

Scripturient

Language as data



Barbara Lemon is a curator, historian and library professional who specialises in oral history and digital collections for national libraries. See more at www.lemonbell.press

The Language Data Commons of Australia (LDaCA) is a government-funded project that brings together language collections from Australia and the Pacific. Catherine Travis is Professor of Modern European Languages at the Australian National University, and one of the project's founding Chief Investigators.

NEARLY twenty years ago, the Australian Broadcasting Corporation (ABC) released a documentary on *The Sounds of Aus*, narrated by the late comedian John Clarke. “You can’t tell an Australian by the way they look”, said the promo blurb, “but the minute they open their mouths, the guessing is over”.

Few of us around the world stop to think about why we speak the way we do, and yet we take all sorts of social cues from one another’s accents, pronunciation, articulation and sentence structure..

It’s this kind of detail – the zeroes and ones of language – that fascinates linguist Catherine Travis and her colleagues. “What we’re interested in is the way that the language is used, be that how words are pronounced, what grammatical constructions are used, what words are used, how people might change the way they speak in one setting versus another, with one interlocutor versus another interlocutor, how the language changes over time. So in that sense we really are looking at the language as primary data.”

An “absolutely central component” of the LDaCA project is documentation of Aboriginal and Torres Strait Islander languages, many of which are at risk. “If the language isn’t being spoken by children, then the language will not continue, so there’s a real urgency there”.

Pacific languages are another priority. As well as Australian English, the project also looks at migrant languages, including their impact on Australian English, how they have changed over generations, and how bilingual speakers combine language in their speech.

Catherine recounts the story of a home visit some years ago at the invitation of the late Barbara Horvath, a linguist who recorded interviews with Australians

of Anglo Celtic, Greek and Italian background in the 1970s. It was one of those golden moments for a researcher. “We did a kind of archaeological dig through her garage. Of the 177 interviews on something like 130 cassettes, 125 of them we were able to digitise.” That enabled detailed comparison of recordings with Italian and Greek Australians today.

Her current project, *Voices of Regional Australia*, was devised during the horrific Black Summer bushfires of 2019-20 and seeks to capture voices outside of major urban centres. It draws on the widely used sociolinguistic interviewing technique known as ‘danger of death stories’.

“The idea is that if you get a person recounting a story when they’re in danger of their life, they will get so engrossed in the story that they won’t be thinking about how they’re speaking and you’ll get a very natural narrative.” Beyond linguistic study, she says, these “first-hand accounts of amazing resilience” can be invaluable for community healing.

But how can this kind of data be usefully recorded and made accessible for analysis? Linguists used to rely a lot more on handwritten vocabulary lists, phonetic

transcriptions, reel-to-reel cassette recordings and translation methodology (asking the speaker of another language, “how do you say ‘the dog chases the kangaroo?’”).

“Today I think people are keen to use more continuous speech and more natural speech and the tools that are available allow us to do that.” Those tools include digital recording devices, software for machine-generated transcription in multiple languages, and customisable computing platforms such as Jupyter Notebook for sophisticated data analysis. The challenge is to track, list and preserve what is now an abundance of data.

“LDaCA does not have long term archiving capability, and currently there really is no solution for that in Australia. Obviously the National Library can only hold so much. One of the major driving forces behind LDaCA is to catalogue the huge numbers of collections that are out there, so at least they’re findable.”

One wonders, how many other garages across the world are just waiting for an archaeological dig? **IP**

● For more about the project, see www.ldaca.edu.au



Wildfire, Australia 2019.