

# Using a large annotated historical corpus to study word-specific effects in sound change

**Jennifer Hay**

School of Languages, Cultures and Linguistics  
University of Canterbury  
jen.hay@canterbury.ac.nz

## **Abstract**

The Origins of New Zealand English Corpora (ONZE) at the University of Canterbury contain recordings spanning 150 years of New Zealand English. These have all been force-aligned at the phoneme-level, and are stored with many layers of annotation some which have been automatically generated, and some which have been manually annotated. We interact with the corpus via our custom LaBB-CAT interface (LAnguage, Brain and Behaviour Corpus Analysis Tool). I will begin the talk by describing and demonstrating the corpus, and its associated LaBB-CAT tool. I will then focus on one particular recent study which has used the corpus, which aims to understand processes of sound change.

The combination of the time-depth of the ONZE collection, and the degree of careful annotation it contains, makes it an ideal data-set for the study of mechanisms underlying sound change. In particular, we aim to address the question which has been the subject of long-standing debate in the sound-change literature do sound changes proceed uniformly through the lexicon, or are there word-specific changes, with some words more ahead in the change than others? I describe a study which aimed to investigate this question by focusing on the mechanisms underpinning the New Zealand English front short vowel shift, of the vowels in words like bat, bet and bit. We automatically extracted formant values for over 100,000 tokens of words containing these vowels, We show that this data contains good evidence for word-specific effects in sound change, and argue that these are predicted by current models of speech production and perception, in combination with well-established psycholinguistic processes.