, in turn, attracted further response. As mentioned earlier, topics of mutual interest also resulted in more sustained dialogue.

Some users responded totally in Māori; others used a mixture of Māori and English. However, when tweets were posted bilingually—in both Māori and English—there were few responses. Given the short time frame in which this strategy was deployed, the low responses may have just been due to timing, or due to topics that were not interesting to te reo Māori tweeters. In the final weeks a topic of interest from the Māori news programme Te Kāea was tweeted. Nine topics were tweeted, two of which were responded to.

Over the 11 weeks 154 tweets were posted by the researcher, as displayed in Table 7.

The purpose of the study was to get a feel for the te reo Māori tweeting environment and to determine if the number of tweets in Māori could be pro-actively increased. The field work was somewhat self-fulfilling: the number of tweets in te reo Māori was increased by the researcher's tweets and responses to these tweets. This was an expected outcome. What this field work did provide was an opportunity to test the validity of the statistics that were being generated by the Indigenous Tweets website. All of the researcher's tweets were identified by the Indigenous Tweets website and correctly tagged as te reo Māori tweets. A difference in time zone meant the time stamps recorded with the tweets were different to the local time that the tweets were posted—again, an outcome that was expected.

In terms of the te reo Māori tweeting environment, the researcher noted that:

- Users will respond to te reo Māori tweets (in te reo Māori) if the topic is of interest to them.
- Te reo Māori tweeters are willing to begin conversations with other te reo Māori tweeters.
- Some users will use English in te reo Māori tweets.
- It appears that tweeting is more likely at certain times in the day.

Discussion from Māori tweeting field work

It became evident that much of the conversation needed to be initiated by the researcher, and users at times were slow to reply. When the topics raised were mutually interesting there was some extended engagement. However, quiet patches were experienced. It appears that the time of day might influence responses from others; for example, a reduction in Twitter conversations during the 5pm to 7pm period may be due to

TABLE 7. Twitter activity in 11-week study.

	Researcher tweets & retweets	Te reo Māori tweets & retweets	%
Week1	36	346	10.4
Week2	28	344	8.1
Week3	17	282	6.0
Week4	10	161	6.2
Week5	11	235	4.7
Week6	0	177	
Week7	0	192	
Week8	12	240	5.0
Week9	13	171	7.6
Week10	18	187	9.6
Week11	9	231	3.9
Total	**154**	**2,566**	**7.0**

post-work commuting, after-work activity or dinner preparation. The Christmas period also had some impact on tweeting probably due to that time being largely family-oriented. It is also possible that the topics of conversation and the questions weren't engaging enough.

It was noticeable that mixtures of Māori, English and bilingual tweets tended to extend some discussions. This might be a point to remember if it is suspected that users are shying away from tweeting because their grasp of the language isn't entirely fluent. A lack of fluency may also apply to people who followed but didn't directly participate. Arguably, strategies that would motivate these users to engage, even if only retweeting, would seem to have benefits for developing an online language community. One such strategy might be to seek help or advice on certain matters. It appears that people are only too willing to share information and resources and, in some cases, to share some advice. One example, when a user noticed the lack of a tohutō (macron) over the long vowel, was the sharing of an app name to enable use of the tohutō.

Future work

The potential exists for much future work in this medium. A similar analysis to the one described in this paper for te reo Māori could be undertaken on any of the other 150+ indigenous languages listed on the Indigenous Tweets website. Longitudinal studies over a longer time period could be undertaken to ascertain current trends. Specific language communities—for example, education institutes—could be encouraged to use the Twitter platform to assist with communications in educational programmes.

Further analysis could be undertaken on the content of the indigenous-language tweets:

- Which technical terms are and are not translated?

- How are diacritics used and does this depend on the device that is used to tweet?
- What topics are discussed most?
- How much content is conversational (bidirectional) and how much is just dissemination (unidirectional)?
- How often does language switching occur and why?
- What are the motivations for choosing to tweet in an indigenous language?

For promoters of an indigenous language there exist possible avenues to contact and encourage language usage amongst low percentage tweeters of a language. Also, followers of indigenous-language tweeters are a potential source of people who are interested in a language but for some reason choose not to tweet, or not to tweet very often in that language.

Conclusion

The Indigenous Tweets website has identified that indigenous languages are using Twitter to converse. Some languages such as Basque, Haitian Creole and Welsh appear to be using the platform to actively converse in a "natural" manner. Other languages have a small number of users tweeting and only show a small number of tweets.

