Learning From Failure: Data Capture in an Australian Aboriginal Community

Éric Le Ferrand,[1,2] **Steven Bird,**[1] **and Laurent Besacier**[2]
[1]Northern Institute, Charles Darwin University, Australia
[2]Laboratoire Informatique de Grenoble, Université Grenoble Alpes, France

Abstract

Most low resource language technology development is premised on the need to collect data for training statistical models. When we follow the typical process of recording and transcribing text for small Indigenous languages, we hit up against the so-called "transcription bottleneck." Therefore it is worth exploring new ways of engaging with speakers which generate data while avoiding the transcription bottleneck. We have deployed a prototype app for speakers to use for confirming system guesses in an approach to transcription based on word spotting. However, in the process of testing the app we encountered many new problems for engagement with speakers. This paper presents a close-up study of the process of deploying data capture technology on the ground in an Australian Aboriginal community. We reflect on our interactions with participants and draw lessons that apply to anyone seeking to develop methods for language data collection in an Indigenous community.

1 Introduction

For decades, the work of collecting data for Indigenous languages has been the province of documentary and descriptive linguistics (Bouquiaux and Thomas, 1992; Vaux and Cooper, 1999; Meakins et al., 2018). This work has involved various kinds of elicitation, e.g. of word lists, phrases, etc, to support description of the phonology, morphosyntax, and grammar of the language. It has also involved the collection of unrestricted text, through recording and transcription. In most cases, the result is audio with aligned text. Many software tools have been developed for supporting these activities (Boersma, 2001; Clark et al., 2008; Hatton, 2013; Sloetjes et al., 2013).

Within the field of natural language processing, established practice is to support the linguist's work (Michaud et al., 2018; Seifart et al., 2018; Foley et al., 2018; Cox et al., 2019). In some cases, this includes the participation of speakers in activities using apps controlled by linguists (Bird et al., 2014; Hanke, 2017; Bettinson and Bird, 2017). However, the premise is basically the same: obtain a substantial quantity of audio and transcribe it, or post-edit the output of an automatic transcription system.

We believe that these approaches do not adequately address a fundamental reality of small languages: they are *oral*. There may be an official orthography, but it has no place in the local language ecology where any written business takes place in a language of wider communication. As a result, local people are usually not confident in the orthography of the language. Furthermore, there may be low confidence in using computers and text editors, and inadequate support for the language in terms of keyboarding and spelling correction. Add to all this the fact that the whole space of rendering an oral language into standardised orthography can be alienating (Dobrin et al., 2009; Hermes and Engman, 2017).

There is no particular reason for NLP approaches to Indigenous languages to follow the long-established practices of linguists. After all, there is an equally long history of algorithmic approaches being profoundly different to the human tasks they replicate. For instance, a human sorting a hand of cards may use insertion sort, but a machine might use Quicksort, with better average-case complexity (Levitin, 1999). Computational approaches may be inspired by analogy, e.g. simulated annealing, genetic algorithms, neural networks, but they are not required to adhere to the human defined process. Accordingly, we can ask, what is an idiomatic computational approach to collecting data for Indigenous languages that is a better fit to the capabilities of human participants? In the case of associating text and speech, we believe that the answer might be keyword spotting. This is because, in our experience, speakers and learners are attuned to identifying whole words, rather than obsessing

Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics
Volume 1: Long Papers, pages 4988 - 4998

about the idiosyncratic phonetic makeup of individual tokens as required for phone transcription (cf. Bird, 2020b, 718f).

Accordingly, we investigate an approach to transcription based on word spotting known as "sparse transcription" (Bird, 2020b). This would seem to be an easier, less specialised task than direct, contiguous transcription. If more people can participate, we can hope to establish a virtuous circle with more data, better models, less correction, even more data, and so on. The idea is that transcription can be accelerated by identifying the tokens of high-frequency terms all at once, then playing them back in quick succession for confirmation by participants.

This paper reports on the deployment of a lexical confirmation app which supports human confirmation of system hypotheses. We begin by describing the background to this work (Sec. 2), including related work on designing technology for use in Indigenous places. We also describe the site where we work and the design of the lexical verification app. Next, we report what happened when we deployed the app in two field tests, including detailed accounts of interactions with participants (Sec. 3). In the discussion section, we reflect on the field experience from a variety of perspectives, trying to draw out lessons that may be applicable to other places where NLP researchers seek to design technologies for language data collection (Sec. 4). The paper concludes with a summary and prospects for further research.

