Microsoft Translator Hub for Māori Language

Te Taka Keegan[1] and Jasmin Cairns[2]

[1] Computer Science Department, University of Waikato, Hamilton, New Zealand

[2] Computer Science Department, University of Waikato, Hamilton, New Zealand

Abstract

Recent improvements in Machine Translation (MT) software has opened new possibilities for applications of automatic language translation. But can these opportunities exist for the smaller, minority languages of the world? Training data was collected for the language pair of Māori and English which was used to build an MT system using Microsoft's Translator Hub software. A comparative analysis was undertaken with this system and Google Translate. Various MT metrics and analysis software was considered before deciding to use the Asiya toolkit to undertake the comparative analysis. Māori language experts were also used to provide a human perspective on the output translations. The overall results showed no significant difference in the translation quality produced by the two systems. Despite some misgivings around the accuracy of the translations these results do show promise for MT usage by minority languages.

Keywords: *Machine translation, Māori language, Minority languages.*

1. Introduction

With the recent application of neural networks in statistical Machine Translation (MT) systems, the accuracy of translations that is being output is increasing all the time [1]. While much research has been undertaken on this technology as it applies to the translation of larger languages of the world, the same amount of application and interest has not been generated to the smaller, minority languages of the world. The Māori language of New Zealand is one such example of a minority language. With the exception of Google's Translate system [2], no (other) major MT systems have been activated for this language. This paper reports on the activating a second MT system for the Māori language; Microsoft's Translation Hub.

Microsoft have created software called Microsoft Translator Hub (MTH) that lets anyone collect documents and train a translation system for free [3]. This system has the potential to be an important resource as it encourages communities to gather electronic documents in their own language and add them to a collective source. This collective source can then be used over time to improve the quality of translations that the MTH produces.

The research discussed in this paper was able to access a new source of documents, translated sentence pairs of Māori/English language texts, which had not previously been available for MT training. Consequently there was interest to determine if the new source of training data, coupled with the new MT system that Microsoft had built, could produce better translations than those currently being available through Google Translate. This paper will discuss a number of areas that are relevant to minority language use of MT systems. The metrics used to evaluate the accuracy of MT systems is discussed first, followed by the methodology used in this research. The activation of the Microsoft Translator Hub for the Māori language is explained, followed by a section on considerations that were important for this particular comparative study. A wide spectrum of results are shown providing a number of different metrics to compare the translation quality of the Microsoft Translator Hub with Google Translate. The paper concludes with a summary of the comparative analysis that has been undertaken and some implications for minority languages.

2. Evaluating Machine Translations Systems

The task of translating between two spoken languages is surprisingly difficult. An accurate translation requires deep knowledge of both languages including understandings of syntax, semantics, connotations, the polysemy of words, tense, being able to deal with ambiguity, and being able to deal with idiomatic expressions. There is no one to one translation for every word or phrase. Therefore, to process parallel translations it is not simply enough to align one word at a time, it must be done with whole sentences [4].

Just as there can be a number of different ways to say something, there can be a number of different translation outputs that are correct. This makes the task of comparing translation outputs difficult. Copyright issues, among others, make it difficult to directly compare the algorithms that Google Translate and Microsoft Translator Hub use to train their systems. Consequently comparative analysis must be undertaken on the translation outputs from these two systems.

2.1 How are MT programs compared?

"The closer the machine translation is to a professional human translation, the better it is" [5]. To compare the two systems the output translations from both systems will be rated and the ratings will then be compared. This comparison can be undertaken by using Machine Translation evaluation metrics, and by employing language specialists to evaluate the outputs.

2.2 Machine Translation Evaluation Metrics

Machine translation evaluation metrics give machine made translations a rating when compared to a reference translation that a person makes. There are several international conferences each year where machine translation evaluation metrics are submitted and tested against each other. The conferences aim to get researchers in the field of machine translation together to combine efforts to work towards a more standardised system of evaluation. Some conferences are focused on setting or reforming benchmarks and metrics, such as 'Cracking the Language Barrier' [6]. Whilst others, such as the Conference on Machine Translation [7], are focused on evaluating research and performance of Machine Translations systems. These conferences discuss a wide range of MT evaluations metrics; three of the most prominent metrics are described below.

