



Records of Pacific Languages: Where Are They and Who Can See Them?

Nick Thieberger

DOI: 10.25911/4Q9D-VM92

5 September 2025

Australian researchers have a long tradition of working in the Pacific and Papua New Guinea. University-based linguists in particular have been travelling since the 1960s to remote parts of the region to learn some of the nearly 2,000 languages spoken there. Most of these languages continue to be spoken, in defiance of the colonial monolingual enterprise that considers them an impediment to progress. A language may provide a source of identity, but multilingualism, the ability to speak or understand other languages, is normal, and is valued. Despite this value, rising rates of migration and urbanisation are contributing to the decline in use of local languages. This increases the urgency of creating records for future reference, and of preserving existing records where they exist.

Since early this century our project, PARADISEC (Pacific and Regional Archive for Digital Sources in Endangered Cultures), has been locating tape recordings made in the region by Australian linguists, musicologists, and anthropologists, but held in Australia. PARADISEC is a collaboration between three Australian universities: the University of Sydney, the University of Melbourne, and The Australian National University. We have been working to find relevant analogue records and to digitise them, putting them into an online system that makes them discoverable, joining increasing numbers of born-digital field recordings. Depositors specify what kind of access and use can be made of the materials; we always aim to make items as accessible as possible, but we also use a takedown principle in case we are advised of inappropriate content.

It was not the responsibility of any Australian institution to locate and preserve these records, which were stored in offices or in deceased estates — often the only recordings ever made in a given language. They provide a snapshot of the way a language was spoken in the past, and can contain oral tradition or musical performances that have now been forgotten. PARADISEC now holds over 18,800 hours of audio material, representing 1,400 languages from more than 100 countries. We consider this work to be part of our responsibility to the speakers recorded in the past and to their descendants — a post-colonial effort to return their cultural material. As mobile phones become more common in even the most remote parts of the Pacific, their users are finding items in our collections, sometimes the only information available in their language.

There are two main goals in doing this work: the first is to get material into a shape that allows it to be found and accessed by the people most closely associated with the content: the speakers and their descendants. We have had a number of people get in touch with us to express their joy at hearing their ancestors, as can be seen in this example:

It is most heartwarming to know that it is possible to sustain the life of my language. Thank you once again for the opportunity to listen to the records! As I see it now, it is 'US OR NEVER!' I believe that if we don't teach ourselves now and pass on, the language will be totally gone in the next 10 to 20 years.¹

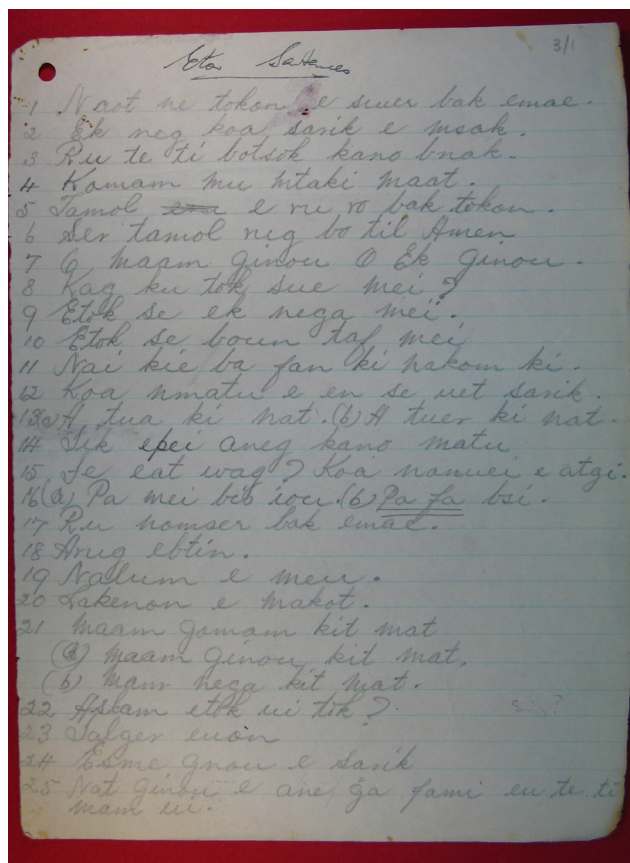
In this way, we aim to restore this material to the people recorded who may reasonably have an expectation of being able to access it, having spent their own time in teaching the foreign researcher. In some cases, it is clear that the speakers are well aware that the recording offers a way of speaking to the future. For example, when I asked people in Erakor Village in Vanuatu to record a story it was not uncommon for them to directly address the children of the future in the recording, or to record customary stories in order that future generations should have access to them. It is a responsible act of reciprocity on the part of the Australian research community that we can care for this legacy and make it accessible.