2 Background

2.1 Designing in an Indigenous context

Designing in the Indigenous space is a small but growing area within the field of Human-Computer Interaction (HCI). Projects in this space often begin with ethnographic research to identify local priorities. Co-design is advocated as a way to establish a "culturally-tailored, culturally-enriched and trustworthy environment for participation" (Peters et al., 2018). The focus of this work includes traditional knowledge (Verran, 2007), language revitalisation (Hardy et al., 2016) or media sharing (Soro et al., 2017). Recent research mentioned the need to involve stakeholders in a system design (Lynch and Gregor, 2004) highlighting the challenges related to the transparency of the mechanism of a given system, specifically when machine learning is involved (Loi et al., 2019) and the difficulty to explain to the users such mechanism (Abdul et al., 2018). The lack of published accounts of experiences collecting language data in Indigenous contexts, specifically in the intersection of NLP and documentary linguistics, makes it difficult for newcomers like us to devise approaches that are likely to work. We address this shortcoming by reporting and reflecting on our field experience.

Deploying speech technologies in remote Aboriginal communities is challenging, not primarily because of low technological literacy on the part of local people, but because of low interactional literacy on the part of NLP researchers who enter indigenous places to gather data.

2.2 Working in an Indigenous place

Our work is grounded in Bininj country in Arnhem land in the north of Australia. The biggest town is Gunbalanya with 1,100 inhabitants where we can find primary and secondary schools in which teaching is done in English. A few remote satellite communities, or "outstations," can be found throughout this country in which education of young people takes place in a bi-cultural environment both in Kunwok and English.

Kunwok (ISO gup) is the main language of communication here, and Kunwinjku is the prevalent dialect. It is spoken by some 2,500 people and is one of the few Australian languages which is gaining speakers (Evans et al., 2003). While a standard orthography exists, most community members do not write at all. When pressed, some of them are able to leverage their knowledge of English literacy in order to decode Kunwok texts (cf. Feinauer et al., 2013; August et al., 2009).

In prior work in Bininj country, we discussed our work with traditional owners (heirs of a given tract of Aboriginal land and leaders of the community). We described and demonstrated prior work involving transcription, and how it can be used to transcribe Kunwok. They raised their concerns about intergenerational knowledge preservation and transmission and access to the resources created by westerners. While it is not clear to us that the nature of our work had been thoroughly understood, we could identify through this interaction topics which are addressed by current speech processing and HCI research projects (San et al., 2021; Taylor et al., 2020). Our work took place in Gunbalanya and Manmoyi, a remote community situated 5 hours drive from Gunbalanya.

Australian Aboriginal communities are far from uniform. The experiences and challenges we describe here may be relevant for the Australian Top End, but they cannot be directly applied to Indigenous communities in other places.

Figure 1: Screenshot of the Spoken Term Detection based Lexical Verification App

2.3 Lexical verification app

We designed a lexical verification app that bridges the output of a spoken term detection system to people. It was built following the design of Bettinson and Bird (2017). We focused on a simple design without any textual component besides the transcription of the query term. The idea is to first load in a web app the query/utterance pairs generated by our spoken term detection system. We then ask speakers of the target language to confirm for each pair if the query word (i.e. the term we are trying to retrieve) is pronounced in the search utterance (i.e. the sentence in the speech collection in which the query term was detected).

The participants have six buttons available to perform the task. They have two play buttons at the bottom left: One to play the query term, the other to play the search utterance. Once the two audio files have been listened to, two feedback buttons appear at the bottom right to allow the user to confirm if the query term is included or not in the utterance. We also added two arrows on each side of the top of the screen to allow the user to jump to the previous or the next example. When a new example is displayed on the screen, the query term is played automatically When the utterance is played, the transcription of the query term is highlighted around the timestamps in which the query term was detected. The terms are spotted in the utterances beforehand following the parameters of the sparse transcription simulation proposed by Le Ferrand and Bird (2020). Because of the challenges posed by the remote Aboriginal context such as the lack of reception or proper working facilities (e.g. a table), we needed to find solution in terms of data storage and activity design. Based on the work of (Bettinson and Bird, 2021), we stored the query/utterance pairs output by our spoken term detection system in a JSON file and loaded them in a Raspberry Pi with the app. The Pi acts as a WiFi hotspot to which any device can connect. We can then then connect a tablet to the Pi and, doing so, the feedback provided by the participant can directly be stored in the associated database.

3 Fieldwork

We tested our approach with two trials in two Aboriginal towns, with three people in each place. While the number of participants seems small, larger trials are difficult to arrange in Aboriginal contexts due to the small number of speakers. At the beginning of each elicitation session, the first author explained our intention to teach a machine to transcribe the language automatically, and that we wanted help to correct system guesses. There is actually no direct translation of transcription in Kunwok and the concept is usually given by the formulation *karribimbun kure djurra*, “we’re drawing on paper”.