2.2.1 BLEU – Bilingual Evaluation Understudy.

BLEU was presented in 2002 as an inexpensive way to rate translations. The output from this metric is simply a number between 0 and 1, with a rating closer to 1 being better. BLEU simply can tell you how many words between reference and candidate sentences were directly correct [5].

2.2.2 ROGUE – Recall-Oriented understudy for Gisting Evaluation.

ROUGE is a set of multiple metrics and a software package. There are 4 main metrics; ROUGE-N is based on recall of n-grams between the candidate and reference texts. ROUGE-L identifies longest co-occurring in sequence n-grams which allows it to take sentence structure into account. ROUGE-W is like ROUGE-L, but it gives more common subsequence's more weight. ROUGE-S measures the overlap of bi-grams between candidate and reference translations. All four metrics perform well in singular document evaluations [8].

2.2.3 METEOR – Metric for Evaluation of Translation with Explicit Ordering.

Meteor puts much more emphasis on recall, it is based on the harmonic mean of unigram precision and recall. Meteor allows for synonym matching which helps to lessen the problems with metrics like BLEU where words are compared strictly as being the same or not. BLEU performs best at a document or large text level, whereas METEOR has shown accurate correlation with human evaluation at the sentence level [9].

2.3 Language Specialist Evaluations

Language specialists should also be used to evaluate the output of translation software. Language specialists provide the 'human eye' and are able to discern the subtleties and historical and cultural references that are often embedded in minority languages. Language specialists, having had a longer and arguably deeper association with the language will more accurately discern ambiguities of the language, and as such are in the best position to judge language accuracy.

MT evaluation metrics have advantages in that they are quick to produce, they provide consistent ratings and there is a low cost to generating the metrics. Using humans to evaluate takes more time as each translation output has to be manually considered. This subsequently means a higher cost for evaluators' time. However, as the output from MT is a human language, humans should be involved with manually evaluating the output.

3. Methodology

Encouragement and support from Microsoft meant the Microsoft Translator Hub (MTH) was a viable alternative as a Machine Translation system for the Māori language. The online Help files and documentation showed how a system could be built without direct engagement with Microsoft MT engineers. Once the system was built it needed to be evaluated. The activity discussed in this paper involved four distinct stages.

The first stage was to collect as many files as possible that contained translated sentence pairs of Māori-English or English-Māori texts. These files were sourced from a number of different collections and institutions. The files were created from a number of different genre, including the bible, computer interface translations, reading material translations and translations of conversational texts. In total 7 separate files were used containing over 300,000 sentence pairs that were stored on 25 Gigabytes. These files needed to be converted into a format that the MTH could easily read.

The second stage involved training the MT system and adjusting various settings. Two systems needed to be trained; a Māori-English system, and an English-Māori system. Once trained and tested the MTH systems gave translated outputs that could be used for testing.

The third stage involved taking the translation inputs that were used in the Microsoft MTH system, and running them through Google's Translate system, this gave some translated output data that was used for comparison analysis.

The final stage, as reported below in the Results Section, involved comparing the translated outputs from the Google Translate and the Microsoft Translator Hub, both Māori-English and English-Māori.

4. Microsoft Translator Hub Activation

4.1 File Formats

The first step was to gather all the parallel documents into one place and one format. The documentation for the Microsoft Translator Hub states that it supports a number of file formats including DOCX, PDF, HTML, TMX, and UTF encoded text files [10]. However, to upload two documents that are translations of each other they need to be identical in structure. Having these sorts of files would be the easiest way to train a system, but this is labour intensive, as it requires manual segment alignment of translated documents. The documents gathered for this research are in the format of phrase books with both languages in one file, as shown in Figure 1.

according to what has been said	e ai ki nga kōrero

Figure 1. Phrase Book example

In phrase book creation, the lines need to be copied from the files into another format where they are aligned on the same line. This was easiest to achieve by using a spreadsheet file. The spreadsheet file allows for the different topics to be separated into tabs, the parallel lines to be right next to each other and all the material in one place. Some of the documents that were obtained needed to be pre-processed. This was as simple as writing a script in java that separated the lines at specific places into separate files that could then be copied into a spreadsheet sheet. Each file set had different line tags and format consequently pre-processing had to occur for each individual file.

4.2 Uploading Documents to the Microsoft Translator Hub

Uploading documents to the Microsoft translator Hub, a task that should have been straight forward, took a while to figure out. On the MTH documents tab, as shown in Figure 2, there is no 'upload' button. This is a usability issue with MTH that needs addressing.

