Fashioning Local Designs from Generic Speech Technologies 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

An increasing number of papers have been addressing issues related to low-resource languages and the transcription bottleneck paradigm. After several years spent in Northern Australia, where some of the strongest Aboriginal languages are spoken, we could observe a gap between the motivations depicted in research contributions in this space and the Northern Australian context. In this paper, we address this gap in research by exploring the potential of speech recognition in an Aboriginal community. We describe our work from training a spoken term detection system to its implementation in an activity with Aboriginal participants. We report here on one side how speech recognition technologies can find their place in an Aboriginal context and, on the other, methodological paths that allowed us to reach better comprehension and engagement from Aboriginal participants.

1 Introduction

A consistent theme in recent NLP research has been *doing more with less* (Wiesner et al., 2022; Gao et al., 2021; Baevski et al., 2021; Schneider et al., 2019; Menon et al., 2019). It is popular to describe new pipelines to solve a wide range of tasks for under-resourced languages (Godard et al., 2018; Anastasopoulos et al., 2018; Settle et al., 2017; Mitra et al., 2016; Lane and Bird, 2019). However, the motivations behind the design of a computational method are not systematically well justified according to the needs of the target speech communities.

The category of under-resourced languages encompasses a wide range of contexts, not simply in terms of the quantity of data available but also in terms of local speech communities' sociolinguistic and political situation (Bird, 2022). Often, the focus has been to generalise a given method across languages, where the proposed system is at the core of the argument instead of the benefits that it could have for the speakers. We could ask whether the same language technology would be equally applicable to Marathi, spoken by millions in a major metropolis, and Miriwoong, with only a few elderly speakers in a remote Australian Aboriginal community (cf. Kuhn, 2022).

Universal solutions dominate NLP: research and results are often provided without taking into account the global situation of the languages involved or the views of the speech communities about the preservation of their language. Instead, it is common to assert that an improvement in Word Error Rate yielded by a given speech recognition system is the answer to the transcription bottleneck and, therefore, the problem of scaling up language documentation (van Esch et al., 2019; Foley et al., 2018).

Most of the world's languages are primarily oral (Ong, 1982; Walsh and Yallop, 1993). Writing is often not a priority, and very few people are skilled in transcribing their language. Written resources often only exist in limited spaces where there is a collaboration between westerners and local communities, such as schools, ranger programs, tourism, and academia. In such cases, writing would seem to primarily serve institutional agendas (cf. Dobrin et al., 2009; Perley, 2012; Nevins, 2013). Accordingly, we must ask ourselves to what extent automatic transcription technologies have a place in research that respects local self-determination. Bird (2022) calls for a *local turn*, for the need to work with local speech communities from the ground up. In other words, outsiders who enter communities with their expertise need to begin with local concerns and local knowledge practices, and only later begin to explore ways in which language technologies can be added into the mix. For example, a local person might want non-indigenous colleagues to learn and use the local language, rather than assuming that all work is conducted in English. We have found that such an approach enlarges the opportunities

Proceedings of the 29th International Conference on Computational Linguistics, pages 4274–4285
October 12–17, 2022.

for collaboration, while simultaneously generating language resources.

This paper extends our previous work on collaborative transcription (Le Ferrand et al., 2022), where the language documentation pipeline we designed failed. We were confronted with different ways of knowing and different expectations in terms of language work. In this work, learning from our past failure, we describe our approach, from the training of a transcription system to the design of collaborative transcription activities with Aboriginal participants. We first describe our speech recognition method based on syllable spotting. We then present the design of the app used that bridges the output of the syllable spotting system to the people, taking into account existing practices. We also explain our method to engage with participants to address their interests in terms of language work. Finally, we detail the application of the proposed transcription activities and discuss the success and flaws of this work.

2 Background

2.1 Decolonising practices

Research contributions around speech processing for low-resource languages have often followed the work of documentary linguistics, where some automation is added to support manual annotations (Adams et al., 2018; Godard et al., 2018; Foley et al., 2018). The 7000+ world languages are often mentioned and language technologies appear as a way to prevent their loss (Adda et al., 2016; Duong, 2017; Jimerson and Prud'hommeaux, 2018). Special workshops like the zero resource challenge[1] and the introduction of a surprise language have pushed in this direction allowing the creation of computational solutions that bypass the need of the speech communities of language experts. Recent studies have also shown that the languages (Schwartz, 2022) or the speech communities (Caselli et al., 2021) are rarely ate the core of the argument in the ACL anthology's publications.

Documentary linguistics is often the preliminary step of language description and analysis (Hanke, 2017). Documentation and description communicate with each other to allow western scholars to have a better comprehension of Indigenous languages. There are no clear benefits for the speech community, and extra work needs to be provided to share the benefits of a research project (Chelliah and De Reuse, 2010). The NHMRC Guidelines[2] for Ethical Conduct in research with Aboriginal and Torres Strait Islander Peoples and Communities set out principles of equity and reciprocity, where the outcome of the research should benefit both parties. Recent research practices, including documentary linguistics, started to fully commit to these standards by adopting a community-based approach (e.g. Rodríguez Louro and Collard, 2021; Ryder et al., 2021; Taylor et al., 2020). Community-based research has the community at its core and is meant to be conducted for and with the participation of community members (Rice, 2011).