In both places, we recruited the participants with the support of two local institutions, the art centre in Gunbalanya and the ranger organisation in Manmoyi. At the start of our trips, the first author introduced himself to the communities and explained that he was looking for people to support him for language work. Then the people interested came to find him throughout the day. Each session lasted approximately 15 minutes and was part of other language work including recordings or language learning. Each participant was paid at the regular rate for language work.

3.1 Trial 1: Gunbalanya

For our first trial, we recorded source audio from a three hour guided tour of a local site. We transcribed a few minutes of this recording and used this transcription to build a lexicon. We used voice activity detection to segment the recording into breath groups. Finally, we automatically spotted terms from the lexicon in these breath groups. Since the speaker of the lexicon and the speech collection overlap, most of the terms spotted by the system were correctly retrieved. In the data presented to participants, the query term was present in the supplied phrase in 57% of the instances.

This configuration was tested with three Gunbalanya residents: SB (20s), TM (30s), and RB (40s). This last participant was also the speaker of the recordings.

SB appeared nervous and said little in response to our explanations and questions. When an audio clip was played, he translated, even though this was not the instruction. It was as if he projected his assumption about the purpose of the task, namely for the researchers to understand the content. At one point he respoke the query term and the target phrase in a single utterance, before explaining his knowledge about the associated place. The interface itself was not legible to him: faced with a choice of two play buttons – one for the query term and one for the phrase – he was never clear which one to press. He never used the thumbs up/down feedback buttons.

Here is an example of the confusing situation set up by our approach (we use "App" to indicate audio produced by the app, along with speaker initials, and ELf for the first author. "Play1" refers to the button that plays the query term and "play2" the utterance).

ELF	<press play1>
App	*manyilk*
ELF	<press play2>
App	*menekke mandjewk karuy*
ELF	*manyilk? larrh*. Because he says *mandjewk*
SB	*manyilk*, first <press play1>
App	*manyilk*

Notice that the query term *manyilk* "grass" is not contained in the utterance *menekke mandjewk karuy* "this wet season he dug it". When we demonstrate the use of the app by giving the expected response of *larrh* "no", SB asserts that manyilk is present, contradicting us. He presses on the query term play button to show us.

The following day, when we discussed with another participant, we heard that SB thought that our task was an attempt to test his memory.

RB was more confident than SB. He seemed intrigued at hearing his own voice on the device. For each audio segment we played, RB gave an interpretation of the content. We offered the device to him to control, but he declined. After we pressed the two play buttons, he waited, and we had to follow up with overt questions: "does he say <query term>?", or "can you hear <query term> in this sentence?" He answered as expected, with: "yes, <query term>" or "no, he doesn't say <query term>." Consider the following example:

ELF	<press play1>
App	*marnbom* ("he made")
ELF	<press play2>
App	*kumekke* artist *marnbom kadi*
ELF	do you hear *marnbom*?
RB	*marnbom* that's painting making the painting
ELF	but do you hear *marnbom* in the sentence?
RB	yeah

Unlike SB and RB, TM readily took the device and used the controls. Sometimes, when the query term was not contained in the utterance, he not only translated the audio, but he also offered an example sentence containing the query term. In the following example, "confirm" refers to one of the feedback button which automatically display the next example and play the query term:

TM	<press confirm>
App	*karrikadjung* ("we follow it")
TM	*karrikadjung*, ("we are following") <press play2>
App	*karrikadjuy* road ("we followed the road")
TM	he says *karrikadjuy*, it means we went this way road, he should have say we are following this one, *karrikadjung*

In this case, the difference between the query term *karri-kadju-ng* "we-follow-PRES" and the utterance *karri-kadju-y* "we-follow-PAST" is only in verb tense. The whole query term appears in the sentence, except for the tense marker. Should the speaker say yes or no? This points to a shortcoming of the task definition.

When the term was correctly retrieved, TM would respeak the audio and press the thumbs-up button. When the term was not correctly retrieved, TM offered extensive explanations.

3.2 Trial 2: Manmoyi

For the second trial, we visited the Manmoyi outstation. We used five short audio recordings from previous fieldwork, including guided tours and traditional stories. One of the recordings was transcribed and we extracted the words to use as our lexicon. As before, we segmented the source audio into breath groups and ran word spotting against

this set.

Since the speaker of the lexicon and those of the rest of the collection did not overlap, there was much lower precision; often a query term matched noise or mumbling. In the data presented to participants, only in about 10% of cases was the query term present in the supplied phrase.

