

Taking language archiving seriously

A/Prof Nick Thieberger, School of Languages and Linguistics, University of Melbourne, Australia

There are not enough language archives in the world, and there are not enough linguists using the ones that do exist. How can this situation be changed? Fieldwork reciprocity can include an expectation by people we (linguists) work with that we will care for the records we make with them. As a profession linguists have made some progress in acknowledging the need to create documentation in addition to analysis, and data archiving in addition to data storage, to use that documentation as a warrant for our research claims, and to allow others to build on the documentation for new research questions. All of that is predicated on the materials being findable, accessible, licensed for use, and re-usable. This is clearly a broader issue for all research that risks being lost due to a lack of suitable repositories, but here I focus on language documentation.

Despite the development of archives in the past two decades specifically for digital language materials, much more remains to be done. The archives that exist have developed exemplary methods for curation of research materials, but not enough archives have been established. Storage provided by our universities is a necessary first step, but is insufficient to allow future access and re-use of the material. Further, not all linguists are archiving the primary records created in their fieldwork or other research. There are a number of reasons for curating language records that will be outlined in this presentation. I will discuss the gap between documentation efforts and what is archived, comparing, for example, the known number of fieldwork efforts on a language with archived materials for that language.

To some extent, archiving has been driven by what we can call 'coercive' archives, that is, those archives attached to the two major funding programmes, DoBeS and ELDP, who require(d), more or less successfully, materials to be archived as a condition of their grants. Other archives engage in targeted digitisation of analog material, and generally receive materials from researchers who recognise the need to preserve the records they have created. I will discuss some incentives that can be provided to ensure greater participation.

I will present a brief outline of the international network of language archives and current work to update the Open Language Archives Community (OLAC). I will then focus on recent innovations at the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) that aid access to items in the collection, and make its 260 terabytes more secure.

Nick Thieberger. 2016. What Remains to be Done: Exposing Invisible Collections in the other 7,000 Languages and why it is a DH Enterprise Digital Scholarship in the Humanities.  
<http://dx.doi.org/10.1093/llc/fqw006>