Collecting Bilingual Audio in Remote Indigenous Communities

Steven Bird
Dept of Computing and Information Systems, University of Melbourne; Linguistic Data Consortium, University of Pennsylvania

Lauren Gawne
Department of Linguistics and Multilingual Studies, Nanyang Technological University, Singapore

Katie Gelbart
School of Oriental and African Studies, University of London

Isaac McAlister
Department of Languages, Literatures, and Cultures, University of Massachusetts, Amherst

Abstract

Most of the world's languages are under-resourced, and most under-resourced languages lack a writing system and literary tradition. As these languages fall out of use, we lose important sources of data that contribute to our understanding of human language. The first, urgent step is to collect and orally translate a large quantity of spoken language. This can be digitally archived and later transcribed, annotated, and subjected to the full range of speech and language processing tasks, at any time in future. We have been investigating a mobile application for recording and translating unwritten languages. We visited indigenous communities in Brazil and Nepal and taught people to use smartphones for recording spoken language and for orally interpreting it into the national language, and collected bilingual phrase-aligned speech recordings. In spite of several technical and social issues, we found that the technology enabled an effective workflow for speech data collection. Based on this experience, we argue that the use of special-purpose software on smartphones is an effective and scalable method for large-scale collection of bilingual audio, and ultimately bilingual text, for languages spoken in remote indigenous communities.

1 Introduction

Past the top one to three hundred economically significant languages, there are few prospects for resourcing the production of annotated corpora. Advances in natural language processing have relied on such corpora – including treebanks and wordnets – though they are expensive to produce and depend on substantial prior scholarship on the language. An alternative is to collect bilingual aligned text, relating a low-resource language to a high-resource language, and then infer lexical and syntactic information from the high-resource language via alignments (Abney and Bird, 2010; Baldwin et al., 2010; Palmer et al., 2010; Das and Petrov, 2011).

This approach only works for written languages. Over half the world's languages lack a literary tradition. In some cases they have a writing system, but it is not in regular use and so these languages remain effectively unwritten. Collecting data for unwritten languages necessarily involves speech.

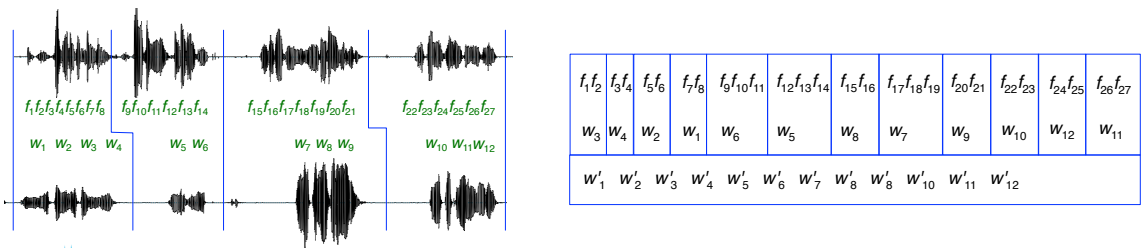


Figure 1: The Vision: phrase-aligned bilingual audio from an unwritten language to a language of wider communication, along with extracted acoustic features and crowdsourced transcription (left); interlinear glossed text with word segmentation, word-level glosses, and sentence-level translations (right).

Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 1015–1024, Dublin, Ireland, August 23-29 2014.

While the physical isolation of these languages presents a logistical challenge, it is still possible to collect hundreds of hours of speech using mobile devices (de Vries et al., 2014). Furthermore, there are promising signs that natural language processing methods and speech processing methods can be integrated (Zhang et al., 2004; Dredze et al., 2010; Vu et al., 2011; Siniscalchi et al., 2012; Lee and Glass, 2012). Thus, the challenge is to collect substantial quantities of bilingual aligned audio, transcribe the translations, extract phonetic features from the source language and, ultimately, produce bilingual aligned text (see Figure 1).