2.2 Community-based projects

Community-based research around software design is a small but growing area. Projects have been based on research Human-Computer Interaction (HCI) or NLP from a language learning perspective. On the HCI side, research has contributed to responding to local issues by designing tools in collaboration with the community (Soro et al., 2017; Hardy et al., 2016; Leong et al., 2019). Cross-cultural collaboration is challenging. From this kind of project have also emerged engagement methods to facilitate the conversation with Indigenous communities about technology design (Zaman et al., 2016; Taylor et al., 2020). On the NLP side, the research contributions have been language-specific or bounded to a specific context. For instance, Pine et al. (2022) have described speech synthesis systems in several Indigenous Canadian languages responding to a call from the language learners. Projects that did not initially have a community-based component sometimes ended up serving community-based projects. Uí Dhonnchadha and Van Genabith (2006) for instance, created a POS tagger for gaellig Irish. The system has been then incorporated into an Irish learning game (Xu et al., 2022). In either case, the majority of the work done in this area is based on writing (e.g. Lane and Bird, 2019; Schwartz et al., 2019; Finn et al., 2022). The only speech-based projects are around speech synthesis (Harrigan et al., 2019; Pine et al., 2022). Speech recognition seems to be rarely involved in community-based projects.

[1] https://www.zerospeech.com/

[2] https://www.nhmrc.gov.au/about-us/resources/ethical-conduct-research-aboriginal-and-torres-strait-islander-peoples-and-communities#block-views-block-file-attachments-content-block-1

2.3 Context

Our work is grounded in Bininj country in West Arnhem, Northern Territory in the Australian Top End. Bininj country is part of the Indigenous Protected Area of Arnhem Land where the land and sea are managed by Aboriginal groups.[3] The main language of communication is Kunwinjku (ISO gup) which is spoken by approximately 2500 people (Marley, 2021). There is a standard orthography that has been introduced by linguists but it is not widely used by the members of the community.

The first and second authors have several years of experience with the Bininj community, have some expertise in Kunwinjku, the local language, and have both been adopted by Traditional owners of the land. In this case, adoption means the attribution of a *skin name* that connects an individual to the rest of the community (cf. Christie, 2008, p.35).

2.4 Learning from failure

This work is the continuation of Le Ferrand et al. (2022). We previously designed a spoken term detection prototype to detect whole words in untranscribed speech collections in Kunwinjku. We then used an app to bridge the output of our prototype to the people to allow local communities to verify the guesses of our system and therefore be part of transcription works. We faced many challenges that we tried to build on in this work.

This previous work focused on the collection of data to enhance the performance of the system. The design ended up being irrelevant and redundant for the participants. From here we realised the need for further discussion with the community to set up activities that are relevant to their agenda, interests and practices.

The app presented displayed only four buttons: one to play the query, one to play the utterance and two to give a feedback on whether the query has been spotted in the utterance or not. While testing the app, we realised how the audio files extracted from their contexts were confusing for the participants. Besides, the fact of validating system guesses in random utterances was disconnected from the idea of transcription which led most participants to overthink the task.

In projects around cross-cultural technology design, shallow information is provided about the extent of the collaboration and the challenges encountered. Yet, studies have described ways of knowing in Indigenous communities that differs from the western approach to knowledge (Descola, 2005; Foley, 2003). Such differences appear as the main reason behind the failed attempt of app design where the proposed task lose all meaning in Bininj context.

From our first failed attempt, the challenges were two-fold. We first needed to figure out a way to solve the comprehension issues we have faced. Then, we needed to improve the relevance of this work for Bininj participants. The key was to find out how to design transcription technologies based on existing practices. From the language learning sessions we had with some of our Aboriginal collaborators, we noticed, for instance, how they teach us breaking down words into syllables to decompose the pronunciation of a given item. This led us to think about replacing word spotting with syllable spotting, allowing participants to reproduce their word decomposition strategy to build up the transcription from the syllables spotted. From here, the focus needed to be given on incorporating this transcription strategy into an activity that matters to the people.

3 Transcription by syllables

3.1 Data

To build the system, we are using a corpus in Kunwinjku built from several sources. The training and validation sets consist of 35.45 min and 7.39 min respectively of spontaneous speech made of guided tours of Aboriginal towns and utterances for language description purposes. Two different sets are used for testing: one set of 19.43 min of spontaneous utterances and one set of 4.43 min of elicited words recorded in isolation.

To build our list of valid syllables, we used a word list built from the Bible in Kunwinjku. We then applied on each word syllable segmentation rules resulting in a set of 584 unique syllables with relative frequency values associated.

