Te Kāhui Roro Reo • New Zealand Institute of
Language Brain & Behaviour

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Phonotactic sensitivity in speakers with a very small lexicon

Yoon Mi OH, Clay BECKNER, Jen HAY, Jeanette KING
NZILBB, University of Canterbury

MARSDEN FUND
TE PŪTEA RANGAHAU A MARSDEN

LabPhon16
June 20, 2018

Background

Language situation in New Zealand

- Māori is the language of indigenous people in New Zealand.
- The size of Māori vocabulary in non-Māori-speaking New Zealanders has been increasing over time, from 40-50 in (Gordon & Deverson, 1998) to 80-90 (Macalister, 2004, 2006) according to surveys.
- Most non-Māori-speaking New Zealanders have some exposure to Māori but do not understand it.

Questions of larger project

- Is it possible to acquire sophisticated phonotactic knowledge through ambient exposure (i.e. **pre-lexical phonotactic knowledge**)?

- If so, how much do non-Māori-speaking New Zealanders (**NMS**) know about the sound patterns of Māori and its lexicon?

Objective of this study

- Before understanding NMS' pre-lexical knowledge of Māori, we first study their vocabulary size and lexical knowledge through two experiments.
- **Exp1** (**Māori word identification task**): to examine what words NMS can recognize and identify as Māori words.
- **Exp2** (**Māori word definition task**): to assess NMS' knowledge of the meanings of Māori words.
- In both experiments, we also aim to study the influence of phonotactic cues on their ability to identify and associate meanings to Māori words.

Materials - stimulus

- **Exp1** (**Māori word identification task**): The list of stimuli consists of 132 Māori words and 209 highly Māori-like nonwords generated from a trigram language model using Pseudomatic (Needle et al., 2014).

- **Exp2** (**Māori word definition task**): The list of stimuli is comprised of the 118 Māori words whose average rating is above 4 (scale ranging from 1 to 5) in Exp1.

Materials - phonotactic score

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Phonotactic score of stimulus is calculated by using:

- words obtained from a Māori dictionary (20k)(Moorfield, 2005).
- a trigram language model. For example, a Māori word, aroha (love), would be split into "#ar, aro, roh, oha, ha#" to calculate its transitional probability.

Phonotactic score refers to the sum of log transitional probabilities normalized by the length of stimulus.

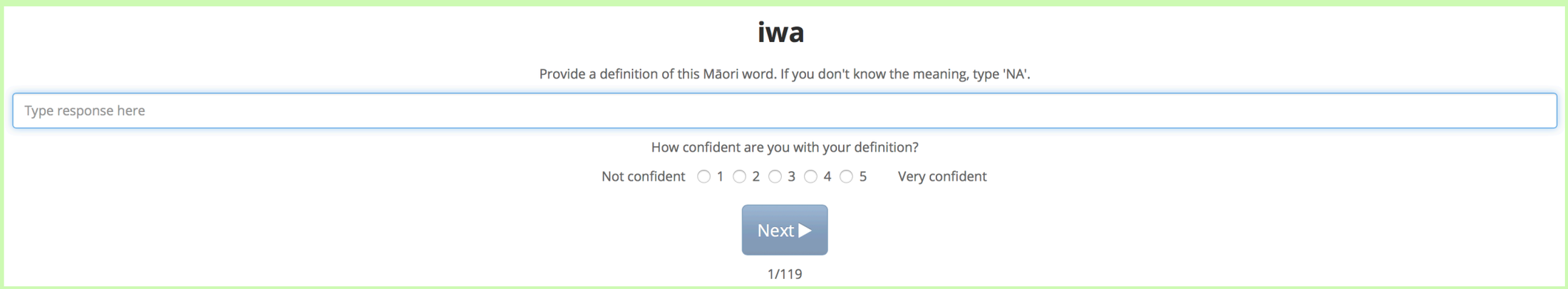
Experiment procedure

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

- **Exp1** (**Māori word identification task**): 102 adult participants rated how confident they are that it is an actual Māori word or phrase, using 1-5 scale. They also complete a post-questionnaire regarding their sociolinguistic profile.