We have chosen to focus on endangered languages because of the interesting and difficult challenges that are faced in collecting data. However, the resource problem exists even for vital languages having large speaker populations. For example, Shanghainese (Wu) is spoken by 77 million people in China, but is almost never written down because written Chinese is based on Mandarin; Oromo is spoken by 17 million people in Ethiopia, but few of its speakers know how to write it. Such languages are collectively spoken by billions, yet remain seriously under-resourced. Thus, while our focus is on endangered languages, the approach applies to under-resourced languages in general.

Several other promising approaches to the problems raised by endangered languages are being actively pursued in computational linguistics, however they typically focus on written language with annotations, often with the goal of making optimal use of human expertise (Probst et al., 2002; Levin et al., 2006; Clark et al., 2008; Palmer et al., 2010; Bender et al., 2012; Beale, 2012; Bender et al., 2013). The research reported here is unique in its focus on securing spoken language data in a form and on a scale that will be usable even once the languages in question are no longer spoken.

This paper explores ways that networked smartphones can be used for collecting bilingual aligned audio. We have used a prototype Android application for collecting audio and phrase-aligned translations (or consecutive interpretations). We took a set of phones to villages in Brazil and Nepal, and worked with languages Tembé, Nhengatu and Kagate. We visited at the invitation of the local communities and collaborated closely with them in each stage of the process, including setting the goals and agreeing on the form of dissemination, cf. (Rice, 2011). We compiled small collections of recorded texts and translations in each language.

We describe and evaluate this novel resource-creation activity, and argue that it can be used effectively for large-scale collection of bilingual aligned audio. This paper is organised as follows. In section 2, we give an overview of the mobile software. The next three sections report the activities in the three communities. We reflect on the work in section 6.

2 Mobile applications for recording and translating endangered languages

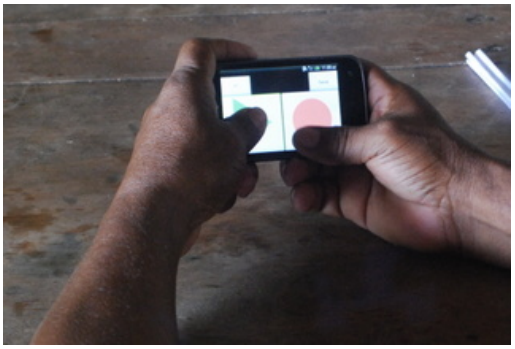
Smartphones are proliferating: they are part of the vanguard of technologies that make it into many isolated communities. Even in the most remote villages, many people own a mobile phone, keep it on their person, and are able to get it charged when mains electricity is unreliable or non-existent. These phones can be inexpensive (US$100-200) and some models have sufficient audio quality to be useful for speech data collection. With suitable software it is possible to collect metadata along with recordings, including location, date, the identity of the speaker, and possibly some information about the content such as the title and genre. The networking capability of a smartphone facilitates wireless sharing and backup.

The speech collection task calls for a variety of individual contributions. The best speakers of the language are not necessarily the best translators; they may be monolingual. Similarly, the best translators may not be the best transcribers; they may be illiterate. Thus, for reasons of skill, not just scale, we need to involve a whole team of people in the data collection activity. In the medium term, we assume that this work would take place under the supervision of a linguist who provides hardware and training, and who monitors the balance of the collection, including coverage of various discourse types, getting everything translated, and so forth).

Aikuma is open source software that supports recording of speech directly to phone storage (Hanke and Bird, 2013; Bird et al., 2014). Recordings are synchronized with other phones that are connected to the same WiFi LAN, so that any user can listen to recordings made on any phone in the same local

network. A user can "like" a recording to improve its overall ranking. A user can also provide a phrase-by-phrase spoken translation of the recording, using the interface shown in Figure 2. This functionality is based on the protocol of "Basic Oral Language Documentation" (Reiman, 2010; Bird, 2010).