This configuration was tested with three residents of Manmoyi: LY (60s), LB (50s), RG (50s).

LB and LY participated together, with LB taking an active role and LY only participating by talking to LB during the task. With each round, LB listened to the query term and the utterance then appeared to associate them as a single linguistic event, and he would recount a story that included both the term and the utterance. After this, he would give feedback (thumbs up or down) depending on how easy he found it to link the two semantically:

LB	<press play1>
App	*wirrihmi* ("dislike/wrong")
LB	that's "wrong one"
LB	<press play2>
App	*wanjh manjbekkan manmanjmak*
LY	it tasted sweet
LB	it tasted like you know this, it might have been a little bit funny or something like that
LB	yeah like for us they say: "no I can't eat" because he tasted it and they say "try it" and they gave it, and he says "aah yeah it tasted nice"
LB	*yoh*, that's the one, that's good, *kamak*

LB often interpreted the audio segment. At one point, he recognised the speaker for the queries, and he told us about her and began to recount the same story:

ELF	<press play1>
App	*nawernwarre* ("big brother")
ELF	<press play2>
App	*birribonguni birri...* ("they were drinking, they...")
LY	*nawernwarre*
LB	*yoh, nawernwarre*
LY	*nawernwarre*, or *manekke* might be... lonely boy (story)
LB	lonely boy *yoh* that's the lonely boy (story)

Towards the end of the session, we asked about LB's understanding of the task:

ELF	Can you tell me in English what do you think I am trying to do?
LB	You are trying to... you are making like Kunwok and English translating, but if you are making straight like Kunwok you're making straight and English making straight, that's the all same.
ELF	well, not really
LB	no it's real, we are talking, we know everything. Not all these, we've seen these people, they don't know anything about it, myself and LY we know everything about it.

LB understood this to be a translation activity. When we disagreed, he re-asserted his standing as a knowledge authority. Later, we explained our ultimate purpose of automatically transcribing the language. LB rephrased transcription as "make it together." We realised afterwards that LB may have been referring to his semantic linking process.

RG was our final participant, and this session revealed many issues. Given the low number of correct query-utterance pairs, we found ourselves needing to manually skip over utterances that were too hard to understand out of context. Each time we abandoned a round and moved on to the following round, the next query term played automatically (this feature was added before any testing with the assumption that it would speed up the verification process).

Such automation turned out to be confusing for RG. For a few instances, RG responded "yes" when the query term was not literally present in the utterance, maybe because the query term was morphologically related to a term that was present, e.g. *birri-m-h-ni* "they-towards-immediate-were" (query) and *birri-ni* "they-were". Another interpretation of this behaviour is that RG was focussing on meanings not forms. In this and other cases, it seems that RG was not clear about what we were asking for.

RG	The old woman is talking about country and the young fellow is talking about what creation was.
RG	It's all a bit confusing. They are not even saying *kunred* it means home, the young other fellow is talking about dreamtime story, so it is not, well it's connect but it is not pronouncing.

Sometimes, RG asked about the speakers and the

overall context of the out-of-context audio segment, asking, e.g. "Is this <name> speaking? I don't know what they're talking about here."

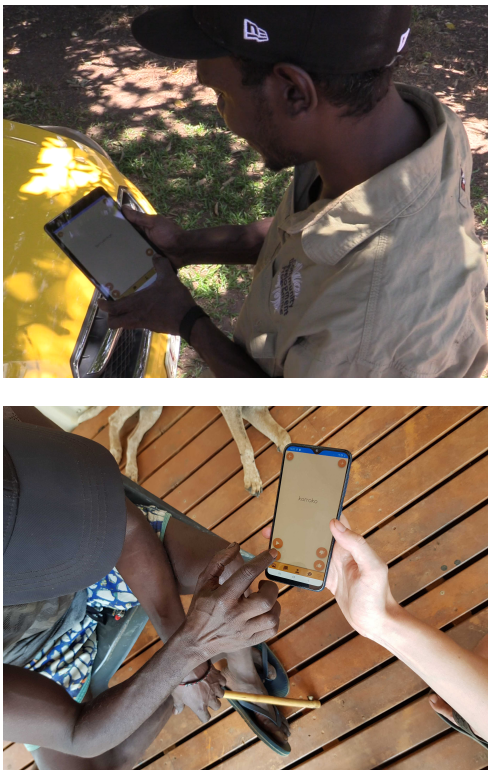

Figure 2: Use of the Word verification app

4 Discussion

There were many issues in the design and conduct of this elicitation activity, and it is clear that our approach need to be completely rethought. In this section, we analyse the above interactions and try to identify some principles to inform NLP elicitation methodology, hoping to avoid such problems occurring in future.