Experiment procedure

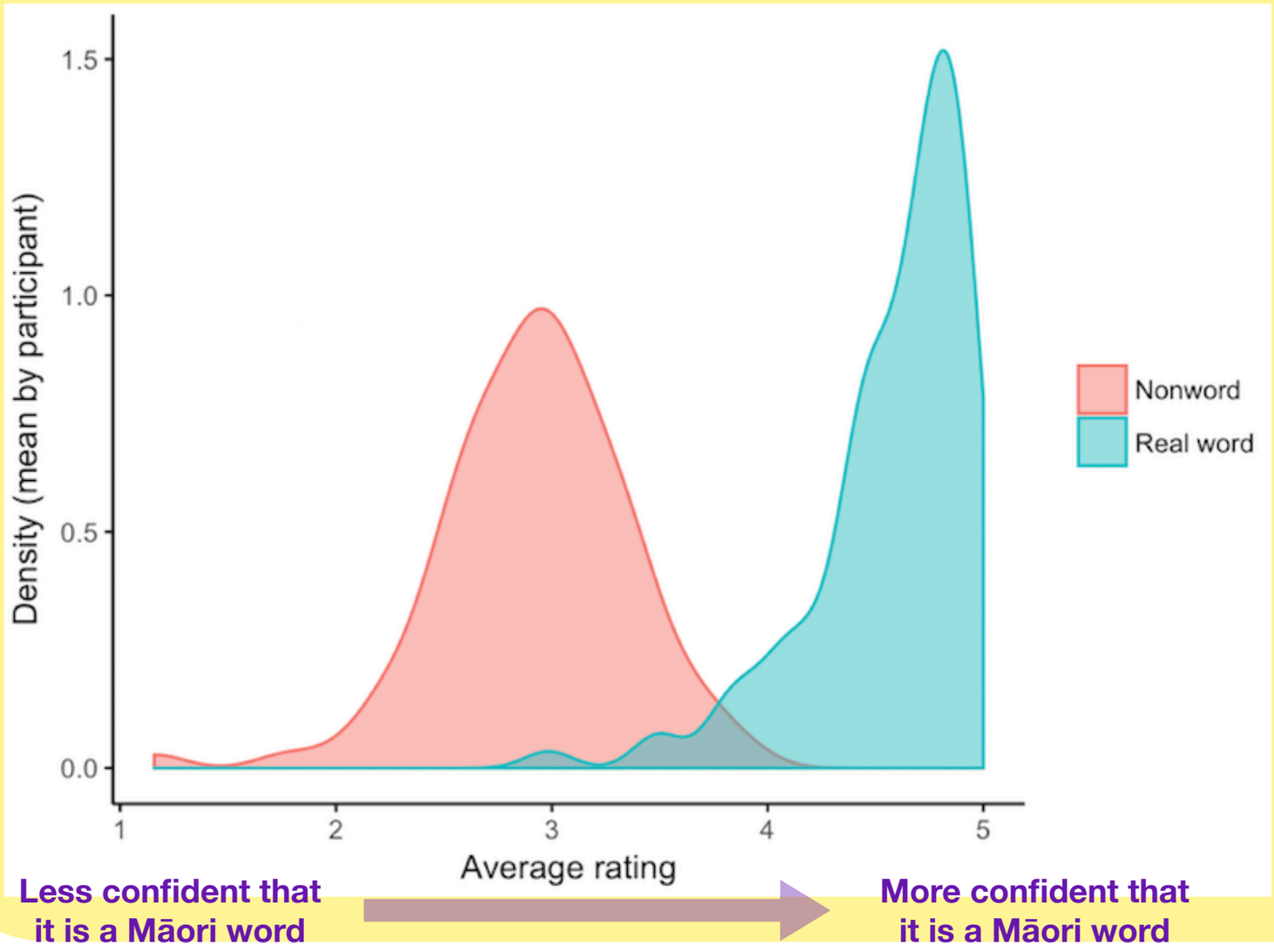
- **Exp2** (**Māori word definition task**): 124 adult participants type a definition of a Māori word they see in the screen and type 'NA' if they don't know its meaning. To obtain the best answer, we ask them to rate how confident they are with their definition.