(a) Recording a Tembé narrative



(b) Translating Tembé into Portuguese

Figure 2: Recording and translating using the Aikuma Android app

Users press and hold the left *play* button to hear the next segment of audio source. They can press it multiple times to hear the same segment over again. Once ready, they press the right *record* button and speak a translation of the source. This process continues until the source has been fully translated. It generates a second audio file that is time-aligned with the source (cf Figure 1). The app supports playback of the source, the translation, and the source with interleaved translation.

Aikuma maintains a database of speakers and synchronizes this to the other phones along with the recordings and titles, and keeps track of which speaker provided which recording. In this way, basic metadata resides with the recordings, and recordings are effectively backed up several times over. If the contents of one phone are periodically archived, then we have a permanent copy of all the recordings and metadata from all of the phones.

We used HTC Desire C and HTC Desire 200 phones which cost US$160 each. We chose these phones for their support of Android 4 and their recording quality. Unlike a professional audio set-up, mobile phone audio recording includes built-in noise suppression that is optimised for near-field voice sources and attenuates background noise. The software stores audio in uncompressed 16kHz 16-bit mono. The quality of the audio from these phones is more than sufficient to support phonetic analysis (Bettinson, 2013). We expect these materials to be considered of archival quality in those cases where the original recording environment was quiet and where the content itself has linguistic and cultural value.

Another advantage of smartphones compared with professional recording equipment is ease of recharging. Many remote indigenous communities without mains electricity are still able to keep phones charged with the help of generators and car batteries. By choosing to use mobile phones, we can piggy-back on the existing infrastructure.

The cost and usability of smartphones relative to professional recording equipment makes it easy to consider giving them out to people to make recordings in their own time. Apart from significantly increasing the amount of recorded and translated material that a linguist can collect, this gives speakers direct control over the content and context of the recordings, and it may lead to the collection of more naturalistic materials. In some cases, speakers already own an Android phone and can simply install the software and get started.

In the following three sections we report on our experience of using these phones with indigenous communities in the Amazon and the Himalayas.

3 Tembé, Pará State, Brazil

The Tembé language is spoken by approximately 150 people amongst a larger community numbering about 1,500, in a group of villages in the Reserva Alto Rio Guam in the vicinity of Paragominas in the Pará state of Brazil. Bird, Gelbart, and McAlister spent five days in the village of Cajueiro (*Akazu'yw*),

the gateway to several other Tembé villages that can only be accessed by river. Like many Indian villages, Cajueiro is laid out around a soccer field. The village was connected to the electricity grid ten years ago.

We recorded 14 texts from 8 speakers, mostly personal narratives but also a song and a dialogue. Most texts were orally translated; some were translated twice. Of two hours of source audio, 35 minutes were orally translated, producing an extra 25 minutes of audio.

Our visit to Cajueiro is mostly interesting for the great variety of unanticipated challenges, and how we were still able to use the platform to collect data.

Previous contact with the Tembé community was mediated by staff at the Goeldi Museum in Belém. The Tembé community had been discussing prospects for installing an antenna in Cajueiro to enable an Internet connection. On arrival, the chief asked about our plans to set up Internet access, and we explained that we were not able to do this because there was no signal for 100km. After this, the chief lost interest in our activities and we were not able to hold a village meeting as we had hoped, in order to discuss our work, invite participation, and demonstrate the use of the technology for recording and translation. Instead, we could only work one-on-one.

Our first 24 hours in the village was spent on video documentation of a coming-of-age ceremony. More elaborate versions of this ceremony had been filmed in the past, so there was minimal documentary value in recording this event. However, it was the basis for our invitation to the village, cf. (Crowley, 2007, 80), and it enabled us to meet the whole community and to observe the limited social interaction, almost exclusively conducted in Portuguese.

In the following days, we went around the village showing the app to people, explaining our work, playing existing recordings in Tembé and other languages, and trying to find fluent speakers who were motivated to preserve Tembé linguistic heritage. Few people claimed to be fluent and we only found six who were willing to be recorded, all men. No women would consent to being recorded until a Tembé man, trained as a computer technician, learnt how to use the app and took a handset and found two female speakers and recorded them. They were in their thirties, less confident with the language, and could only read haltingly from a small storybook. For the fluent speakers we were able to find, the documentary activity proceeded naturally; they easily recounted histories and gave phrase-by-phrase translations. We prepared a selection from our recordings and made audio CDs to give away for people to play on their personal stereos.