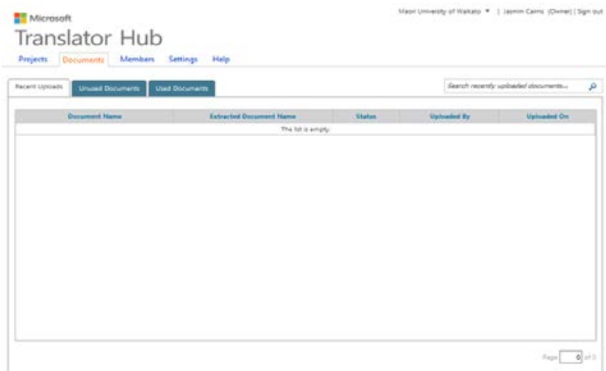

Figure 2. Microsoft Translator Hub Document Page

To get a community uploading documents to the MTH there needs to be a quick way to add them. To add documents to the system a user must go to the project tab, click "+ Train New System", and then "+ add documents". After adding documents, the user exits out of the train new system page and the documents can then be found in the documents tab.

Having a button on the documents page would let you compile the needed documents for the system by clicking "Upload", selecting the file and clicking "Submit".

4.3 Training Systems on the Microsoft Translator Hub

After uploading all the documents to the MTH, the combined total showed 604,220 parallel sentences. However, this was a double up; the files used for the Māori-English translation system were the same files used for the English-Māori translation system, with the language columns reversed.
The MTH allows the user to set out specific documents for training, tuning and testing. Since the base language data was derived from such a wide range of genre it was decided to use a random selection of this original data for training. While the MTH software offered a facility for tuning the system, time did not allow this facility to be investigated in any detail. Although there is a facility

IJCSI International Journal of Computer Science Issues, Volume 15, Issue 3, May 2018
ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784
www.IJCSI.org https://doi.org/10.5281/zenodo.1292400

within the MTH for testing, due again to time constraints, it was decided to test the translated outputs external to the MTH software.

5. MT Evaluation Considerations

Due to the nature of how the MT systems and evaluation metrics are developed, there is a high degree of fragmentation between tools and the data they output. This makes it hard to know where to look for a widely accepted evaluation. For this research, it was decided that a wide range of options would be considered after a preliminary evaluation of translation output.

5.1 Preliminary Check

The MTH provides a facility to evaluate its results before the system is deployed and made public. 'Evaluate Results' screens are produced as displayed in Figure 3 for Māori-English and Figure 4 for English-Māori. The first column contains the source text to be translated, the second column contains the reference text and the output text from the MT system. For the Māori-English system shown in Figure 3 an additional colour scheme is displayed; text in blue is a correct translation, matching the reference text is green, while the text in red is where the systems indicates low confidence in its translation.

Māori	English
Mauria he hāmarara, kai te tīmata te kōua.	**Ref:** Take an umbrella, it's starting to spit. **MT:** Take an umbrella and start the kōua.
Ki te iwi o Ngāti Porou, he maunga rangatira a Hikurangi.	**Ref:** To the people of Ngāti Porou, Hikurangi is a grand mountain. **MT:** To the people of Ngāti Porou, a captain of the Sky.
Ka murua te whenua i runga i te raupatu.	**Ref:** The land was taken by right of conquest. **MT:** The Earth will be purged on the confiscated.
I tonoa mai māku e torō te nama wini.	**Ref:** I was asked to draw the winning number. **MT:** Sent me draw a win.
E ngākau kore ana ahau ki te whai atu i tēnā ara.	**Ref:** I am disinclined to follow that course of action. **MT:** I have no heart to follow that path.

Figure 3. Microsoft Translator Hub 'Evaluate Results' Example Māori-English

The colour scheme was not available for the English-Maori Evaluate Results screen, as displayed in Figure 4.