3.2 Experimental setup

Le Ferrand et al. (2021) introduced a method of spoken term detection for very low-resource languages based on phone recognition. Their method is based on Allosaurus (Li et al., 2020), a universal phone recognizer. We preferred this method in this work due to its flexibility in terms of query selection and its speed compared to Dynamic Time Warp-

[3] https://www.awe.gov.au/agriculture-land/land/indigenous-protected-areas

ing, which is usually used for very low-resource languages.

We first trained the phone recognizer using our train set and generated confusion matrix from the validation and test sets. A confusion matrix consists of a phone transcription and the top k (we use k=5) most likely alternatives per phone with a likelihood score associated. To spot syllables, we expressed the syllables extracted from the bible as a finite state automaton after conversion from graphs to phones and explored every possible path in the phone matrix that corresponded to a valid transition in the lexicon. Ultimately we extracted the resulting syllables with the mean of the phones' scores that are used as a likelihood measure to filter the syllables spotted based on a threshold *T*.

To increase the accuracy of the method of Le Ferrand et al. (2021) which only relies on the likelihood scores output by allosaurus, we used the frequency information in our syllable list to more precisely select our candidates. To do so, we average the likelihood score L_s of a detected syllable with its unigram probability P_s weighted with a constant α as:

$$L_s + \alpha P_s \quad (1)$$

We then optimised, on the validation set, α varying a range of values between 0 and 10 with a 0.1 step and a syllable detection threshold T between 0 and 10 with a step at 0.01. We then spotted syllables on the test set with the parameters which provided the best F-score on validation. We also report results without the frequency where only the threshold T is optimized on the validation set.

3.3 Experimental results

Our best results on the validation set have been obtained with $T = 0.39$ when unigram probability is added. For our baseline without unigram probability, the best threshold has been obtained with $T = 0.35$ We report the results in Table 1.

We can see here that the frequency information has an impact on the overall performances in both scenarios with an F-score nearly 4 points higher in the results with frequency. Better performances are obtained on the test set made of utterances. Two elements can explain it. First, the phone recognition model has been trained on similar data to the utterance test set which leads to better phone recognition performances. Then, the chance of a given syllable being pronounced several times is higher in longer utterances which means that it has higher chance to be spotted.

Results with likelihood score alone			
Sets	Recall	Precision	F-score
Words	41.71%	24.40%	30.79%
Utterances	47.21%	36.26%	41.02%
+ unigram probability			
Sets	Recall	Precision	F-score
Words	43.08%	28.09%	34.23%
Utterances	46.56%	41.50%	43.88%

Table 1: Experimental results (syllable spotting) on the two test sets

4 App design

4.1 Prototype

We designed a simple interface to display the syllables from our spoken term detection systems to our participants (see Figure 1). Our goal here was not to design a final product but to present a simple interface that works well enough to see if the proposed syllable concatenation mechanism makes sense from a Bininj perspective. We bridged the output of the system to a transcription interface by creating one button per syllable spotted for a given audio recording. The buttons display the orthography of the syllables spotted. They play the corresponding pronunciation when clicked. There is one *play* button to play the audio to transcribe and one text area with an associated *play* button to look for syllables that have not been spotted. The user needs to use the keyboard to make guesses on missing syllables and needs to click on the *play* button to check the pronunciation of their guesses.

Figure 1: Preliminary version of the app

We organised a testing session with one participant in Gunbalanya: IG, a 25 year old local artist and tour guide. We spotted syllables in a 3.35min recording made of elicited speech of Bible stories. Because of the quality of the audio, most of the syllables were correctly spotted. We explained to IG that we wanted to write down Kunwinjku and we needed his help to spell the words.

IG rapidly understood the task and started point-

ing syllables on the screen while we were writing with pen and paper IG's feedback. He clicked several times on the different syllables displayed and progressively gave feedback. When a syllable was not spotted, he could with some hesitancy, write with the keyboard syllables guesses in the text area.

The main observations made during the pilot study were IG's quick comprehension of the task, his hesitancy while using the keyboard and his confidence while reporting the orthography. At the end of the activity, he told us that he was expecting the text area to produce a new syllable button he could use.

4.2 Design and features

The quick comprehension of IG showed the potential of the proposed transcription mechanism which made us pursue this direction. Based on the first trial, we designed a proper transcription interface based on syllable spotting (see Figure 2). The core of the interface was the same that our first trial: we have a play button on the top of the screen playing the target audio to transcribe. We have one button per detected syllable associated with a wav file containing their pronunciation. The syllables can be dragged and dropped to the black box at the bottom of the screen. The user can listen to the final concatenation of the syllables with the associated play button and validate the transcription created with a thumb up button.

We needed to find a way to allow the user to add undetected syllables manually. To do so, we initially added a side menu accessible through a plus button on the side of the screen. The menu consisted of a scrolling list that contained the 584 syllables. We added a text area at the top of the list that allowed the user to retrieve a syllable from its first letters (see Figure 3). Following the principle of the regular syllable button, the user could click on the syllable to hear the pronunciation and click on the associated plus button once their choice was made. The syllable was then added as a regular syllable button. To avoid the use of the keyboard, we changed this syllable search mechanism by removing the text area and by replacing the list of syllables with expandable sub-lists labelled with the first graph[4] of the syllables it contains (see Figure 4). The user can then search for a syllable by expanding the lists and select a syllable by listening to it and clicking on the associated plus button. The app and databases were stored in a laptop accessed remotely by a tablet with wifi.