Results - Exp1

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Density plot of average rating for categories by participant

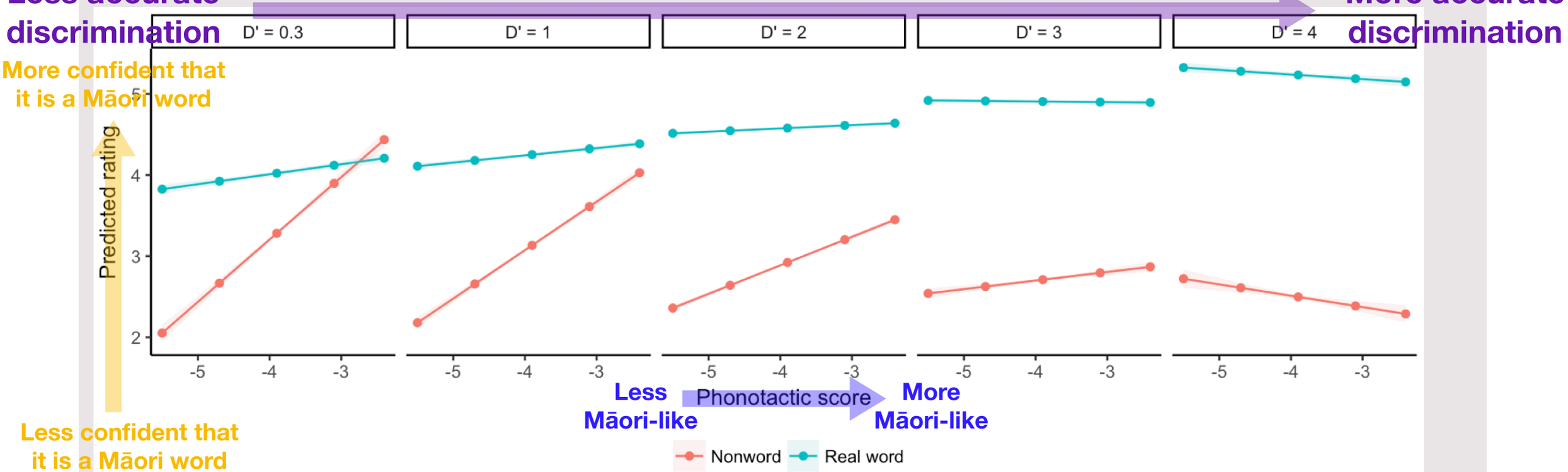


- Participants can discriminate between real Māori words and highly Māori-like nonwords.
- Among 132 Māori words, 11 words are scored below 4 on average.

Results - Exp1

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Interaction between phonotactic score, the distinction between non vs. real word stimuli, and d' (=sensitivity)