We experienced a variety of technical difficulties with the smartphones, none of which had been apparent during lab testing. The most obvious were due to people's unfamiliarity with smartphones. Signing in required entering the participant's name using a touchscreen keyboard, then selecting an ISO 639 language code via a search interface, then taking a photo using the phone. The photo could not be taken easily by the participant as the phones lacked a front-facing camera. Consequently, we generally took care of these tasks on behalf of speakers. Similarly, upon completion of a recording, the participant was prompted to enter a name for the recording, and we would reclaim the phone and enter a title after a brief discussion with the participant about a suitable choice.

Further problems concerned the translation task. A couple of participants began to give a Portuguese paraphrase immediately on finishing a story. Despite the obvious value of capturing an immediate paraphrase from the same speaker, the software was not designed for this and we had no way to capture the paraphrase as a separate audio file and link it back to the original. The thumb-controlled interface (Figure 2b) was also slightly problematic. Often a speaker would still be holding down the play button with his left thumb at the moment he went to press the record button with his right thumb. Sometimes, speakers would begin to speak and then notice that playback was still continuing, and only then release the play button. By the time they had pressed the record button again, they had already spoken a word or two, and this speech was not captured by the app. This problem happened often enough to interfere with the flow of the translation task. Possible solutions are to have the controls operated by a single thumb, or else to change the behavior of the app so that the most recent thumb gesture overrides a existing button press. Several other interactional issues with the software were identified and resolved with similar minor changes to behavior.

A final set of issues concerned dissemination. Many Indian villages are now equipped with computer

rooms and have desktop machines with CD burners, though mains electricity may be intermittent, or else depend on a generator. We were able to transfer files from the phones to a local machine using a USB connection, though it was a slow process to identify the recordings of interest to the participants and to compile an audio CD. Instead, we realised that any user of a phone should be able to export selected recordings to a local folder that could be burnt to CD.

The key problem for us, however, was lack of participation. The main reason for this, we believe, was the limited local interest in the Tembé language. A secondary factor was the misunderstanding about our contribution ("bringing the Internet") and the fact that the product, a CD of stories, was not necessarily something that the community wanted.

4 Nhengatu, Amazonas State, Brazil

The Nhengatu language is a creole spoken by 10,000 people across a wide area, including the village of Terra Preta, 50km NW of Manaus. Nhengatu used to be the language of communication amongst Indians from different tribes along the Rio Negro, and between Indians and non-Indians in the Brazilian Amazon. Although most of the inhabitants of Terra Preta are ethnically Baré, the only indigenous language spoken in the village is Nhengatu. Younger generations are monolingual in Portuguese. Unusually, there are also some non-Indians living in the village. The villagers were open to receiving us, partly due to their proximity to Manaus and the fact they were accustomed to meeting tourists and showing white people around and selling handcrafts. Compared with Cajueiro, there was a stronger sense of community in Terra Preta: on weekends they would have breakfast together in a communal meeting place, and agree on community service tasks for the weekend.

We made a preliminary visit and presented our work at a public meeting. We called for a volunteer to tell a story to the group and then invited another volunteer to provide an oral translation. Both individuals did a perfect job even though neither one had used the software before. One of them, a former village chief, addressed the group and explained the significance of our work. He then asked if we would help in the preparation of a DVD. Since we did not have the necessary equipment, we offered to create a bilingual storybook instead. They agreed, and said this could be used in their local school. We had already intended to propose this as our contribution to the community after our experience with Tembé, where most people did not grasp the value of us only leaving audio recordings. A booklet would be a natural extension to our documentary goals, and it offered to draw in the whole community including the children who could provide illustrations.