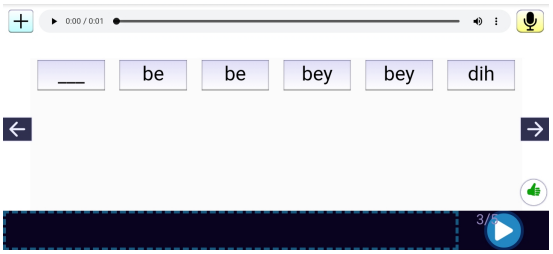

Figure 2: Final version of the app

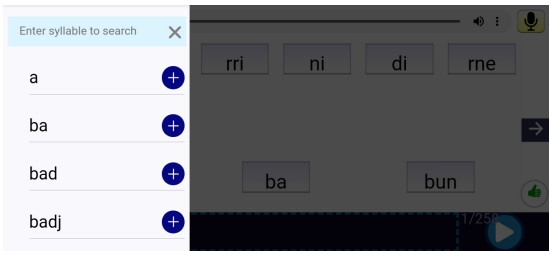

Figure 3: Initial syllable search mechanism

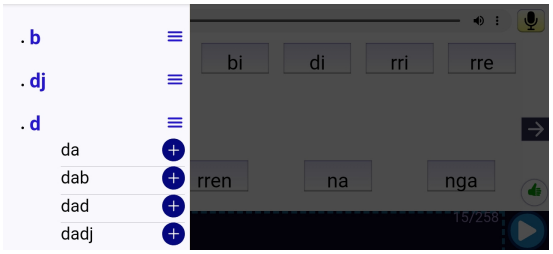

Figure 4: Updated syllable search mechanism

5 User testing

Due to Covid-19 restrictions, no trips to remote communities were possible. However, we have been able to work individually with Kunwinjku speakers in transit in Darwin at the university. We incorporated our syllable spotting based transcription task in a more global resource creation workflow. We could test it with two participants from Bininj country. In order to engage with the participants, we organised the testing phase in two sections. In the first one, we discussed and elicited knowledge about topics of interest based on previous conversations, in the second, we used the interface to transcribe the knowledge recorded. Therefore, besides the focus given to the design of the app and spoken term detection system, time of this

[4]We are not talking in terms of individual letter but graph or group of graphs that correspond to a single phone in Kunwinjku

project has been dedicated to the study of cultural elements to enable more efficient collaboration.

5.1 Activity description

Elicitation of knowledge

> *Ngabenbekken nahni wurdwurd nawu kabirrihre minj Kundebi kabirrikarrme. Burrkyak. Kabirridjalngeybun. Minj kabirridebikarren, burrkyak.*
>
> "I hear these children going about – they don't have Kundebi. No. They just use people's names. They don't use Kundebi with each other, no." (Etherington, 2006)

Language shift is not a new phenomenon. Language variation in Kunwinjku has been the subject of recent research (Marley, 2021) and has been one of the concerns raised by Bininj Elders. *Kundebi* specifically has been described as a language feature that the community is proud of and that is being progressively lost by the young generation (Garde, 2013; Etherington, 2006). It has also been mentioned in the same terms by some Elders during some of our fieldtrips. *Kundebi* refers to the way a speaker A refers to an individual C while talking to an addressee B. For example, a speaker A is talking to their elder sister's child B about their elder sister C. A is usually referring to B using the term *djedje* "nephew" and to C using the term *yabok* "sister". Listener B however usually refers to C using the term *morlah* "mother's elder sister". The kundebi term *berlunghkowarre* is then used to summarize these three relationships and could be translated as "my sister, your mother's elder sister, you are my sister's child" (Garde, 2013).

In order to respond to people's priority in terms of language work, we have decided to first focus the activity on the creation of written resources around *Kundebi*. To do so, while working with a Kunwinjku speaker, we would talk about common acquaintances, identify the way we both would refer to them and then identify and record the corresponding *Kundebi* terms. We used an activity sheet (see Figure 5) to draw the relationships we wanted to elicit (for instance, E for first author, G for the participant and J for the person we are talking about). The recording is directly stored on our laptop. The speed of the pipeline, described in Section 3.2, also allowed us to directly spot the syllables in the audio. Some of our participants expressed the fact that they were not confident with

