Reflections on the Design of Indigenous PhD Projects

Steven Bird
Northern Institute
Charles Darwin University
Darwin, Australia

Much of the NLP research community approaches indigenous languages just like any other language, as a source of data to be exchanged for promises of future language technologies, particularly machine translation and speech recognition. PhD students entering this space are often motivated by the desire to use their skills for social good (cf. Jin et al., 2021). What better way than to open another door to participation in the global information society, while saving a language in the process. The challenge is to recruit the speech community and extract as much data as possible. Such work has been increasingly called out as Eurocentric and neocolonial (Bird, 2020; Meighan, 2021; Liu et al., 2022; Schwartz, 2022).

A different starting point is indigenous self-determination (United Nations, 2007). The implications for PhD projects are far-reaching, but include attention to many responsibilities that follow from self-determination, e.g., indigenous leadership, impact, sustainability, and accountability (Fig. 1).
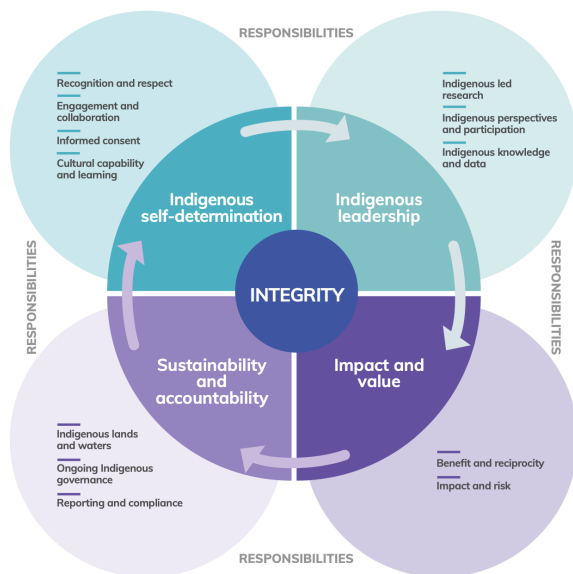

Figure 1: AIATSIS Research Ethics Framework (AIATSIS, 2022)

Background. Extractive approaches work in many places, and I have participated in such work in Cameroon, Papua New Guinea, and Brazil (Bird, 1997, 2003, 2010; Bird et al., 2014). The basis for local participation is usually the assent of community leadership coupled with work for hire by suitably qualified language workers (e.g., Adda et al., 2016; Amith et al., 2021). It is within the self-determination of indigenous groups to participate in this way.

However, I have found that the situation in many indigenous Australian contexts does not work in this way. Resilient communities enact their sovereignty by shaping outside engagements to serve local needs and aspirations (Bird and Yibarbuk, 2024). After three years of finding my way in indigenous Australian contexts I began supervising PhD projects, discussed below.

Indigenous PhD Projects at the Top End Language Lab. I give a brief overview of six PhD projects involving Indigenous research conducted by non-indigenous scholars. All projects have been conceived as decolonising, and have been brought under Indigenous authority in the places where they have been conducted, and all have been approved by our institutional human research ethics committee.

The first two PhD projects concerned technology-driven engagements, supporting text entry and transcription respectively (2017–21) (Lane, 2023; Ferrand, 2023) A common theme was to shift away from batch processing of speech (whether prerecorded or spontaneous) to interactive, collaborative transcription. In reflecting on what worked on the ground, it became abundantly clear that local participants were not fundamentally interested in a textual product, but in knowledge transmission. When teaching the newcomer, the starting point was a pre-existing relationship and the local desire to school the newcomer in

local life ways. Newcomers repeatedly wanted to see words written down, and so this motivated some locals to use written forms. When teaching children, the starting point was transmission of life-crucial knowledge (Lewis and Simons, 2016), particularly traditional ecological knowledge conveyed through stories of the country. Here, the practice of inscription, or writing down just enough in order to prompt an oral retelling, proved useful.