A large collection of notes made by the late Professor Arthur Capell was held on paper and only accessible by visiting the home of his executor. We digitised some 15,000 pages of language records and put them online, as can be seen in the example page of sentences from Eton in Efate, Vanuatu (Figure 1).

The Vanuatu Cultural Centre has thousands of tapes, and other agencies, like the Solomon Islands National Museum, have similar numbers. These tapes are no longer playable as the machines to play reel-to-reel tapes are hard to find and to maintain without specialist knowledge, and similarly, well-maintained cassette decks are necessary for getting the best result in digitising cassettes. PARADISEC has, over the years, digitised some 1,000 of the most at-risk tapes for these agencies, and continues to do this work with whatever funding we can arrange.

Recently, the State Archives of Yap in Micronesia asked us to digitise 120 of their tapes. We got funding from UNESCO Korea² to assist with this work and the

Figure 1: Digitised page of Eton sentences from Arthur Capell's fieldnotes.



Source: [Arthur Capell fieldnotes](#) ©.

tapes and files have been returned to Yap. The collection is listed in our catalogue, but is 'private' which means it is not visible to the world; essentially, it is a safe backup of the files. In this way, we can refresh collections if a disaster occurs and they are lost locally.

We have also been exploring the use of Raspberry Pi computers for return of materials, taking advantage of their local wi-fi transmitter to create a catalogue of the set of files relevant to that particular location (see McDougall 13/2/2024).

The second goal of our work is to ensure that research can be based on citable data, and can build on that data. We are all familiar with citation forms for published sources, and a repository like PARADISEC provides a similar service for primary records, with persistent identifiers for images, audio, or video files and their transcripts. You cannot cite data that only

exists in your office or on your laptop and expect others to be able to access your citations, so primary records in PARADISEC can be cited via digital object identifiers (doi). This is good research methodology, and allows new research to be based on the primary records. There is always more in a recording than the original recorder has analysed. For example, specialists in phonetics, or intonation, could analyse texts and their media to understand more about the way a language works. Similarly, an ornithologist could identify bird calls in field recordings.

In the past, there were archives and libraries that would take the output of a linguist's fieldwork. But someone needed to deposit them, and to describe them, which was rarely done. And, if they were deposited, often they were not described in sufficient detail to allow them to be found. And, on top of all that, they were even more rarely digitised for access outside of the holding institution. Digitising is the necessary first step but is not sufficient in itself. We have all had the experience of being unable to open digital files on our own computers over time. The work of PARADISEC includes converting files to archival standards and ensuring they will be usable in future. We regularly ask researchers to let us know about collections that need preserving.

PARADISEC is academic research infrastructure that, by curating unique records in many of the world's Indigenous languages, is also providing a community service in a public-facing collection with deep connections with agencies in the Pacific.

Author notes

Nick Thieberger is an associate professor in the School of Languages and Linguistics at the University of Melbourne, and is the director of PARADISEC. PARADISEC has been funded by the Australian Research Council and by the Language Data Commons of Australia.

Endnotes

1. E'ava Geita 23/3/2015. Personal communication.
2. Administered by the ACC-MOWCAP (Asia Culture Center and the UNESCO Memory of the World Committee for Asia and the Pacific) grants team, Exchanges & Public Relations Division, ACC.

Reference

McDougall, D. 13/2/2024. [Using Raspberry Pi in Ranongga](#). Endangered Language and Cultures blog.



Pacific Research Program *An Initiative of the Australian Aid Program*



Australian
National
University

Department of
Pacific Affairs

Development
Policy Centre

**LOWY
INSTITUTE**

The **Department of Pacific Affairs (DPA)** is the leading international centre for applied multidisciplinary research and analysis concerning contemporary state, society and governance in the Pacific. DPA acknowledges the Australian Government's support for the production of the In Brief series. The views, findings, interpretations and conclusions expressed in this publication are those of the authors and not necessarily those of the Australian National University or the Australian Government. The Australian Government, as represented by the Department of Foreign Affairs and Trade (DFAT), does not guarantee, and accepts no legal liability whatsoever, arising from, or connected to, the accuracy, reliability, currency or completeness of any information herein. This publication, which may include the views or recommendations of third parties, has been created independently of DFAT and is not intended to be, nor should it be, viewed as reflecting the views of DFAT, or indicative of its commitment to a particular course or courses of action.

dpa@anu.edu.au

bellschool.anu.edu.au/dpa

[DepartmentofPacificAffairs](https://www.facebook.com/DepartmentofPacificAffairs)
 [@anudpa](https://twitter.com/anudpa)

[DepartmentofPacificAffairs](https://www.linkedin.com/company/DepartmentofPacificAffairs)

ISSN 2209-9557 (Print)

ISSN 2209-9549 (Online)