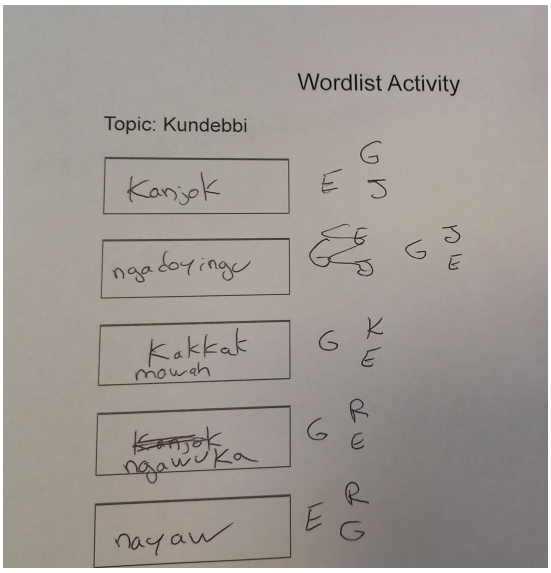

Figure 5: activity sheet filled

Kundebi and would feel more comfortable talking about *Kunbalak*. *Kunbalak* is a sub-language used for forbidden relationships to show respect. It is identical to regular Kunwinjku syntactically but would use different lexical items. For instance *Birriwam* "they went" becomes *birridokang* in *Kunbalak* (Manakgu, 1996). To elicit *Kunbalak* we would just ask for the conversion of regular Kunwinjku terms.

Use of the app

After recording a few terms with a speaker, we presented the transcription interface to them. The terms previously recorded and the syllables spotted have been automatically loaded into the app database. After showing the interface's different features to the participants, we asked them to drag and drop the syllables to build the transcription of the previously recorded terms. After actively working around Kunwinjku and building expertise about the proper way to write the language through the years, we let the participants use their own expertise on what they think is the orthography without questioning their authority.

5.2 Fieldwork

We tested our pipeline with two participants. JB (30s) and GB (30s).

We could present our activity to JB on three different occasions. We could identify and record some *kundebi* terms during the first trial. The activity has then been interrupted by upset child. During this first trial, she briefly started to point syllables on the screen without properly using the app.

She told us afterwards that the *kundebi* terms we recorded should be double-checked by an Elder, and she would feel more confident talking about *Kunbalak* instead. During the second and third trials, we could easily identify and record some *Kunbalak* terms. While using the app, we faced minor technical difficulties with the manual syllable addition feature. However, JB could take control of the tablet to transcribe some of the terms recorded. One of the issues we faced was the playback of syllables that include a glottal stop which was hard to identify in syllables in isolation (the difference between *ma* and *mah*, for instance). The activity was trialled with the first version of the syllable search (see Figure 3). The keyboard generated by the text area would take most of the space on the screen. JB needed to ask for our support to know how to proceed. At the end of the second trial, while no instruction had been explicitly given, she started to drag and drop the syllables available on the screen to explore the different words that are possible with them. We asked about her thoughts about the activity, and she responded that she liked it and would like to get more confident in writing in Kunwok and download the app later.

We could test the activity with GB, the second participant. We first recorded a few *Kundebi* terms. We wrote on paper the relationship to elicit, which made him understand the activity was about constructing a word list. After recording a few terms, we gave GB the tablet and asked him to transcribe the words. For each term, he listened to the audio first and pressed the syllable displayed on the screen. He was able to add new syllables manually without too much difficulty. For one particular term: *nayaw*, we discussed rather the term should be written *nayaw* or *nayawu*. While listening for a given syllable, he sometimes asked for confirmation about what he heard (for instance, "Is this *ka*?"). We discussed his thought about the task at the end of the activity. He showed enthusiasm about the incremental construction of the transcription. During the activity, he rephrased the syllable concatenation process by "putting pieces of language together".

No more participants were available for the time for this project. However, to sustain this work in the future, we deployed it in a laptop to be brought to the community by future scholars or language workers, as soon as COVID-19 restrictions are eased.

Figure 6: Picture of a participant using the app

6 Discussion and Limitations

The design and testing of the activity have shown promising results among a few participants, which gave us a glance at the potential of syllable spotting for the design of language related activity for Aboriginal people.

Syllable Spotting: It has been shown in the literature that traditional ASR is hardly applicable to Aboriginal languages due to the lack of resources available to train robust systems. Sub-word detection has been seen as a way to avoid out-of-vocabulary (Szoke et al., 2008; Parlak and Saraclar, 2008; Van Heerden et al., 2017) and, in our case, to allow a denser transcription than word spotting specifically for a polysynthetic language like Kunwinjku. Adding information on frequency, not surprisingly, allowed us to boost our performance (F-score) from 40% to nearly 44% for the syllables displayed on the screen for a given utterance.

Enabling mutual comprehension: Our main objective, starting from our previous work, was to enable a better comprehension in our cross-cultural setting. Part of this process consisted of getting familiar with cultural components that have been raised by the community (namely, *Kundebi* and *Kunbalak*). This also consisted of finding methods to trigger a conversation about these topics. For the rest, strategies have been found to help the participants to understand our contribution is this work. For instance, the support of the activity sheet made clear that the ultimate goal of the activity was to build a word list. Then the syllable concatenation mechanism allowed the participants to leverage existing language patterns from the aural space into writing.

