# Non-Māori speaking New Zealanders show surprisingly sophisticated Māori phonotactic knowledge

**Yoon Mi OH, Clay BECKNER, Jen HAY & Jeanette KING**

New Zealand Institute of Language, Brain and Behaviour, University of Canterbury

## ABSTRACT

❍ **Background:** Language speakers can rate the gradient well-formedness of nonwords in their language. Such phonotactic knowledge is assumed to have been acquired from statistical learning over speakers' lexicons [1-3]. Most New Zealanders (NZers) are exposed to Māori in their daily lives but do not speak it.

❍ **Objective:** What phonotactic knowledge do NZ-based non-Māori-speakers (NMS) have?

❍ **Results:** Non-speakers of a language can develop sophisticated phonotactic knowledge through ambient exposure. Phonotactics need not arise as a generalization over a large explicit lexicon. Non-fluent speakers seem to possess implicit knowledge of the statistical properties of the lexicon.

## GOALS OF THE STUDY

⇨ Collect NZers' well-formedness ratings of Māori-like nonwords to assess their phonotactic knowledge of Māori.

⇨ Compare the phonotactic knowledge that NZ-based non-Māori-speakers (NMS) and Māori-speakers (MS) have.

⇨ Explore possible explanations for NMS' surprisingly sophisticated phonotactic knowledge of Māori.

## MATERIALS

❍ **Stimulus preparation**

⇨ 1760 Māori-like nonwords generated from a trigram model using a pseudoword generator [4]

⇨ The trigram model trained by means of a Māori dictionary [5] and two Māori running speech data (RS) [6-7]

⇨ A list of stimuli consisting of 240-320 nonwords with the same phoneme length ranging from 3 to 8

❍ **Participants**

⇨ 41 Māori-speakers (MS): 5-9 participants per length

⇨ 137 non-Māori speakers (NMS): 20-25 participants per length

❍ **Measures of linguistic knowledge**

⇨ **Phonotactic score:** sum of log transitional trigram probabilities normalized by length using a language model (LM) based on:

- a Māori dictionary [5]
- segmented Māori RS [6-7]
- a list of 963 Māori words in NZ English [8]

⇨ **Word shape score:** sum of log transitional trigram probabilities normalized by length using a LM obtained by identifying each segment as a consonant, vowel, or long vowel, and calculating probabilities over sequences of those categories.

⇨ **Presence of macron** in nonwords: used as long vowel markers in written Māori.

## EXPERIMENTS

① **Online Māori well-formedness rating task (for MS)**

⇨ 2 parts: rating task & post-questionnaire

⇨ In a rating task in Māori, participants rate Māori-like nonwords for how good they would be as Māori using a 1-5 scale.

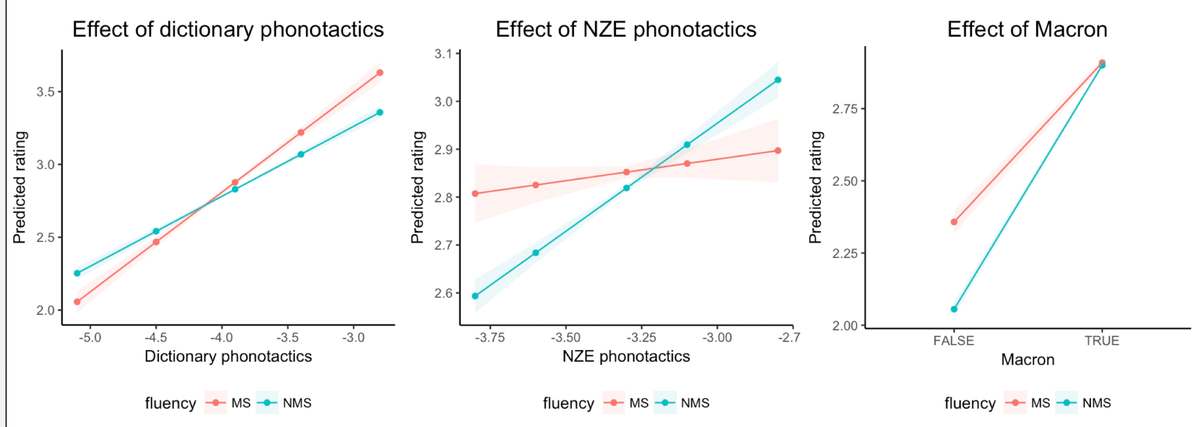

⇨ After the rating task, participants complete a post-questionnaire containing 19 questions in English regarding their sociolinguistic profile.