With regards to te reo Māori, there does seem to be a large number of tweets in te reo Māori (over 90,000); however, many of these tweets have been posted for religious or commercial reasons. Encouragingly, there does appear to be a number of individuals who are using Twitter to converse in te reo Māori. The numbers suggest there are two types of conversational te reo Māori tweeters: those that are passive and post a low percentage of Māori tweets, and those that are active and post a high percentage of Māori tweets. It seems that the passive tweeters become more active if there is a purpose for them to do so; Māori Language

Week and the kapa haka (traditional Māori performing arts) competitions of Te Matatini being the two examples that appeared prominently in the statistics. The active users were willing to converse with our researcher through retweeting, the following of tweets, the marking of favourites and in personal conversations, especially when the topic of the tweet was interesting to them. In general it appeared that the active tweeters of te reo Māori were acting in isolation; they were choosing when and when not to respond to tweets in Māori. There did not appear to be a large community of Māori-language speakers using the Twitter platform to converse in a "natural" manner, but instead a small community that has doubled in size each year for the past two years.

From a technical perspective Twitter does appear to be an appropriate medium to support indigenous-language conversation; however, to assist with language revitalization this needs to be realized by the indigenous-language communities themselves.

Glossary

kapa haka	traditional Māori performing arts
reo Tahiti	Tahitian language
te reo Māori	the Māori language
tohutō	macron—a symbol to mark long vowels

References

Alexa. (2014). *The top 500 sites on the web*. Retrieved from www.alexa.com/topsites

Austin, P., & Sallabank, J. (Eds.). (2011). *The Cambridge handbook of endangered languages*. Cambridge, England: Cambridge University Press.

Cormack, M. (2007). The media and language maintenance. In M. Cormack & N. Hourigan (Eds.), *Minority language media: Concepts, critiques and case studies* (pp. 52–68). Clevedon, England: Multilingual Matters.

Eisenlohr, P. (2004). Language revitalization and new technologies: Cultures of electronic mediation and the refiguring of communities. *Annual Review of Anthropology*, *33*(1), 21–45.

Fishman, J. (1991). *Reversing language shift: Theoretical and empirical foundations of assistance to threatened languages*. Clevedon, England: Multilingual Matters.

Fishman, J. (2001). *Can threatened languages be saved? Reversing language shift, revisited: A 21st century perspective*. Clevedon, England: Multilingual Matters.

Fox, Z. (2013). *Top 10 most popular languages on Twitter*. Retrieved from http://mashable.com/2013/12/17/Twitter-popular-languages/?utm_cid=mash-com-Tw-main-link

Harrison, K. D. (2007). *When languages die: The extinction of the world's languages and the erosion of human knowledge*. New York, NY: Oxford University Press.

Janson, T. (2012). *The history of languages: An introduction*. Oxford, England: Oxford University Press.

Jones, R. J., Cunliffe, D., Honeycutt, Z. R. (2013). Twitter and the Welsh language. *Journal of Multilingual and Multicultural Development*, *34*(7), 653–671.

Katzner, K. (1995). *The languages of the world*. London, England: Routledge.

Mato, P., & Keegan, T. T. (2013). Indigenous tweeting for language survival: The Māori-language profile. *International Journal of Technology and Inclusive Education*, 2(2): 184–191.

Scannell, K. (2007). The Crúbadán project: Corpus building for under-resourced languages. In C. Fairon (Ed.), *Building and exploring web corpora (WAC3-2007): Proceedings of the 3rd Web as Corpus Workshop* (pp. 5–15). Louvain-la-Neuve, Belgium: Presses univ. de Louvain.

Scannell, K. (2011). *Welcome/Fáilte!* [Blog post].

Retrieved from http://indigenoustweets.blogs-pot.co.nz/2011/03/welcomefailte.html

Scannell, K. (2013). *Mapping the Celtic Twittersphere* [Blog post]. Retrieved from http://rising.globalvoicesonline.org/blog/2013/12/31/mapping-the-celtic-Twittersphere/

Seshagiri, A. (2014, 9 March). The languages of Twitter. *New York Times: Bits*. Retrieved from http://bits.blogs.nytimes.com/2014/03/09/the-languages-of-twitter-users/?_r=0

Statistic Brain. (2014). *Twitter statistics*. Retrieved from http://www.statisticbrain.com/Twitter-statistics/

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2003). *Language vitality and endangerment*. Report by the UNESCO Ad Hoc Expert Group on Endangered Languages. Paris, France: Author.

Ungerleider, N. (2011). *Preserving indigenous languages via Twitter*. Retrieved from http://www.fastcompany.com/1747283/preserving-indigenous-languages-twitter