Aligning agendas: Asking the participants about traditional knowledge allowed them to di-

rectly use their expertise and navigate in familiar territory. Talking about *Kundebi* and *Kunbalak* gave a sense of clarity regarding our function in Aboriginal land because of the continuation between previous conversations and the current activity. Yet the extent of our contribution being seen as beneficial for Bininj people from a language preservation perspective is still unclear. Writing in language is not a traditional practice in this community. People are often literate in English but not in their language. We then needed to find a space where the orthography made sense (Lewis and Simons, 2016). Documents written in Kunwinjku exist in Bininj country through the ranger program, the schools or in facilities where exists an interaction between Bininj and westerners (art centres, clinics, etc.). While we thought that the proposed activity could enable the continuation of the creation of these resources by Aboriginal participants, the proposed app has probably mainly been seen as a way to enhance writing skills.

App design: There were two main challenges related to the design of the app. The first one was enabling syllable concatenation, prioritising information from the oral space. Then we needed to efficiently retrieve syllables that had not been spotted. The first challenge was easily solved by the syllable playback features possible with the progressive collection of syllables throughout this project. Then we designed a basic search mechanism. The first search mechanism to add new syllables relied on the keyboard, which we knew was problematic (cf. Section 4.1). We believe that the new design would lead to better efficiency, but it could not be properly tested.

Activity flaws: The lack of good quality data available in Kunwinjku did not allow us to build a robust speech synthesis system that would have been relevant to the interface. Instead, we recorded in isolation syllables which sometimes lacked clarity. While ultimately, some of the most common syllables have been recorded by a native speaker, many were still pronounced by the first author, whose pronunciation might not be accurate. For instance, in the pilot study, while writing the word *djurra* (IPA djura) "paper", first author's pronunciation of the syllable *rra* has not been accepted by IG and selected instead "*da*" which was closer to the pronunciation of the word according to him. The case of the glottal stop has also been mentioned as a challenge in the literature (Wigglesworth et al., 2021). The glottal stops included in some syllables were not clearly audible out of context, which made them hard to differentiate from similar syllables without glottal stops (*ma* and *mah*, for instance).

Limitations: There is a limited number of Kunwinjku speakers, and recruiting a large number of participants for such work was not easy. The current pandemic did not facilitate our work, and we know that it is hard to draw final conclusions with activities conducted with only three participants. Further research needs to be done, including proper testing in Bininj country to consolidate our observations. The activity setup was also grounded for JB and GB in an academic environment with access to facilities that we do not necessarily have access to in remote locations (access to the internet, workplaces etc...). Besides, we can ask ourselves about the sustainability of such a work grounded in an interaction between Aboriginal participants and scholars in a very controlled environment. To be sure that our methods can be used in the long term, we imagine setting up a remote server to enable remote access on tablets so that people can keep interacting with the app without outside intervention.

7 Conclusion

Generic speech recognition methods for under-resourced languages offer the potential to support small speech communities. Yet the translation of such methods into community-based projects is rare. We have presented a study on the creation and testing of a syllable spotting-based transcription interface to enable the creation of written resources by the members of an Aboriginal community in the Australian Top End. Based on the challenges encountered in previous work, we went from word spotting to syllable spotting to reach a denser transcription and enabled a transcription method closer to existing practices. With the help of collaborators, we designed a transcription interface that allowed the users to build the transcription of given audio using the syllable spotted by our system. We reported the testing of the app with three participants at different stages of development, including lessons learnt from their interaction with the transcription activity and the app design.

Research guidelines push scholars to decolonise their practices and to go towards self-determination. Yet the translation of guidelines to real-life applications is unclear, specifically in cross-cultural collaborations with different ways of knowing. This

work allowed us to highlight methodological paths that improved the engagement and comprehension of the participants. The activity sheet, for instance, made clear that the activity was about creating a wordlist which was not necessarily clear based on our explanation. Dividing the activity between an elicitation part and a transcription part allowed us to hook the interest of the participants with a task they were familiar with and allowed us to clarify the context of our work in contrast to the sparse transcription of random sentences explored previously (Le Ferrand et al., 2022). All participants frequently used the playback of the syllables in isolation and their concatenation, confirming its engaging aspect.

Documentary linguistics has often been undertaken by non-indigenous linguists where the collaboration with the community did not go further than the collection of spoken data (First Languages Australia, 2014). In this work, we initially wanted to counterbalance these practices by enabling community-based language documentation. Yet keeping a language strong does not need to be about language documentation, and Bininj people who took part in this work did not seem to buy into documentary linguistics practices. Instead, they seemed to see the interface as a literacy learning tool. Keeping language strong is seen as building capabilities instead of creating and storing language material. Community-based implies an active role of the community in the work we conducted, and following their view in terms of language work is then crucial. The cross-cultural challenges we encountered required extra work to enable a common ground we could build on. Now that comprehension issues are solved, that we have a better comprehension of people agenda and COVID-19 restrictions start to be eased, more iteration can happen to allow the community to take control of the design of the proposed tool to better fit their agenda and practices.