Task motivation. SB, RB and LB understood us to be interested in interpreting the content. SB thought we were testing his memory. TM offered detailed explanations. LB said things that we interpret as asserting authority. It appears that our attempt to explain our purpose in automatic transcription, and the activity of confirming or refuting system guesses, was unsuccessful.

Task definition. Participants were not clear about what we were asking of them. The notion of "word" was not clearly defined, and there were a variety of responses when the query term was not identical yet morphologically or semantically similar to a word in the corresponding utterance.

Naturalness of the task. When it comes to collaboration with western language workers, Aboriginal people in these communities are accustomed to participating in interviews, recordings, transcription, and translation activities. This may explain people's readiness to respeak or interpret the content or supply additional cultural information. We entered with a different task, one where the overt activity of human confirmation/rejection of system guesses was not transparently related to a recognisable transcription task. We explained and demonstrated the activity, but TM was the only participant to instantly grasp this task. Even so, he provided extensive explanations when the system guess was wrong in an effort to teach us.

Utterance context. From our perspective, the components of the device were clear. We have a query term that needs to be detected, and an utterance that should contain the query term. From this, we just need two feedback buttons to confirm whether the query term is included in the utterance. However, to the participant listening to the audio produced by the app and not following our use of the controls, the query term and utterance may be perceived as a single utterance. Everything put into the aural space appears to be concatenated by listeners, and our non-conventional metalinguistic context is not interpretable. When endeavouring to explain the task in Kunwok, we were hampered by the lack of words for "word" and "sentence".

Teaching. The participants generally provided much more information than the simple yes/no response we requested. Each instance was another opportunity to teach us about the language or the country. The design of the task only limited the space for this style of participation. The activity itself was not particularly engaging, taking utterances out of context and asking for a mechanical response to a seemingly pointless question. It seems to be a kind of resilience that participants made the most of the opportunity to pursue their own ends of educating newcomers. Further discussion with community members highlighted their concerns about knowledge preservation, access to archival recordings, and learning literacy.

Knowledge transmission. George et al. (2010) explains that the way in which westerners and Australian Aboriginal people transmit their knowledge varies in that one extracts, identifies and, categorizes while the other needs the information to be

embedded in a system of kinship relationships. For example, in Bininj country, every individual has a kinship relationship to every other individual, and they address each other accordingly (Glowczewski, 1989). Stories do not exist in isolation but are connected to an individual who tells them, and the country it comes from. We ran up against this when participants needed to connect isolated utterances back to their rightful cultural context, not just consider them as arbitrary linguistic material for which they can answer an unmotivated question: "does this utterance contain this word?" We can see this in Trial 2 where LB ignores the utterance and uses his knowledge of the speaker of the query term to link the content back to the story.

Yarning. Recent fieldwork methods research has shown that adopting Aboriginal-led approaches leads to more culturally appropriate practices and better feedback from Aboriginal consultants (Louro and Collard, 2021). Yarning has been described as a research method and the traditional way for Aboriginal people in Australia to pass knowledge. It can be defined as "a conversational process that involves listening to storytelling that creates new knowledge and understanding" (Terare and Rawsthorne, 2020). Adopting this to engage with participants could lead to better participation and a more appropriate way to collaborate. Here, the Aboriginal consultant would occupy a teaching role and the function of the technology would be to capture, support, and organise natural ways of transmitting knowledge.

Spoken term detection performance. The spoken term detection method delivered markedly different results in the two trials. Presenting data with 50% accuracy (first trial) makes the user's task seem most worthwhile, otherwise, the user is mostly confirming or refuting system guesses (refuting in 90% of cases in the second trial). If this reasoning is correct, then we predict that a trial involving 90% accuracy would also be challenging to motivate and teach. The low accuracy of the system probably contributed to the challenges encountered during the second trial. However similar behaviour in both trials was observed (e.g. the systematic translation after an audio was played or the semantic linkage process) which makes us think that the sole performance of a system is not the main source of the misinterpretation of the task.

App design. The design of the app was based on preliminary thinking about how collection could proceed fluidly. We did not consider the confusion that might be caused by having two play buttons on the screen (one for the query term, and one for the corresponding utterance). In the interests of efficiency, with each new round, the query term was played automatically. It was as if the thumbs up/down button from the previous round caused playback, and this turned out to be confusing. When we wanted to skip forward by a few examples using the right or left arrow keys at the top of the display (Fig. 1), the app would play a series of seemingly random words. Such automation should have been avoided, specifically in the early stage of our work when there was a lot of uncertainty regarding people reaction towards our activity.