Three weeks later, once the necessary approvals had been obtained, we arrived in Terra Preta and launched our activities with another public meeting. At this meeting, and again at public meetings on the following two mornings, we invited anyone who was interested to take a phone and record a story. Sometimes a storyteller held a phone while addressing a small group (often involving children), and recounted a folktale.

After three days, we recorded 35 texts from nine speakers (including two children), mostly folklore and personal narratives. Most texts were orally translated. Of 2.5 hours of source audio, approximately one hour of recordings were orally translated (some two or more times), producing an extra two hours of audio. Seven short texts by children or directed at children were delivered in Portuguese, and we did not translate these back into Nhengatu.

During the second half of the visit, four men who were literate in Nhengatu joined us in the task of transcribing the stories, focussing on those that would be most interesting for inclusion in the storybook. They worked in parallel, playing back the recordings on the phones, transcribing them on paper, then bringing the sheets back to be typed and proof-read. This work was arduous, continuing through the heat of the day, but they were keen to process as many stories as possible.

Two weeks after our visit, we published a small booklet of stories and translations and sent copies back to the village, and posted a digital copy in the Internet Archive (Bird et al., 2013).

We encountered some additional technical difficulties that we had not experienced in Cajueiro. First, a bug in the recording app which appeared on the last day caused one recording to overrun and produced a three hour (350MB) file. After this, WiFi synchronisation was too slow to be effective, and it was

necessary to perform synchronisation manually, copying the files from all phones onto a laptop, then copying the collection back onto each phone. Second, the presence of an audience for some stories encouraged the storyteller to speak loudly. Since speakers were holding the phone close to their mouths, this resulted in clipped audio. Third, at the height of our intensive transcription and translation process, we needed to keep track of the activities of several participants, and created a checklist. Finally, there was an issue with the power supply. Unlike Cajueiro, Terra Preta is not attached to the electricity grid, but it has a generator which is turned on for four hours every evening, and sometimes during the mornings for brief periods. We could use this to keep the phones charged and to power the router for long enough to synchronise the phones a couple of times each day. But the village became very noisy when power was available, thanks to an abundance of stereo systems and power tools, and this made it difficult to get good quality recordings during these times.

In spite of these problems, there were some successes. The most notable was that participants took no more than a minute to become adept with the recording functionality and the thumb-controlled oral translation functionality (Figure 2b). Second, the availability of multiple networked recording devices meant that we could collect materials in parallel. For example, we could discuss a story we wanted to record and then send several people off at the same time to record their own versions. Then they could synchronise their recordings and hear what each other said. Finally, automatic synchronisation greatly facilitated concurrent transcription activities. We could assign people to transcribe or translate a particular source recording without having to keep track of device it had been recorded with: it was already available on all of the phones.

Figure 3: Transcribing a spoken translation

5 Kagate, Ramechhap district, Nepal

A third field test with a later version of the app was undertaken in Nepal. Kagate, known to its speakers as Syuba, is a Tibeto-Burman language spoken by around 1,500 people in the Ramechhap district, east of Kathmandu. Handsets with the Aikuma app were taken by Gawne and were deployed in parallel, in the context of a project to video record traditional folk narratives and history. Twelve original recordings were made, totalling 80 minutes. Four of these recordings were translated into Nepali, and two recordings were also carefully "respoken" to aid later written transcription (Woodbury, 2003). Although the recordings represent a more modest total than at other fieldsites, this field test demonstrates that Aikuma can operate in conjunction with, and to the benefit of, more traditional field methods. A number of challenges were addressed.

The first challenge was the lack of mains electricity, with the village only having a number of small solar panels for charging mobile phones and running small lights. Much like at the Nhengatu site, mobile phones enabled work to proceed in the absence of mains electricity. Indeed, this was greatly beneficial because it meant that more recordings could be made without rapidly depleting the video camera battery, which required charging at a village a one hour walk away. The lack of proximal mains electricity meant that it was not possible to run the router and synchronise the data on each phone. As a result of this (and participation issues discussed below) the researcher only kept two devices in use at a time, making it easier to keep track of what was on each device. This field trip demonstrated that even without the data synchronization feature Aikuma is still a useful fieldwork tool.