8 Acknowledgements

We are grateful to the Bininj people of Northern Australia for the opportunity to work in their community, and particularly to artists at Injalak Arts and Craft (Gunbalanya), so as the Bininj and Daluk from Mamadawerre and Kabulwarnamyo who spent time helping us in the university facilities. Our thanks to several anonymous reviewers for helpful feedback on earlier versions of this paper. The final version of the transcription building app has been co-designed by Dr. Cat Kutay, Lecturer at Charles Darwin University and Melko Nguyen as part of her final master's project.

This research was covered by a research permit from the Northern Land Council, ethics approved from CDU and was supported by the Australian government through a PhD scholarship, and grants from the Australian Research Council and the Indigenous Language and Arts Program. All the participants have been paid at the regular rate for Aboriginal people consultancy.

References

Oliver Adams, Trevor Cohn, Graham Neubig, Hilaria Cruz, Steven Bird, and Alexis Michaud. 2018. Evaluating phonemic transcription of low-resource tonal languages for language documentation. In *LREC 2018 (Language Resources and Evaluation Conference)*, pages 3356–3365.

Gilles Adda, Sebastian Stüker, Martine Adda-Decker, Odette Ambouroue, Laurent Besacier, David Blachon, Hélene Bonneau-Maynard, Pierre Godard, Fatima Hamlaoui, Dmitry Idiatov, et al. 2016. Breaking the unwritten language barrier: The bulb project. *Procedia Computer Science*, 81:8–14.

Antonios Anastasopoulos, Marika Lekakou, Josep Quer, Eleni Zimianiti, Justin DeBenedetto, and David Chiang. 2018. Part-of-speech tagging on an endangered language: a parallel Griko-Italian resource. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2529–2539.

Alexei Baevski, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. 2021. Unsupervised speech recognition. *Advances in Neural Information Processing Systems*, 34:27826–27839.

Steven Bird. 2022. Local languages, contact languages, and other high-resource scenarios. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 7817–7829.

Tommaso Caselli, Roberto Cibin, Costanza Conforti, Enrique Encinas, and Maurizio Teli. 2021. Guiding principles for participatory design-inspired natural language processing. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 27–35.

Shobhana Chelliah and Willem De Reuse. 2010. *Handbook of Descriptive Linguistic Fieldwork*. Springer Science & Business Media.

Michael Christie. 2008. Yolngu studies: A case study of aboriginal community engagement. *Gateways: International Journal of Community Research and Engagement*, 1:31–47.

Philippe Descola. 2005. *Par-delà nature et culture*, volume 1. Gallimard Paris.

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.

Long Duong. 2017. *Natural Language Processing for Resource-poor Languages*. Ph.D. thesis, University of Melbourne.

Steve Etherington. 2006. *Learning to be Kunwinjku: Kunwinjku People Discuss their Pedagogy*. Ph.D. thesis, Charles Darwin University.

Aoife Finn, Peter-Lucas Jones, Keoni Mahelona, Suzanne Duncan, and Gianna Leoni. 2022. Developing a part-of-speech tagger for te reo māori. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 93–98.

First Languages Australia. 2014. Angkety Map: Digital resource report. Technical report. https://www.firstlanguages.org.au/images/fla-angkety-map.pdf; accessed Oct 2021.

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.

Dennis Foley. 2003. Indigenous epistemology and indigenous standpoint theory. *Social Alternatives*, 22(1):44–52.

Tianyu Gao, Adam Fisch, and Danqi Chen. 2021. Making pre-trained language models better few-shot learners. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 3816–3830.

Murray Garde. 2013. *Culture, Interaction and Person Reference in an Australian Language: An Ethnography of Bininj Gunwok Communication*, volume 11. John Benjamins Publishing.

Pierre Godard, Marcely Zanon Boito, Lucas Ondel, Alexandre Berard, François Yvon, Aline Villavicencio, and Laurent Besacier. 2018. Unsupervised word segmentation from speech with attention. In *Proceedings of Interspeech 2018*, pages 2678–2682.

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.

Atticus Harrigan, Timothy Mills, and Antti Arppe. 2019. A preliminary plains cree speech synthesizer. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*, volume 1, pages 64–73.

Robbie Jimerson and Emily Prud'hommeaux. 2018. ASR for documenting acutely under-resourced indigenous languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pages 4161–4166.

Roland Kuhn. 2022. The Indigenous Languages Technology Project at the National Research Council of Canada, and its context. In *Language Technologies and Language Diversity*, pages 85–104. Linguapax International.

William Lane and Steven Bird. 2019. Towards a robust morphological analyzer for Kunwinjku. In *Proceedings of the The 17th Annual Workshop of the Australasian Language Technology Association*, pages 1–9.

Éric Le Ferrand, Steven Bird, and Laurent Besacier. 2021. Phone based keyword spotting for transcribing very low resource languages. In *Proceedings of the The 19th Annual Workshop of the Australasian Language Technology Association*, pages 79–86.

Éric Le Ferrand, Steven Bird, and Laurent Besacier. 2022. Learning from failure: Data capture in an australian aboriginal community. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 4988–4998.

Tuck Wah Leong, Christopher Lawrence, and Greg Wadley. 2019. Designing for diversity in aboriginal australia: Insights from a national technology project. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, pages 418–422.