Design improvements. Besides the elements we already mentioned, a few paths can be explored to address the challenges we have faced. Removing the query play button could have the effect of reducing the number of contexts and avoid the linkage process we have observed with LB and SB. Limiting the activity to a single story and playing the utterances in chronological order can make the context clear, and the participant would not need to clarify it. Using bottleneck features instead of MFCCs to spot words could improve the precision of the system (Menon et al., 2019).

Such modifications, however, cannot address the biggest flaw of our proposed task: it does not respond directly to people's agenda in terms of language work, but simply tries to leverage people's skills to respond to westerners' expectations. Pushing the proposed pipeline for several iterations would risk alienating our participants and compromising further collaboration. We believe that a complete reshaping of our method is necessary to enable a sustainable and community-based model for language and knowledge documentation.

5 Further Reflection

Our first attempt in this space was unsuccessful on many levels. Most superficially were issues with the task definition and the app interface. The task focused on the notion of "word" and on deciding whether a given word occurred in a given utterance. Yet the notion of word was not established; as an oral language, there was no *a priori* shared understanding between the participant's notion of spoken

word and our notion of orthographic word.

Throughout our interactions with participants, our attempts to explain the method and the purpose were unsuccessful. Local perception was fixed on the idea that we had entered the community to learn the language and culture, and that the purpose of participating in the study was to teach us and to interpret the texts for us.

Consequently, the narrow focus of our activity on eliciting a binary, thumbs up/down response was unsuccessful. This is hardly surprising as many people have noted that engaging Aboriginal people with direct questions requiring a yes or no response is seen as testing people's knowledge or memory, and potentially irritating (Maar et al., 2011; Ober, 2017). We observed this ourselves, when SB reported that he felt like he was being tested, or when LB responded as if his authority was being questioned.

Clearly, our style of engagement was not the expected kind of collaboration on a linguistic task. Aside from one participant (TM), no one would participate in the abstract and apparently pointless task of confirming whether a word was present in a sentence. Instead, all participants sought to create meaning from any language fragments they were presented with. On the basis of an isolated word, and person, place or story would be detected, and people would seek to teach us about these aspects of their lifeworld. This took various forms: repeating, paraphrasing, translating, interpreting, or offering extensive cultural commentaries.

In retrospect, this response to our approach comes across as resilient and generous. In comparison, our narrow focus on data collection, and on getting across the specialised task of lexical confirmation may have come across as disconnected from local interests, and potentially disrespectful.

Of course, we can hope to recruit more people like TM. However, the story about scalable creation of language resources involves working with whoever is available. The tasks need to be locally comprehensible and motivating. In moving forward, we believe it is necessary to rethink the collaborative transcription task. The starting point is to understand local participants as teachers and cultural guides, occupied with their own knowledge practices and with passing these on. Special focus need to be given on the creation of a third space between the several stakeholders of a project with benefits that serve both Indigenous participants and external actors (Bird, 2020a). Could we view the task of putting an audio recording into textual form as a way to help a newcomer make progress with the language and culture, and with getting the pronunciations and meanings correct? The answer to this question depends on further research.

6 Conclusion

Outside the major languages, the development of language technologies is considered to be held up by the general lack of data (Krauwer, 2003). In the case of the world's small, oral languages, the usual approach has been to follow the long-established practice of linguists and record and transcribe audio and elicit wordlists and paradigms. Many computational tools were developed to support this approach. However, algorithmic approaches to working with small languages do not need to be limited by these past practices, and so we believe it is worth considering other approaches to data collection that might simultaneously support computational methods while engaging effectively with members of the speech community.

Accordingly, we took a recently proposed approach to transcription based on keyword spotting, and developed an app for confirming system guesses. We anticipated that this app would be more accessible to local participants than the conventional linguist-driven tasks. We ran trials in two Aboriginal towns, with speakers of the Kunwok language.

In this paper, we report the description of the several interactions we had with locals around a lexical verification activity. We present the many challenges we encountered, including a reflection around the technical and cultural issues of the task design, and the flaws around our approach in terms of collaborative language work.

For the present, we offer our findings as a candid report on the experience of deploying data capture technology in an Indigenous community, in the hope that others will succeed where we have failed. We hope others will also follow our lead and share their own experiences of data collection, and make visible more of the real work of NLP (cf. Star, 2007). Perhaps it is possible for an externally-defined task such as transcription to be aligned to local agendas. Just as often, we expect that it will be necessary to let go of such tasks and do something different. Something that makes sense locally.

