

Does Language Technology Offer Anything to Small Languages?

Nick Thieberger

PARADISEC, University of Melbourne/

University of Hawai'i at Manoa

thien@unimelb.edu.au

The effort currently going into recording the smaller and perhaps more endangered languages of the world may result in computationally tractable documents in those languages, but to date there has not been a tradition of corpus creation for these languages. In this talk I will outline the language situation of Australia's neighbouring region and discuss methods currently used in language documentation, observing that it is quite difficult to get linguists to create reusable records of the languages they record, let alone expecting them to create marked-up corpora. I will highlight the importance of creating shared infrastructure to support our work, including the development of Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC), a facility for curation of linguistic data.