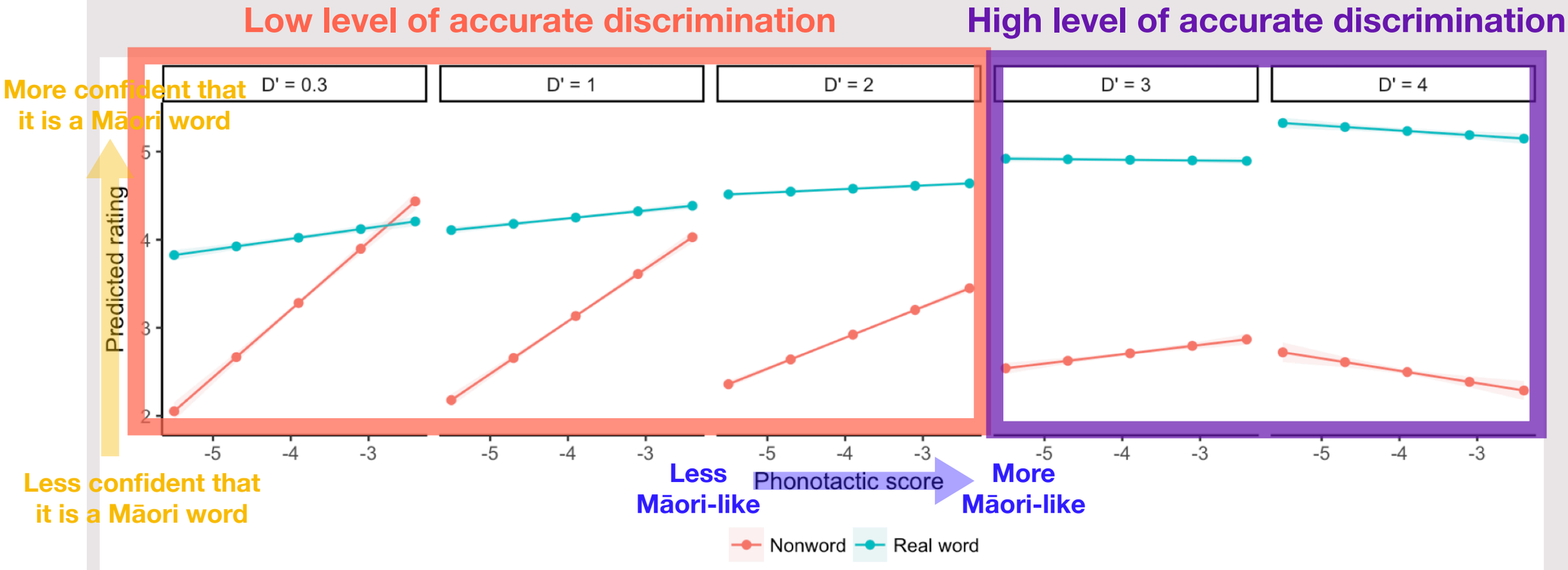


- When distinguishing between real words and nonwords, participants with lower d' are more dependent on phonotactic patterns than those with higher d'.

Results - Exp1

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Interaction between phonotactic score, the distinction between non vs. real word stimuli, and d' (=sensitivity)

