He Tātai Whenua: An Exploratory Analysis of Landscape Terms in the New Zealand Māori Language
Kristin Stock*, Hone Waengarangi Morris, Margaret Forster, Rex Paraku and Ekaterina Egorova
Massey University, Auckland, New Zealand
*k.stock@massey.ac.nz

Introduction
Current GIS do not account for indigenous conceptualisations and worldviews.
In New Zealand, large collections of written text are available in the Māori language.
We explore the use of NLP and corpus linguistics methods to:
• identify landscape senses of ten landscape terms and
• study the semantics of those terms to aid understanding towards the incorporation of indigenous conceptualisations in GIS.

The Māori Worldview
Mana Atua Place names and geographic feature types associated with spiritual elements of the environment. This dimension provides the source of mana “authority” and mauri “life force”.
Mana Whenua Place names and geographic feature types that describe the landscape from a Māori worldview.
Mana Tangata Place names and geographic feature types associated with ancestors and ancestral events that establish tribal authority and identity over specific geographical spaces and natural resources.

The Corpus
Māori language newspapers from the NZ Digital Library:
• Te Ao Hou: 1952 to 1957
• Te Puke ki Hikurangi: 1887-1898

All instances of the 10 landscape terms were manually classified into:

| Sense Code | Place name P 0.2% | General landscape feature (e.g. rivers are sacred) 12.7% | A specific instance of a feature (e.g. this is the river I grew up near) 6.5% | A person’s name connected to the feature N 0.2% | A related landscape term (e.g. weipuke/to flood, literally means water hill) R 3.3% | Non-geographic senses 77.2% |

Analysing Co-occurrence
Selected only the 5 terms used more than 100 times in a geographic sense.
Extracted top 10 (ranked by Mutual Information):
• immediately preceding and succeeding words (1R and 1L) and
• in a 3L-3R window
• words containing the landscape term (compound words are common in Māori).
Manually classified using the Māori worldview.

Classifying Landscape Senses of the Terms
Manually classified each mention of the ten landscape terms in the corpus. Used SVM Bag of Words to automatically classify.

Observations
Place names that included the ten landscape terms were relatively rare in our corpus. Variation in frequency of interactive verbs:
• More frequent and more intimate for kāinga: verbs of occupation, affection, care and connection.
• With motu and awa, more related to physical activities and management.
• With whiti, across the board, reflecting more general meaning.
Possessives relatively rare (surprising given Māori notion of belonging to land).
Body parts rare, c.f. mōrearea “ancient chants” and pārikau “narratives” of whakapapa “genealogy”.

Future Work
Larger corpus (including other kinds of documents).
More sophisticated analyses (e.g. topic modelling) to better document Māori conceptions of land.