The next two PhD projects examined the inner workings of indigenous-led programs (2017-22) (Curtin, 2023; Wiltshire, 2024) A common theme was to shift away from typical western theories of change to local understandings of situated and embodied teaching and learning. The first project considered the programs of a remote indigenous language centre, a culturally safe space where an ancestral language is being revived and taught in local schools. The second project considered indigenous cultural tourism programs and the way that tour guides enacted their agency in creating a safe space for intercultural engagements, including some language teaching. These projects used Realist Inquiry and Constructivist Grounded Inquiry methods (Pawson et al., 1997; Charmaz, 2014).

Two further PhD projects are looking at hooks to engagement in language work (2021-2025). A common theme is to shift away from the direct approach of engaging people in explicit language work like elicitation, transcription, and translation. The first project is examining knowledge of food, including species, seasons, and procedures, and how this knowledge is transmitted via the development of a game (Hlaváčková, ms). The second considers knowledge of climate, including sensing and communication around severe weather events, in collaboration with the Australian Bureau of Meteorology and the Northern Territory Emergency Services (Aquino, ms). These projects are using Participatory Action Research and Appreciative Inquiry (Lenette, 2022; Bushe, 2013).

Lessons learned. There are manifold challenges when working with a remote indigenous community. The most superficial is funding, as accommodation in remote communities can be expensive, and because it is culturally appropriate to work with people in small groups (multiplying the hourly cost). We build relationships and the responsibilities around those relationships (Fig. 1) with the consequence that they are not manipulative, as

means to an end, but part of the ends themselves. We agree on the work and establish governance, and obtain institutional ethics approval. Such work adds significant extra cost, and at least a year to the duration of an office-based PhD project. The first lesson learned is that *the extra time and cost must be taken seriously and supported by the institution.* Often, additional external grant funding is needed on top of a PhD scholarship.

A second challenge is in the area of entering the local lifeworld, which requires significant cultural induction, and ongoing cross-cultural engagement. Here, a lesson learned is that *prospective candidates must show a commitment to cross-cultural living coupled with language learning.*

A third challenge is conducting all research in the context of ongoing relationships, and that these relationships cannot be picked up and set down. Here, a lesson learned is that there is no cold start and no cold finish: *relationships with community members must be already established (either by supervisor or candidate) and must continue after the project ends.*

A correlate of the relational foundation of respectful indigenous engagements is the basis on which locals participate. This is often oriented around the relationship, and in teaching the newcomer how to live well in this place. If this is the local motivation, then any data collection needs to honour this. For instance, if we are collecting plant names, it is because locals want us to learn them, and so we must go beyond mere capture to make a serious attempt at memorisation. Once newcomers show they are serious to learn, locals may go on to share practices and stories associated with those species, leading to further learning. Here, a lesson learned is that *the local agenda of knowledge transmission should not be confused with participation in the external program of data extraction.*

There are many further lessons to discuss, e.g.: the need to continually guard against deficit framings and seek out strength-based approaches (cf. Bushe, 2013); and the need to let go of western disciplinary boundaries and shift into a transdisciplinary mode (cf. Christie, 2006). The presentation will cover a comprehensive range of such lessons and show how they are all tied back to the requirements for ethical indigenous engagement and for centering the speech community (Bird and Yibarbuk, 2024).

References

Gilles Adda, Martine Adda-Decker, Odette Ambouroue, Laurent Besacier, David Blachon, Hélène Bonneau-Maynard, Elodie Gauthier, Pierre Godard, Fatima Hamlaoui, Dmitry Idiatov, Guy-Noël Kouarata, Lori Lamel, Emmanuel-Moselly Makasso, Annie Rialland, Sebastian Stüker, Mark Van de Velde, François Yvon, and Sabine Zerbian. 2016. Innovative technologies for under-resourced language documentation: The BULB Project. In *Workshop on Collaboration and Computing for Under Resourced Languages, International Conference on Language Resources and Evaluation*, pages 59–66. ELRA.