Acknowledgements

This research was covered by a research permit from the Northern Land Council, and ethics approved from Charles Darwin University. We are grateful to the Australian government for a PhD scholarship to the first author, and for grants from the Australian Research Council and the Indigenous Language and Arts Program to the second author.

The recruitment of participants was done with the support of the local organisation: Injalak Arts and Craft in Gunbalanya, and Warddeken Land Management in Manmoyi. The shape and purpose of the work was explained in English and oral consent has been obtained by all the participants. Additional approval has been given by Manmoyi traditional owners, concerning the collection and use of the data. All the participants have been paid at the regular rate for Aboriginal people consultancy. We would like to thank Mat Bettinson for his involvement in the design of the lexical verification App and Joshua Yang for the video recording of the trials.

References

Ashraf Abdul, Jo Vermeulen, Danding Wang, Brian Y Lim, and Mohan Kankanhalli. 2018. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–18.

Diane August, Timothy Shanahan, and Kathy Escamilla. 2009. English language learners: Developing literacy in second-language learners—report of the national literacy panel on language-minority children and youth. *Journal of Literacy Research*, 41:432–452.

Mat Bettinson and Steven Bird. 2017. Developing a suite of mobile applications for collaborative language documentation. In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 156–164.

Mat Bettinson and Steven Bird. 2021. Collaborative fieldwork with custom mobile apps. *Language Documentation & Conservation*, 15:411–432.

Steven Bird. 2020a. Decolonising speech and language technology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3504–3519.

Steven Bird. 2020b. Sparse transcription. *Computational Linguistics*, 46:713–744.

Steven Bird, Florian Hanke, Oliver Adams, and Haejoong Lee. 2014. Aikuma: A mobile app for collaborative language documentation. In *Proceedings of the Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 1–5. ACL.

Paul Boersma. 2001. Praat: A system for doing phonetics by computer. *Glot International*, 5:341–345.

Luc Bouquiaux and Jacqueline M. C. Thomas. 1992. *Studying and describing unwritten languages*. Dallas: Summer Institute of Linguistics.

Jonathan Clark, Robert Frederking, and Lori Levin. 2008. Toward active learning in data selection: Automatic discovery of language features during elicitation. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.

Christopher Cox, Gilles Boulianne, and Jahangir Alam. 2019. Taking aim at the 'transcription bottleneck': Integrating speech technology into language documentation and conservation. Paper presented at the 6th International Conference on Language Documentation and Conservation, https://instagram.com/p/Buho4Z0B7xT/.

Lise M Dobrin, Peter K Austin, and David Nathan. 2009. Dying to be counted: The commodification of endangered languages in documentary linguistics. *Language Documentation and Description*, 6:37–52.

Nicholas Evans et al. 2003. *Bininj Gun-wok: a pandialectal grammar of Mayali, Kunwinjku and Kune*. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Erika Feinauer, Kendra M Hall-Kenyon, and Kimberlee C Davison. 2013. Cross-language transfer of early literacy skills: An examination of young learners in a two-way bilingual immersion elementary school. *Reading Psychology*, 34:436–460.

Ben Foley, Josh Arnold, Rolando Coto-Solano, Gautier Durantin, T. Mark Ellison, Daan van Esch, Scott Heath, František Kratochví, Zara Maxwell-Smith, David Nash, Ola Olsson, Mark Richards, Nay San, Hywel Stoakes, Nick Thieberger, and Janet Wiles. 2018. Building speech recognition systems for language documentation: The CoEDL Endangered Language Pipeline and Inference System. In *Proceedings of the 6th International Workshop on Spoken Language Technologies for Under-Resourced Languages*, pages 205–209. ISCA.

Reece George, Keith Nesbitt, Patricia Gillard, and Michael Donovan. 2010. Identifying cultural design requirements for an australian indigenous website. In *Proceedings of the Eleventh Australasian Conference on User Interface-Volume 106*, pages 89–97.

Barbara Glowczewski. 1989. *A topological approach to Australian cosmology and social organisation*, volume 19. Proquest Social Sciences Journals.

Florian Hanke. 2017. *Computer-Supported Cooperative Language Documentation*. Ph.D. thesis, University of Melbourne.

Dianna Hardy, Elizabeth Forest, Zoe McIntosh, Trina Myers, and Janine Gertz. 2016. Moving beyond "just tell me what to code" inducting tertiary ict students into research methods with aboriginal participants via games design. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*, pages 557–561.

John Hatton. 2013. SayMore: Language documentation productivity. Paper presented at the Third International Conference on Language Documentation and Conservation, http://hdl.handle.net/10125/26153.