M Paul Lewis and Gary F Simons. 2016. *Sustaining Language Use*. SIL International Publications.

Xinjian Li, Siddharth Dalmia, Juncheng Li, Matthew Lee, Patrick Littell, Jiali Yao, Antonios Anastasopoulos, David R Mortensen, Graham Neubig, Alan W Black, et al. 2020. Universal phone recognition with a multilingual allophone system. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8249–8253. IEEE.

Andrew Manakgu. 1996. Kunbalak. *Stories for Kunwinjku young people in mother-inlaw language, ordinary Kunwinjku and English. Kunbarllanjnja (Oenpelli): Kunwinjku Language Centre*.

Alexandra Marley. 2021. "I speak my language my way!"—young people's kunwok. *Languages*, 6(2):88.

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.

Vikramjit Mitra, Andreas Kathol, Jonathan D Amith, and Rey Castillo García. 2016. Automatic speech transcription for low-resource languages-the case of Yoloxóchitl Mixtec (Mexico). In *Proceedings of Interspeech 2016*, pages 3076–3080.

M. Eleanor Nevins. 2013. *Lessons from Fort Apache: Beyond Language Endangerment and Maintenance*. Wiley.

Walter Ong. 1982. *Orality and Literacy: The Technologizing of the Word*. Routledge.

Siddika Parlak and Murat Saraclar. 2008. Spoken term detection for Turkish broadcast news. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5244–5247. IEEE.

Bernard Perley. 2012. Zombie linguistics: Experts, endangered languages and the curse of undead voices. *Anthropological Forum*, 22:133–149.

Aidan Pine, Dan Wells, Nathan Brinklow, Patrick Littell, and Korin Richmond. 2022. Requirements and motivations of low-resource speech synthesis for language revitalization. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7346–7359.

Keren Rice. 2011. Documentary linguistics and community relations. *Language Documentation and Conservation*, 5:187–207.

Celeste Rodríguez Louro and Glenys Collard. 2021. Working together: Sociolinguistic research in urban Aboriginal Australia. *Journal of Sociolinguistics*, 25:785–807.

Courtney Ryder, Tamara Mackean, Kate Hunter, Julieann Coombes, Andrew JA Holland, and Rebecca Ivers. 2021. Yarning up about out-of-pocket healthcare expenditure in burns with aboriginal families. *Australian and New Zealand journal of public health*, 45(2):138–142.

Steffen Schneider, Alexei Baevski, Ronan Collobert, and Michael Auli. 2019. wav2vec: Unsupervised pre-training for speech recognition. In *Proceedings of Interspeech 2019*, pages 3465–3469.

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.

Lane Schwartz, Emily Chen, Benjamin Hunt, and Sylvia LR Schreiner. 2019. Bootstrapping a neural morphological analyzer for st. lawrence island yupik from a finite-state transducer. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*, volume 1, pages 87–96.

Shane Settle, Keith Levin, Herman Kamper, and Karen Livescu. 2017. Query-by-example search with discriminative neural acoustic word embeddings. *Proceedings of Interspeech 2017*, pages 2874–2878.

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.

Igor Szoke, Lukás Burget, Jan Cernocky, and Michal Fapso. 2008. Sub-word modeling of out of vocabulary words in spoken term detection. In *2008 IEEE Spoken Language Technology Workshop*, pages 273–276. IEEE.

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.

E. Uí Dhonnchadha and J. Van Genabith. 2006. A part-of-speech tagger for Irish using finite-state morphology and constraint grammar disambiguation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, pages 2241–2244, Genoa, Italy.

Daan van Esch, Ben Foley, and Nay San. 2019. Future directions in technological support for language documentation. In *Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 14–22.

Charl Van Heerden, Damianos Karakos, Karthik Narasimhan, Marelie Davel, and Richard Schwartz. 2017. Constructing sub-word units for spoken term detection. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5780–5784. IEEE.

Michael Walsh and Colin Yallop, editors. 1993. *Language and Culture in Aboriginal Australia*. Aboriginal Studies Press.

Matthew Wiesner, Desh Raj, and Sanjeev Khudanpur. 2022. Injecting text and cross-lingual supervision in few-shot learning from self-supervised models. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8597–8601. IEEE.

Gillian Wigglesworth, Melanie Wilkinson, Yalmay Yunupingu, Robyn Beecham, and Jake Stockley. 2021. Interdisciplinary and intercultural development of an early literacy app in Dhuwaya. *Languages*, 6:106.

Liang Xu, Elaine Uí Dhonnchadha, and Monica Ward. 2022. Faoi gheasa an adaptive game for irish language learning. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 133–138.

Tariq Zaman, Heike Winschiers-Theophilus, Franklin George, Alvin Yeo Wee, Hasnain Falak, and Naska Goagoses. 2016. Using sketches to communicate interaction protocols of an indigenous community. In *Proceedings of the 14th Participatory Design Conference: Short Papers, Interactive Exhibitions, Workshops-Volume 2*, pages 13–16.