English	Māori
They turned up several adzes when they were excavating.	**Ref:** I a rātau e karikari ana, rawaka tonu nga toki i kōhurea. **MT:** Ka tahuri rātau i maha nga toki no nga excavating ai ratou.
Would you like to continue using the new setting?	**Ref:** Aromatawai i whakaaetia e te Arotake ā-Hoaako **MT:** Kei te hiahia koe kia haere tonu te whakamahi i te tautuhinga hōu?
They were distressed when they heard the news of the death.	**Ref:** No te rongonga mo te mate, ka pōkeka rātau. **MT:** Ka manahau mātau i te whakaahuareka ki a ratou e rongo ana i nga kōrero mo te matenga.
One combatant from each side was selected to settle the dispute.	**Ref:** Kotahi anō kairiri i kōwhiria o ia taha hai whakatau i te **MT:** I kōwhiria combatant kotahi i ia taha ki te whakatikatika i te tautohetohe.
He had a plate fitted by the dentist.	**Ref:** I whakamaua e te pou niho he tako mōna. **MT:** Niho he i whakamaua e te rata niho.

Figure 4. Microsoft Translator Hub 'Evaluate Results' Example English-Māori

A preliminary check on the 'Evaluate Results' outputs highlights some general translations issues and some issues with orthography.

5.1.1 General Translation Issues

The first example in Figure 3 shows the word 'kōua' has not been translated. The first example in Figure 4 shows the word 'excavating' has not been translated. This suggests that the dictionaries generated, or used are insufficient and could be expanded to cover all candidate words. The second, third and fourth examples in Figure 3 show poorly translated phrases with an only a minimal understanding discernible from the MT text. The fifth example in Figure 3 shows a translation that has been generated literally, where the intent of the original text is not completely conveyed.

5.1.2 Orthography

The standard orthography for the Māori language uses a macron, or bar, over a vowel to indicate when a particular vowel is lengthened. This is important in terms of pronunciation and meaning. One Māori word may have several variants depending on which vowel is lengthened, with each variant have a completely separate meaning. For example; 'taua' means that particular object, 'tāua' means both you and I, and 'tauā' means a war party.

Many of the MT outputs from the English-Māori system had macrons missing when they should have been present. In the first example on Figure 4 the MT output is given as 'Ka tahuri rātau i maha nga toki no nga excavating ai ratou.' Four macron characters are missing. The output should be listed as 'Ka tahuri rātau i maha ngā toki nō ngā excavating ai rātou.' While in this case the absence of macron characters does not hinder the understanding of the meaning, examples may arise where it does.

5.2 Obtaining Outputs from Google Translate

To be able to compare the Microsoft Translator Hub with Google Translate both systems needed to translate a similar set of sentences. When building the translation systems of the MTH two test files was created using 2500 source texts of Māori and 2500 source texts of English. Each file was 280kB. Examples of these source sentences are shown in the first columns of Figure 3 and Figure 4, with the results output shown in the second columns.

Having already translated these files in the MTH the source sentences were subsequently uploaded to the document translator of Google Translate. This should have been an easy process but Google's online translator tool only translated 25kb of any given file at a time. This meant that the resulting translation file ceased translating at a certain point in the file. This point had to be found, the file separated, and then uploaded for translation again. This process was repeated 11 times per 280kb file that needed to be translated with GT. Once completed this gave a set of translated outputs that could be used for comparative analysis.

5.3 Machine Translation Evaluation Software

Three software systems were considered for evaluating and comparing the accuracy of the translated outputs from the two MT systems.

5.3.1 MTEval Toolkit

The MTEval Toolkit is written in C++ and has been updated as recently as April 7th, 2017. The developer Yusuke Oda is a researcher from the Japan National Institute of Information and Communications Technology (NICT). This toolkit includes the metrics BLEU, NIST, RIBES, and WER which can be run on the document and sentence levels. An output example of the MTEval Toolkit is shown in Figure 5.

```
$ cd /path/to/mteval
$ build/bin/mteval-corpus -e BLEU RIBES -r data/ref -h data/hyp1
BLEU=0.666113   RIBES=0.969124

$ build/bin/mteval-sentence -e BLEU RIBES -r data/ref -h data/hyp1
BLEU=1.000000   RIBES=1.000000
BLEU=0.759836   RIBES=0.955443
BLEU=0.000000   RIBES=0.975310
BLEU=0.000000   RIBES=0.945742
```

Figure 5. MTEval Toolkit Output Example

5.3.2 MultEval

MultEval is written in Java and was last updated in 2013. The developer Jonathan Clark is a member of the Microsoft Research Translator Team, all code seen is from his graduate work for the Pennsylvanian Carnegie Mellon University. This project includes the metrics BLEU, METEOR, TER, and Length which can be run at the document level. An output example of MultEval is shown in Figure 6.