Mary Hermes and Mel Engman. 2017. Resounding the clarion call: Indigenous language learners and documentation. *Language Documentation and Description*, 14:59–87.

Steven Krauwer. 2003. The Basic Language Resource Kit (BLARK) as the first milestone for the Language Resources Roadmap. *Proceedings of the International Workshop on Speech and Computer*, pages 8–15.

Eric Le Ferrand and Steven Bird. 2020. Enabling interactive transcription in an indigenous community. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3422–3428.

Anany Levitin. 1999. Do we teach the right algorithm design techniques? In *The Proceedings of the Thirtieth SIGCSE Technical Symposium on Computer Science Education*, pages 179–183.

Daria Loi, Christine T Wolf, Jeanette L Blomberg, Raphael Arar, and Margot Brereton. 2019. Co-designing AI futures: Integrating AI ethics, social computing, and design. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion*, pages 381–384.

Celeste Louro and Glenys Collard. 2021. Working together: Sociolinguistic research in urban aboriginal australia. *Journal of Sociolinguistics*.

Teresa Lynch and Shirley Gregor. 2004. User participation in decision support systems development: influencing system outcomes. *European Journal of Information Systems*, 13(4):286–301.

MA Maar, NE Lightfoot, ME Sutherland, RP Strasser, KJ Wilson, CM Lidstone-Jones, DG Graham, R Beaudin, GA Daybutch, BR Dokis, et al. 2011. Thinking outside the box: Aboriginal people's suggestions for conducting health studies with aboriginal communities. *Public Health*, 125(11):747–753.

Felicity Meakins, Jenny Green, and Myfany Turpin. 2018. *Understanding Linguistic Fieldwork*. Routledge.

Raghav Menon, Herman Kamper, Ewald van der Westhuizen, John Quinn, and Thomas Niesler. 2019. Feature exploration for almost zero-resource ASR-free keyword spotting using a multilingual bottleneck extractor and correspondence autoencoders. *Proceedings of Interspeech 2019*, pages 3475–3479.

Alexis Michaud, Oliver Adams, Trevor Cohn, Graham Neubig, and Séverine Guillaume. 2018. Integrating automatic transcription into the language documentation workflow: experiments with Na data and the Persephone Toolkit. *Language Documentation and Conservation*, 12:481–513.

Robyn Ober. 2017. Kapati time: storytelling as a data collection method in indigenous research. *Mystery Train*, 2007.

Dorian Peters, Susan Hansen, Jenny McMullan, Theresa Ardler, Janet Mooney, and Rafael A Calvo. 2018. "participation is not enough" towards indigenous-led co-design. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*, pages 97–101.

Nay San, Martijn Bartelds, Mitchell Browne, Lily Clifford, Fiona Gibson, John Mansfield, David Nash, Jane Simpson, Myfany Turpin, Maria Vollmer, et al. 2021. Leveraging neural representations for facilitating access to untranscribed speech from endangered languages. In *Proceedings of the Automatic Speech Recognition and Understanding Workshop (ASRU)*.

Frank Seifart, Harald Hammarstroöm, Nicholas Evans, and Stephen C. Levinson. 2018. Language documentation twenty-five years on. *Language*, 94:e324–45.

Han Sloetjes, Herman Stehouwer, and Sebastian Drude. 2013. Novel developments in Elan. Paper presented at the Third International Conference on Language Documentation and Conservation, http://hdl.handle.net/10125/26154.

Alessandro Soro, Margot Brereton, Jennyfer Lawrence Taylor, Anita Lee Hong, and Paul Roe. 2017. A cross-cultural noticeboard for a remote community: design, deployment, and evaluation. In *IFIP Conference on Human-Computer Interaction*, pages 399–419. Springer.

Susan Leigh Star. 2007. Living grounded theory: Cognitive and emotional forms of pragmatism. *The Sage Handbook of Grounded Theory*, pages 75–94.

Jennyfer Lawrence Taylor, Wujal Wujal Aboriginal Shire Council, Alessandro Soro, Michael Esteban, Andrew Vallino, Paul Roe, and Margot Brereton. 2020. Crocodile language friend: Tangibles to foster children's language use. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–14.

Mareese Terare and Margot Rawsthorne. 2020. Country is yarning to me: Worldview, health and well-being amongst australian first nations people. *The British Journal of Social Work*, 50(3):944–960.

Bert Vaux and Justin Cooper. 1999. *Introduction to Linguistic Field Methods*. Lincom Europa.

Helen Verran. 2007. The educational value of explicit non-coherence. *Education and Technology: Critical Perspectives, Possible Futures*, pages 101–124.