The second challenge was fostering participation. As a number of anthropologists working in related communities have observed, the centre of village life for Kagate people is the household (Fürer-Haimendorf, 1964; Desjarlais, 1992). Relationships beyond this are negotiated through extended familial relations of reciprocity. Therefore, there were no opportunities to arrange community meetings as in

Terra Preta, or even to find an individual who was an officially designated leader. As a result, much time was spent engaging a small number of enthusiastic participants and working with them to engage other members of the community through existing social networks. The benefit of the mobile devices was that they could be carried about and then demonstrated to people during a lull in other activities. Because of this portability and ease of demonstration, the mobile phones became a key part of negotiations with all participants, even those who the community members wanted to video record. Having the handsets meant that we could immediately show people the outcome of a recording session. Sometimes, even after this demonstration, people were reluctant to participate in recording with video cameras or phones. We took this as a positive sign that participants had a better level of informed consent with which to make this choice than they otherwise would have. Many community members were reluctant to take the phones, as even basic smartphones that we chose for their affordability are an expensive commodity and out of the price bracket of many. A small number of people became comfortable enough to take the phones away to work with, but would return them immediately after a specific task had been completed. With a longer period of presence in the village it is likely more people would become more comfortable with the process.

The final issue, like at other sites, concerned the process of saving recordings once they had been made. Processes that are taken for granted with some audiences, like naming a recording, presume a great deal of cultural knowledge about iconography, the layout of keyboards, and spelling conventions. It was only on the final day that one of the more frequent participants saved a file without assistance. Fortunately, an import feature had been built into the app, which meant that when participants returned with files that they had not managed to save they could still be loaded into the list. While some of the issues faced can be overcome through further refining the design, others are useful educational tools to help familiarize participants with key features of digital literacy.

Throughout the above discussion we have touched on some benefits to using Aikuma at this field site. There are some other advantages that are also worth noting. The first is that the portability of the handsets meant that there was a wider range of participants recorded. The limited electricity available for the parallel video documentation, and community attitudes about who was a suitable participant in that work, meant that only a small section of the community (mostly older males) would have been documented. The lower formality of using the phones, compared with a bulky video camera, meant that people also felt quite relaxed, often telling stories with an audience present.

The use of phones also meant that there were fewer missed opportunities for recording. One evening we used the phones when the light was too poor for video. Another morning when the researcher was unwell, she gave one of the handsets to a member of the community who recorded some traditional stories with an older man who had not been able to remember them the day before. On yet another occasion, a man took one of the handsets away and recorded a translation while the researcher was filming a video with another participant. Although the linguist was still needed for the saving of recordings, people became less dependent on her presence to do their own documentation work.

6 Discussion

Reflecting on our experience in the Tembé, Nhengatu, and Kagate communities, further issues warrant discussion.

The mobile device was a major attraction. People gathered round to see how it was used, then explained it to others in their native language. They brought elders to see the work, and encouraged them to tell stories. This impact convinced us that the mobile phone is an effective platform for engaging with participants and helping them quickly grasp the collection and dissemination aspects of language documentation work, cf. (Rice, 2011). Note that the phones were not equipped with SIM cards, and so there was no distraction of them being used for voice calls or for downloading extraneous software.

However, the device was also an obstacle. Although some people had used smartphones, few had experienced touchscreens. Creating a user profile required entering a name using the touchscreen keyboard. It seemed like overkill to train individuals to use a keyboard and to go through a process they would only perform once. Moreover, the language selection process displayed a searchable list of 7,000 languages,

and it would have been easier to have a small selection of local languages to choose from. In Tembé, the man who was trained as a computer technician learned to create user profiles for other people. By the time of the Kagate experiment, we added support for default languages, and set these as Kagate and Nepali. This simplified the task, though it also meant that we did not capture information about people's competencies in other languages. These issues with the device only occurred at the outset, and highlight the need to simplify the metadata collection process. The impact of the problem would be reduced with improved software design.