Metric	System	Avg	$\bar{s}_{sel}$	s_{Test}	p-value
BLEU ↑	baseline	18.5	0.3	0.1	-
	system 1	18.8	0.3	0.3	0.00
	system 2	18.5	0.3	0.1	0.00
METEOR ↑	baseline	29.3	0.1	0.0	-
	system 1	30.3	0.1	0.1	0.00
	system 2	29.3	0.1	0.0	0.00
TER ↓	baseline	65.7	0.4	0.2	-
	system 1	64.8	0.4	0.6	0.00
	system 2	65.7	0.4	0.2	0.00
Length	baseline	107.5	0.4	0.1	-
	system 1	107.7	0.3	0.7	0.09
	system 2	107.5	0.4	0.1	0.00

Figure 6. MultEval Output Example

5.3.3 Asiya Toolkit

The Asiya Toolkit is a website that was made in 2010 and was last updated in 2014. Asiya was developed in at the TALP Research Centre NLP group at Barcelona Tech. This toolkit includes the metrics BLEU, NIST, METEOR, ROUGE, GTM, O, WER, TER, and PER. A run of the software gives results at the sentence and document levels with optional graphing and further searches such as highest scoring translations. An output example of Asiya Toolkit is shown in Figure 7.

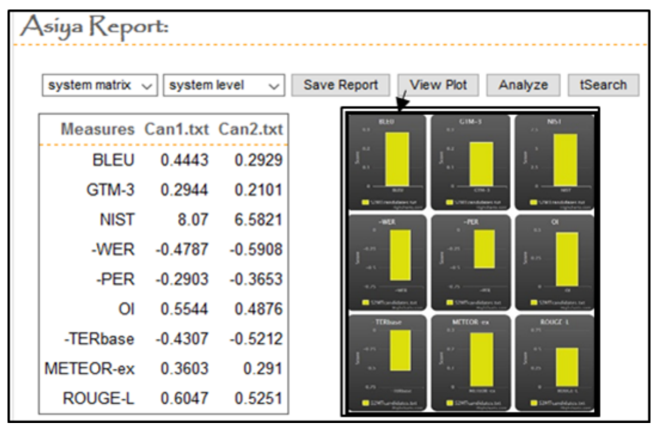

Figure 7. Asiya Toolkit Output Example

To verify that the metrics were giving a similar score between different evaluation systems, the BLEU results were compared. The scores were all within 0.05 of each other indicating consistencies across the systems. MTEval Toolkit and MultEval are both based on Linux installs and command line controls. Whereas, Asiya is an online tool that does not require installation and has MOOCs available to aid in the use of the tool. It also includes 9 metrics whilst the others only include 4. Asiya allows the user to upload multiple reference and candidate translation files. The output can be viewed at the sentence and document level. Consequently the Asiya Toolkit was used to compare the output translation from the MTH and GT.

The Asiya Toolkit produced a series of graphs but also some raw output. This raw output was preferred as it allowed us to build our own graphs and have a better control on what was displayed as can be shown in Figure 8 and Figure 9.

6. Results

6.1 MT Evaluation Metrics

Raw data from the Asiya Toolkit was used to produce the graph displayed in Figure 8 and the graph displayed in Figure 9. Figure 8 is a graph of nine metrics that compare translations from Google Translate with translations from the Microsoft Translator Hub. Both MT systems translated the same 2500 sentences from English to Māori.

Six of the metrics, BLEU, GTM, IO, METEOR-ex, ROUGE and NIST have the GT system performing on average 0.05 better. But the error rate based metrics WER, PER and TERbase show that GT has a higher error rate.

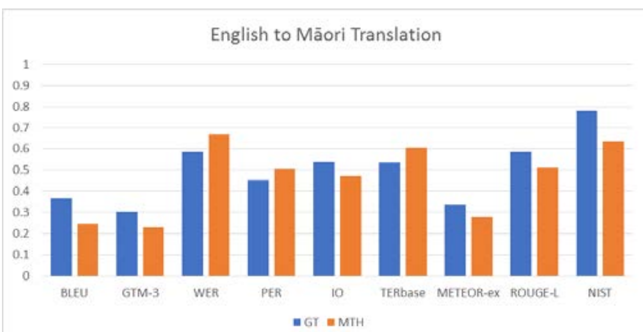

Figure 8. Comparing English-Māori Translation Metrics

Figure 9 is similar to Figure 8 in that it is also a graph of nine metrics that compare translations from Google Translate with translations from the Microsoft Translator Hub. Both MT systems translated the same 2500 sentences but in this case the translations were from Māori to English.