② **Online Māori well-formedness rating task (for NMS)**

⇨ The same instruction in English and set of stimuli as in the Exp for MS


## RESULTS

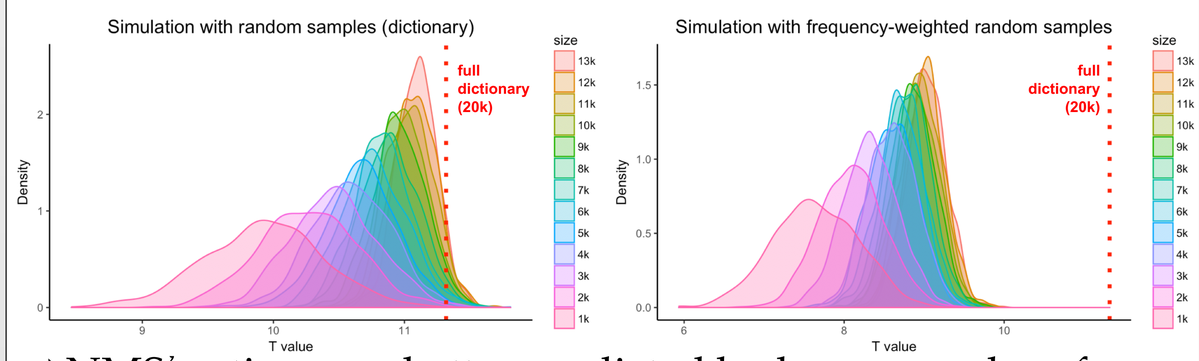
❍ **Modeling ratings with a linear mixed effects model**




⇨ When rating nonwords, Māori speakers (MS) are more sensitive to dictionary-based phonotactics than non-Māori speakers (NMS).

⇨ NMS are more influenced by phonotactics based on the list of Māori words in NZ English (i.e. NZE phonotactics) and macron of nonwords than MS.

❍ **Monte Carlo simulation with varying sample size to predict NMS' ratings of nonwords**




⇨NMS' ratings are better predicted by large samples of a Māori dictionary [5] (on the left). Phonotactics derived from frequency-weighted random samples perform even worse (on the right).

## DISCUSSION

❍ MS are more sensitive to the overall statistical patterns while NMS are more driven by the phonotactics of a set of Māori words in NZ English and the presence of macron.

❍ NMS' Māori phonotactic knowledge is best explained if we assume they have a large implicit lexicon acquired by ambient exposure.

❍ To study NMS' knowledge of Māori, we further explore their implicit lexicon by testing whether they can identify infrequent Māori words as well as frequent Māori words.

## REFERENCES

[1] Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: effects of segment probability & length on the processing of nonwords. *Journal of Memory & Language*, 42(4), 481–496. https://doi.org/10.1006/jmla.1999.2692

[2] Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics & neighborhood activation in spoken word recognition. *Journal of Memory & Language*, 40(3), 374–408.

[3] Adriaans, F., & Kager, R. (2010). Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory & Language*, 62(3), 311-331.

[4] Needle, J., Pierrehumbert, J. & Hay, J. (2014). Phonotactic probability & wordlikeness: a flexible pseudoword generator with triphones, Madison, WI, USA: *19th Mid-Continental Phonetics & Phonology Conference*, 12-14 Sep 2014.

[5] Moorfield, J. C., (2005). *Te aka: Māori-English, English-Māori dictionary & index,* Auckland, N.Z: Pearson Longman.

[6] King, J., Maclagan, M., Harlow, R., Keegan, P. & Watson, C. (2011). The MAONZE project: changing uses of an indigenous language database. *Corpus Linguistics & Linguistic Theory, 7*(1), 37-57.

[7] Boyce, M. (2006). *A corpus of modern spoken Māori.* PhD Thesis, VUW.

[8] Macalister, John (2005). *A Dictionary of Maori Words in New Zealand English*. Melbourne: Oxford University Press.