The device helped with the process of obtaining informed consent. We played an existing recording, either one collected during an earlier phase of documenting the language, or one from another endangered or extinct language. In this way we communicated the idea that language recordings can be preserved and transmitted over distance and time, even once the language is no longer spoken. We also asked what people thought about the idea of others hearing their language, and they were generally enthusiastic. In the case of a further Brazilian language, one community leader asked for substantial donations of hardware and another cited intellectual property concerns, and so we did not record this language. A related open issue concerns the process for *documenting* informed consent, particularly when working with monolingual speakers.

Most of the collected material consisted of personal narratives, folklore, and a limited amount of singing. Other discourse types that we did not collect include dialogue, oratory, and procedural discourse, cf. (Johnson and Aristar Dry, 2002). On many occasions, people listened to a traditional narrative and then asked to recount their own version. Consequently, we see the possibility for achieving substantial lexical overlap in recordings by different speakers, which could help with speech modelling, dialect identification, and lexicon production.

7 Conclusions

We have investigated the use of Aikuma, an Android app designed for recording and translating unwritten languages. We taught members of indigenous communities in Brazil and Nepal to use smartphones for recording spoken language and for orally interpreting it into the national language, and we collected a sample of bilingual phrase-aligned speech in the languages. We collected approximately 8.5 hours of audio, approximately 100,000 words, and in the process, we demonstrated that the platform is an effective way to engage indigenous communities in the task of building phrase-aligned bilingual speech corpora. The built-in networking capability of the phone was used to good effect in Nhengatu for leveraging the contribution of multiple members of the community who have differing linguistic aptitudes.

We identified several areas for additional functionality: support for adding a paraphrase as soon as a story has been told; support for exporting playlists to CD; a checklist that shows which recordings have been translated; permitting handwritten transcriptions to be photographed and linked back to the original audio; and redesigning the interface to remove some remaining English prompts and confusing icons. These and other enhancements are being developed in our open source project.[1]

Above all, we have found that this approach to linguistic data collection greatly facilitates work on indigenous languages that are falling out of use. It bypasses the need for expensive equipment by piggybacking on the burgeoning adoption of mobile phones and wireless broadband networks. We are optimistic about the prospects of using this approach to collect substantial new corpora for supporting linguistic research and language technology development, even for some of the most isolated linguistic communities in the world.

Acknowledgments

This research was supported by NSF Award 1160639 *Language Preservation 2.0: Crowdsourcing Oral Language Documentation using Mobile Devices* (Bird and Liberman), ARC Award 120101712 *Language Engineering in the Field* (Bird), and Firebird Foundation project *Documenting the Traditional Songs and Stories in Kagate, a language of Nepal* (Gawne). Bird, Gelbart, and McAlister are grateful to Dr Denny Moore and the Goeldi Museum (Belém) for facilitating their work in Brazil.

[1] `https://github.com/aikuma`

References

Steven Abney and Steven Bird. 2010. The Human Language Project: building a universal corpus of the world's languages. In *Proceedings of the 48th Meeting of the Association for Computational Linguistics*, pages 88–97. Association for Computational Linguistics.

Timothy Baldwin, Jonathan Pool, and Susan Colowick. 2010. PanLex and LEXTRACT: Translating all words of all languages of the world. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 37–40, Beijing, China.

Stephen Beale. 2012. Documenting endangered languages with Linguist's Assistant. *Language Documentation and Conservation*, 6:104–134.

Emily Bender, Robert Schikowski, and Balthasar Bickel. 2012. Deriving a lexicon for a precision grammar from language documentation resources: A case study of Chintang. In *Proceedings of the 25th International Conference on Computational Linguistics*, pages 247–262.

Emily Bender, Michael Wayne Goodman, Joshua Crowgey, and Fei Xia. 2013. Towards creating precision grammars from interlinear glossed text: Inferring large-scale typological properties. In *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 74–83. Association for Computational Linguistics.