In a similar manner to the previous results, six of the metrics, BLEU, GTM, IO, METEOR-ex, ROUGE and NIST have the GT system performing better than the MTH system but this time the difference on average is only 0.01. Again, the error rate based metrics WER, PER and TERbase show that GT has a higher error rate but again the differential is much less.

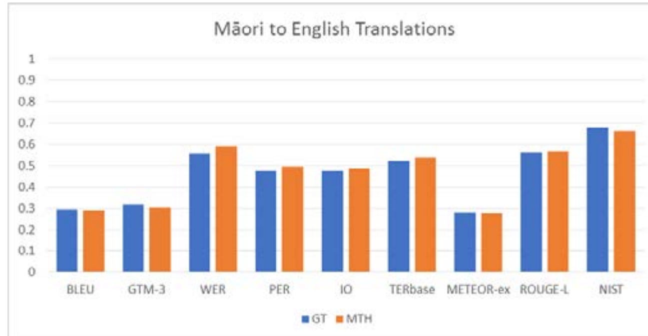

Figure 9. Comparing Māori-English Translation Metrics

6.2 Language Specialist Evaluations

Two Māori language specialists were employed to evaluate the translated output of the MT systems. Each language specialist was given 1000 translated sentences to evaluate. This consisted of 250 English sentences that were translated by GT and MTH, and 250 Māori sentences that were translated by GT and MTH. The source data was selected at random from the 2500 sentence test source, with sentences ranging in length of 50 characters to 320 characters. The evaluators were given a different set of translations to evaluate. The translated sets were labelled A and B rather than GT and MTH to ensure that the no prior personal biases were introduced by the evaluators.

The evaluators were asked to grade the translations based on two separate ratings; a preference rating and an accuracy rating. The preference rating was a simple choice, which was the better translation; A or B, or were they both rated the same? The accuracy rating was based on a five point scale; 1 if the translation was nonsense, 2 if some words were correct, 3 if the main context was correct, 4 if it was a good translation but with minor errors and 5 if it was an accurate translation.

The results of the language specialist evaluations are displayed in the next four figures.

6.2.1 Evaluator 1 Preference Rating

The preference feedback from evaluator 1 is shown in Figure 10. The graph shows a distinct preference for the translations from GT for English-Māori at 53% (265/500). However for Māori-English there was no significant preference for either GT; 38.6% (193/500) compared with MTH 36.0% (180/500).

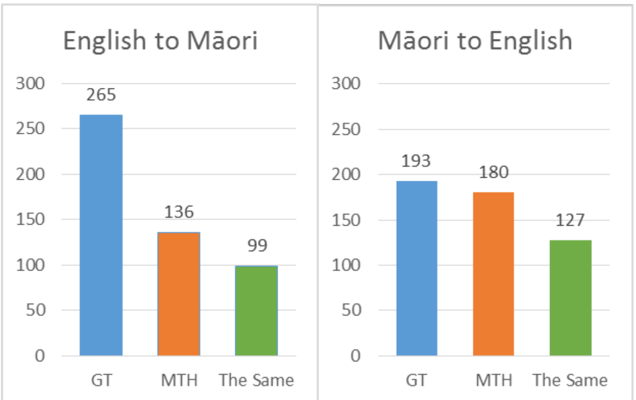

Figure 10. Evaluator 1 Preference Ratings

6.2.2 Evaluator 1 Accuracy Rating

The accuracy feedback from Evaluator 1, as displayed in Figure 11, showed a similar result to the preference rating. For the English-Māori translations, GT has a higher number of translations rated 4 and 5, whilst MTH translations show higher numbers at ratings 2 and 3. For Māori to English, Evaluator 1 indicated both systems have reasonable similar numbers across the different ratings.

Figure 11. Evaluator 1 Accuracy Ratings

6.2.3 Evaluator 2 Preference Rating

The preference feedback from evaluator 2 is shown in Figure 12. In contrast to the previous evaluator this evaluator has a distinct preferences for MTH translations over GT translations. For the English-Māori system the preference was 37.2% (186/500) compared to 28.6% (143/500). For the Māori-English system the preference was much higher; 36.4% (182/500) compared to 21.2% (106/500). The graph shows a distinct preference for the translations from GT for English-Māori at 53% (265/500). Another difference highlighted in these graphs is that Evaluator 2 was more likely to rate the translations the same than Evaluator 1.