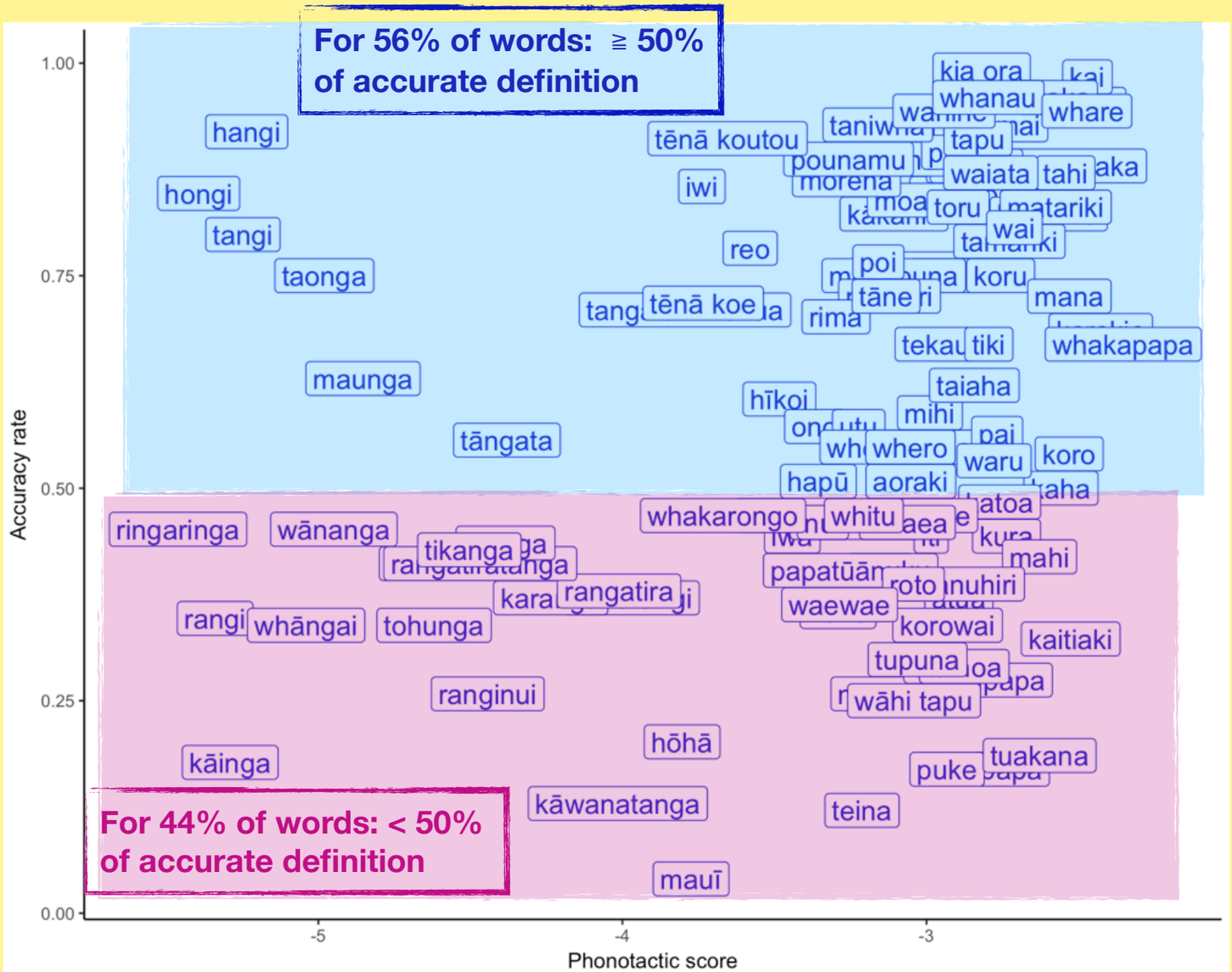


- When distinguishing between real words and nonwords, participants with lower d' are more dependent on phonotactic patterns than those with higher d'.

Results - Exp2

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Accuracy rate per word (y axis) & phonotactic score (x axis)