Mat Bettinson. 2013. The effect of respeaking on transcription accuracy. Honours Thesis, Dept of Linguistics, University of Melbourne.

Steven Bird, Katie Gelbart, and Isaac McAlister, editors. 2013. *Fábulas de Terra Preta*. Internet Archive.

Steven Bird, Florian R. 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*. Association for Computational Linguistics.

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

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*.

Terry Crowley. 2007. *Field Linguistics: A Beginner's Guide*. Oxford University Press.

Dipanjan Das and Slav Petrov. 2011. Unsupervised part-of-speech tagging with bilingual graph-based projections. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 600–609. Association for Computational Linguistics.

Nic de Vries, Marelie Davel, Jaco Badenhorst, Willem Basson, Febe de Wet, Etienne Barnard, and Alta de Waal. 2014. A smartphone-based ASR data collection tool for under-resourced languages. *Speech Communication*, 56:119–131.

Robert R. Desjarlais. 1992. *Body and emotion: the aesthetics of illness and healing in the Nepal Himalayas*. Philadelphia: University of Pennsylvania Press.

Mark Dredze, Aren Jansen, Glen Coppersmith, and Ken Church. 2010. NLP on spoken documents without ASR. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 460–470. Association for Computational Linguistics.

Christoph von Fürer-Haimendorf. 1964. *The Sherpas of Nepal: Buddhist highlanders*. London: John Murray.

Florian R. Hanke and Steven Bird. 2013. Large-scale text collection for unwritten languages. In *Proceedings of the 6th International Joint Conference on Natural Language Processing*, pages 1134–1138. Asian Federation of Natural Language Processing.

Heidi Johnson and Helen Aristar Dry. 2002. OLAC discourse type vocabulary. `http://www.language-archives.org/REC/discourse.html`.

Chia-ying Lee and James Glass. 2012. A nonparametric bayesian approach to acoustic model discovery. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 40–49. Association for Computational Linguistics.

Lori Levin, Jeff Good, Alison Alvarez, and Robert Frederking. 2006. Parallel reverse treebanks for the discovery of morpho-syntactic markings. In Jan Hajič and Joakim Nivre, editors, *Proceedings of the Fifth Workshop on Treebanks and Linguistic Theories*, pages 103–114.

Alexis Palmer, Taesun Moon, Jason Baldridge, Katrin Erk, Eric Campbell, and Telma Can. 2010. Computational strategies for reducing annotation effort in language documentation. *Linguistic Issues in Language Technology*, 3:1–42.

Katharina Probst, Lori Levin, Erik Peterson, Alon Lavie, and Jaime Carbonell. 2002. MT for resource-poor languages using elicitation-based learning of syntactic transfer rules. *Machine Translation*, 17(4):225–270.

Will Reiman. 2010. Basic oral language documentation. *Language Documentation and Conservation*, 4:254–268.

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

S.M. Siniscalchi, Dau-Cheng Lyu, T. Svendsen, and Chin-Hui Lee. 2012. Experiments on cross-language attribute detection and phone recognition with minimal target-specific training data. *IEEE Transactions on Audio, Speech, and Language Processing*, 20:875–887.

Ngoc Thang Vu, Franziska Kraus, and Tanja Schultz. 2011. Rapid building of an ASR system for under-resourced languages based on multilingual unsupervised training. In *Interspeech*, pages 3145–3148.

Anthony C. Woodbury. 2003. Defining documentary linguistics. In Peter Austin, editor, *Language Documentation and Description*, volume 1, pages 35–51. London: SOAS.

Ruiqiang Zhang, Genichiro Kikui, Hirofumi Yamamoto, Taro Watanabe, Frank Soong, and Wai Kit Lo. 2004. A unified approach in speech-to-speech translation: integrating features of speech recognition and machine translation. In *Proceedings of the 20th International Conference on Computational Linguistics*, pages 1168–1174.