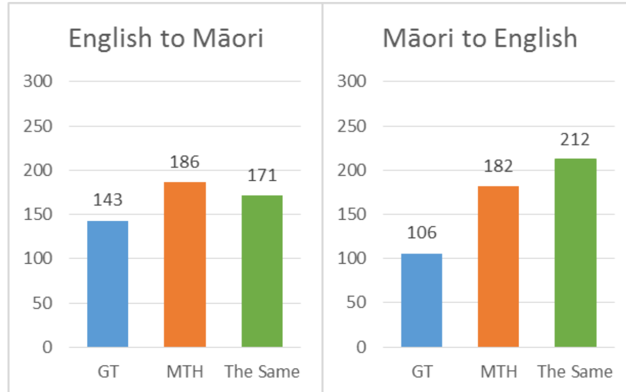

Figure 12. Evaluator 2 Preference Ratings

6.2.4 Evaluator 2 Accuracy Rating

The accuracy feedback from Evaluator 2, as displayed in Figure 13, showed a similar result to Evaluator 2's preference ratings. There are preferences for MTH translations. For the English-Māori translations, there are similar numbers around rating 3 but the higher ratings of 4 and 5 are slightly favoured towards MTH. For the Māori-English systems the MTH has significantly higher numbers in ratings 4, 142 (28.4%) compared with 93 (18.6%) and higher numbers in rating 5, 120 (24.0%) compared with 109 (21.8%).

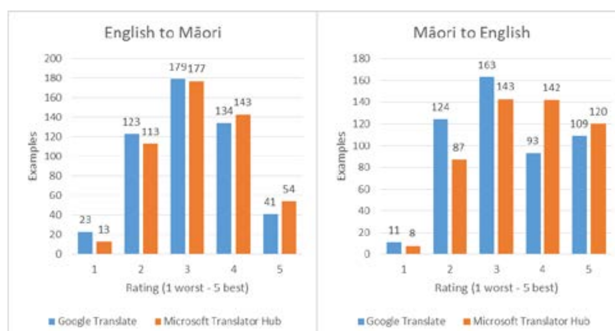

Figure 12. Evaluator 2 Accuracy Ratings

7. Conclusions

This research set out to determine if a new set of language data could be used with the Microsoft Translator Hub to build a comparable Machine Translation system for the language pair of Māori and English. The Microsoft Translator Hub was used to build the MT system and Google Translate was used for a comparative analysis. The Asiya Toolkit was used to compare 9 MT evaluation metrics. As a further comparison two language experts were asked to evaluate the translation output from the two systems.

7.1 Comparison Summary

The results show that there were no significant differences between the two systems.

The MT evaluation metrics generated by the Asiya Toolkit showed the GT system performing marginally (0.05) better in six of the nine metrics when the systems were translating from English-Māori. There was an insignificant difference (0.01) when the systems were translating from Māori to English.

Evaluations from the language experts were contradictory. Results from the first evaluator suggested that the GT system produced translations that are more accurate; results from the second evaluator suggested the MTH system produced translations that are more accurate. There could be a number of factors leading to these conflicting results. The sets that the evaluators examined were not the same translations; this may have skewed the results. The evaluators may not have been rating with the same severity. Further, the evaluators may have been more have been more familiar with certain genre of translations and this may have effected their ratings.

A number of approaches could be taken to improve consistency by language expert evaluators. The evaluators could be asked to rate the same translations. A larger number of translations could be rated. A larger number of evaluators could be used. These last two solutions would of course require additional resources.

7.2 Implications for Minority Languages

The Microsoft Translation Hub software can build MT systems for minority languages. The online software can be used to publish a trained MT system that can be utilised in apps or web pages. The online documentation and support is sufficient to build a system without direct involvement from Microsoft. There is no need for specialized hardware or any additional software.