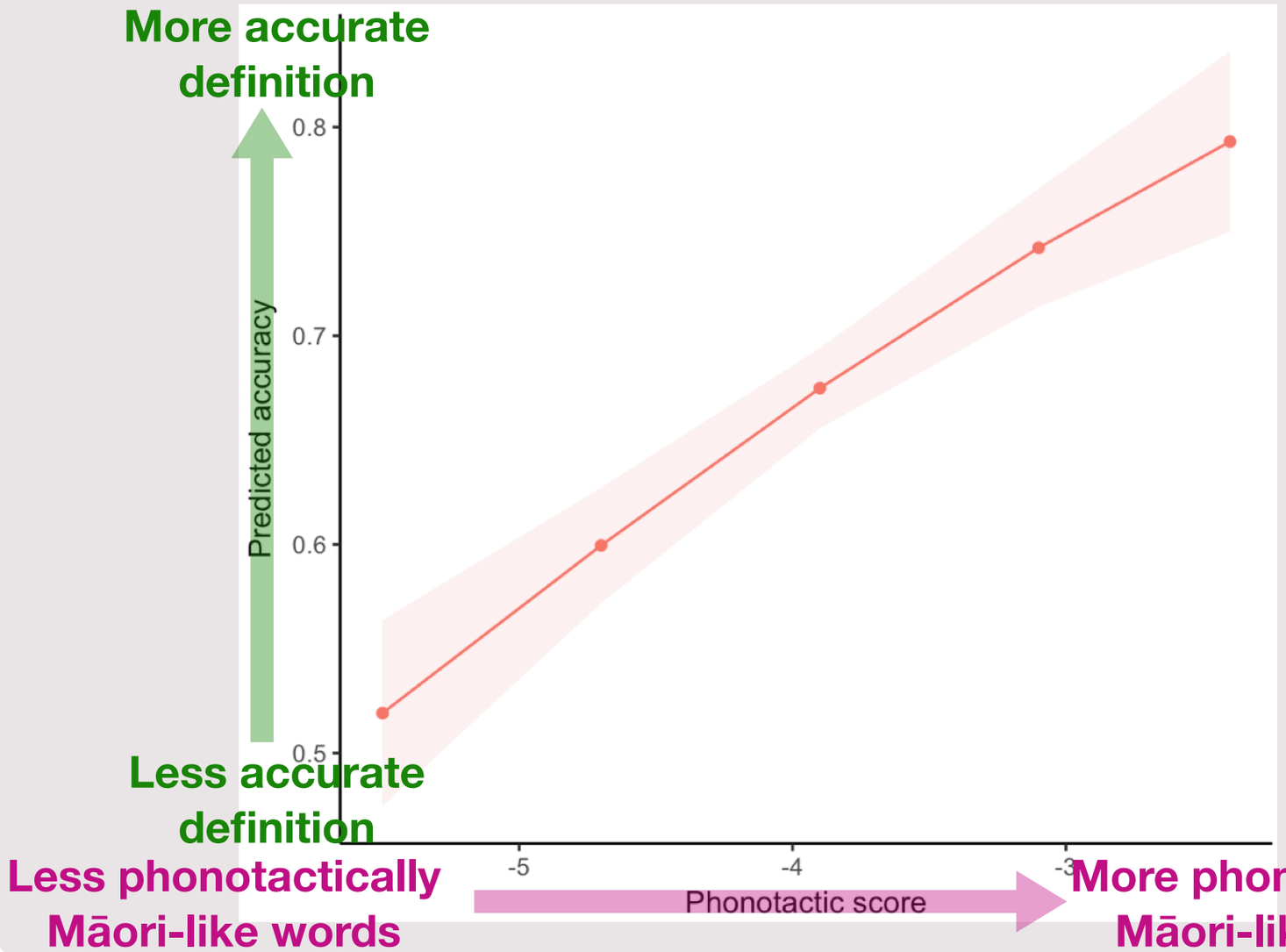


- Participants provided correct definitions with more than 50% chance for 69 words among 118 (55.6%).
- They were not able to provide correct definition for a significant subset of words identified in Exp1.

Results - Exp2

Te Kāhui Roro Reo • New Zealand Institute of Language Brain & Behaviour

Effect of phonotactic score on accuracy of definition



- Phonotactic probability also has a significant effect on participants' ability to define Māori words.
- More Māori-like words in their phonotactics tend to elicit more accurate definitions.

Conclusion

Te Kāhui Roro Reo • New Zealand Institute of
Language Brain & Behaviour

- The results of this study provide evidence that NMS have small vocabulary of Māori that they can both identify and define.
- Despite this, they are able to access statistical phonotactic information about Māori.
- Phonotactics affects which words can be identified and accurately defined.
- Phonotactics affects patterns of borrowings, and perhaps also the earliest stages of the second language acquisition.

Conclusion

- Paradox: If the active vocabulary is so small, where does the phonotactic knowledge apparently used to acquire this vocabulary come from?
- Our ongoing experimental work seems to point to a very large incipient lexicon that has been acquired by ambient exposure.
- This incipient lexicon leads to sophisticated phonotactic knowledge far beyond that which could be generalized from this small vocabulary.

Come and see our poster at the satellite event on June 23rd!

Ngā mihi nui ki a koutou!
Thank you! Any questions?

References

Gordon, Elizabeth, and Tony Deverson. 1998. *New Zealand English and English in New Zealand*. New House Publishers.

Macalister, John. 2004. A survey of Māori word knowledge. *English in Aotearoa 52*. 69–73.

Macalister, John. 2006. The Māori presence in the New Zealand English lexicon, 1850–2000: Evidence from a corpus-based study. *English World-Wide 27.* 1–24.

Needle, Jeremy; Janet Pierrehumbert; and Jen Hay. 2014. Phonotactic probability and wordlikeness: a flexible pseudoword generator with triphones. *19th mid-continental phonetics & phonology conference*.

Moorfield, John Cornelius. 2005. *Te aka: Māori-English, English-Māori dictionary and index*. Longman.