Building a corpus of Maori loan words

By Sally Harper

Research purpose

This project investigates the use of Maori loan words - words from te reo Maori that are used by NZ English speakers - in New Zealand's media. We are interested in exploring the frequency of various loan words, whether the inclusion of certain loans will 'attract' other loan words or a higher frequency thereof, and how these loans are utilised in the context of newspaper articles, a medium which is intended for an audience and thus designs its content around its audience. We hope to produce visual displays of the relationships between loan words, and gain greater insight into how speakers of NZ English use and understand Maori loan words.

Methodology

We chose to focus our research around articles which featured the word 'Matariki', this being the name of the Maori New Year – something culturally salient and important to Maori. Using the New Zealand Herald website for results from all NZME publications, we searched for all occurrences of the term 'Matariki' in articles over a ten year time span, dating from 2016 to 2007. We copied these articles into text files which now make up the Matariki Corpus. From this corpus we coded a random selection of 30 articles to determine things such as how the time of year can affect the number of articles utilizing loans (Matariki occurs at approximately the time of the winter solstice, around June 21, and so it is written about more often at this time), what loans occur the most frequently, and whether variables such as author or region affect which loans occur or how frequently they occur.

We intend to run our coding through the statistical modelling tool ‘R’, and produce statistical models to qualify and predict how loan occurrences are governed by the greater discourse structure present in newspaper journalism.

Outcome

This is the first part of a much larger project looking at the use of Maori loan words across a greater number of domains and greater variety of New Zealanders. The analysis of our results is ongoing.

I would like to thank my supervisor Andreea Calude for her guidance and support, our collaborators Steven Miller and Hemi Whaanga for their contributions and input, and also the University of Waikato's Summer Research Scholarship for this opportunity.