Clearly, the accuracy of the translations produced depend on the accuracy and the amount of translated data that is made available to train the MT system. With 300,000 translated pairs the MTH system was able to match the GT system for the language pair of Māori and English. However, there is an argument that we were not comparing apples with apples. It was not possible to determine exactly what data the GT system used when its MT system was trained.
It is also important to note that both of these systems produced translations that had an average rating of only 0.3 on the BLUE evaluation metric, and produced highly accurate translations only 15.6% of the time. This is not an accuracy level that can be used with any confidence. This issue was made clear to a New Zealand Hamilton mayoral candidate, James Casson, in September 2016 [11]. Wanting to appeal to the Māori community, he had his profile translated into Māori using Google Translate, and then without having this checked by a proficient Māori language speaker, sent it out to everyone household in Hamilton. The translation had many errors and served more to offend rather than appeal to the Māori community.

7.3 Suggested Avenues for Improvements of Minority Language Translations

As has been stated above, the accuracy of the translations produced depend on the accuracy and the amount of translated data that is made available to train the MT system.

One avenue is to improve the quality of the training data that is used to build the MT system. Steps should be taken to pre-process this training data. Section 5.1.2 highlighted errors observed in the orthography of the translated outputs in Māori. For these errors to appear there must have been similar errors in the training data. A facility, called the Māori Macron Restoration Service, is available to check the orthography of Māori language texts [12]. This facility should be used to pre-process Māori language training data.

A suggested avenue to increase the amount of translated data is to seek community involvement. If a number of community language supporters and community language translators are collectively working together to gather high quality language data in the form of translated pairs, and if this data is regularly used to retrain the translation system, then over time the translation accuracy will rise to an acceptable level.

Acknowledgments

With much gratitude to Te Pokapū WhitiReo Charitable Trust and Auckland University of Technology (AUT) who provided tremendous assistance in sourcing translated pairs of Māori-English and English-Māori texts. We are also indebted to Microsoft NZ for their continued support of the Māori language.

References

[1] Cambria, E., Hazarika, D., Poria, S., & Young, T. Recent Trends in Deep Learning Based Natural Language Processing. CoRR, abs/1708.02709. 2017.

[2] Keegan, T., Evas, J. Can computer assisted translation tools improve the productivity and quality of minority languages translations? He Pukenga Korero: a Journal of Māori Studies 10 (2), 2011, 34-41.

[3] Microsoft Translator. New Microsoft Translator Customization Features Help Unleash the Power of Artificial Intelligence for Everyone. 2016. Retrieved from https://blogs.msdn.microsoft.com/translation/2016/01/27 /new-microsoft-translator-customization-features-help-unleash-the-power-of-artificial-intelligence-for-everyone/

[4] Mohaghegh, M., McCauley, M., & Mohammadi, M. Maori-English Machine Translation. NZCSRSC New Zealand Computer Science Research Student Conference-Canterbury University. Unitec Research Bank. 2014.

[5] Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. BLEU: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting on association for computational linguistics 311-318. Association for Computational Linguistics. 2002.

[6] Cracking the Language Barrier. LREC 2016 Workshop. 2016. Retrieved from http://www.cracking-the-language-barrier.eu/mt-eval-workshop-2016/

[7] EMNLP 2017 Second Conference on Machine Translation (WMT17). 2017. Retrieved from http://www.statmt.org/wmt17/

[8] Lin, C. Y. Rouge: A package for automatic evaluation of summaries. Text summarization branches out: Proceedings of the ACL-04 workshop. 2004.

[9] Banerjee, S., & Lavie, A.. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgements. The ACL workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization. Vol 29, 2005, pp. 62-72.

[10] Microsoft Corporation. Microsoft Translator Hub User Guide. 2016. Retrieved from hub.microsofttranslator.com /Help/Download/Microsoft%20Translator%20Hub%20User %20Guide.pdf

[11] Smallman, E. R. Hamilton candidate's Google translation misguided gibberish, Maori likely to switch off. 2016. Retrieved from Stuff: http://www.stuff.co.nz/national/politics /84398001/hamilton-candidates-google-translation-misguided-gibberish-maori-likely-to-switch-off

[12] University of Waikato. The Māori Macron Restoration Service. 2011. Retrieved from http://community.nzdl.org/ macron-restoration/jsp/servlet/DirectInput.

Te Taka Keegan is a senior lecturer with the Computer Science Department at the University of Waikato. He has a DipCompEng from CIT and MA, PhD from Waikato. He has been involved in a number of research activities and projects that involve the Māori language and technology.

Jasmin Cairns qualified with a BCMS in 2016 and BSc and GradCert in 2017. She currently works in IT support at Xcrania..