AIATSIS. 2022. *AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research*. Australian Institute of Aboriginal and Torres Strait Islander Studies.

Jonathan D Amith, Jiatong Shi, and Rey Castillo Garcia. 2021. End-to-end automatic speech recognition: Its impact on the workflow for documenting Yoloxóchitl Mixtec. In *First Workshop on NLP for Indigenous Languages of the Americas*.

Steven Bird. 1997. A lexical database tool for quantitative phonological research. In *Proceedings of the Third Meeting of the ACL Special Interest Group in Computational Phonology*, pages 33–39. ACL.

Steven Bird, editor. 2003. *Grassfields Bantu Fieldwork: Dschang Tone Paradigms*. Linguistic Data Consortium. LDC2003S02, ISBN 1-58563-254-6.

Steven Bird. 2010. A scalable method for preserving oral literature from small languages. In *Proceedings of the 12th International Conference on Asia-Pacific Digital Libraries*, pages 5–14.

Steven Bird. 2020. Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, page 3504–19, Barcelona, Spain.

Steven Bird, Lauren Gawne, Katie Gelbart, and Isaac McAlister. 2014. Collecting bilingual audio in remote indigenous communities. In *Proceedings of the 25th International Conference on Computational Linguistics*.

Steven Bird and Dean Yibarbuk. 2024. Centering the speech community. In *Proceedings of the 18th Conference of the European Association for Computational Linguistics*.

Gervase R Bushe. 2013. The appreciative inquiry model. In Eric H. Kessler, editor, *Encyclopedia of Management Theory*. Sage Publications.

Kathy Charmaz. 2014. *Constructing Grounded Theory*. Sage.

Michael Christie. 2006. Transdisciplinary research and Aboriginal knowledge. *Australian Journal of Indigenous Education*, 35:78–89.

Nicole Curtin. 2023. *Moving Forward while Looking Back: Exploring Reconciliation Pathways through Indigenous Knowledge Sharing in Tourism*. Ph.D. thesis, Charles Darwin University.

Éric Le Ferrand. 2023. *Leveraging Speech Recognition for Interactive Transcription in Australian Aboriginal Communities*. Ph.D. thesis, Charles Darwin University.

Zhijing Jin, Geeticka Chauhan, Brian Tse, Mrinmaya Sachan, and Rada Mihalcea. 2021. How good is NLP? A sober look at NLP tasks through the lens of social impact. In *Findings of the Association for Computational Linguistics*, pages 3099–3113.

William Lane. 2023. *Morphological Complexity and Interactive Language Processing*. Ph.D. thesis, Charles Darwin University.

Caroline Lenette. 2022. *Participatory Action Research: Ethics and Decolonization*. Oxford University Press.

Paul Lewis and Gary Simons. 2016. *Sustaining Language Use: Perspectives on Community-Based Language Development*. SIL International.

Zoey Liu, Crystal Richardson, Richard Hatcher, and Emily Prud'hommeaux. 2022. Not always about you: Prioritizing community needs when developing endangered language technology. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3933–3944. Association for Computational Linguistics.

Paul J Meighan. 2021. Decolonizing the digital landscape: The role of technology in indigenous language revitalization. *AlterNative: An International Journal of Indigenous Peoples*, 17:397–405.

Ray Pawson, Nick Tilley, and Nicholas Tilley. 1997. *Realistic Evaluation*. Sage Publications.

Lane Schwartz. 2022. Primum Non Nocere: Before working with Indigenous data, the ACL must confront ongoing colonialism. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 724–731. Association for Computational Linguistics.

United Nations. 2007. United Nations Declaration on the Rights of Indigenous Peoples. https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html.

Brandon Wiltshire. 2024. *"You gotta feel good first before you can start learning language": Understanding how and why Indigenous language revitalisation programs work according to local perspectives*. Ph.D. thesis, Charles Darwin University.