A Scalable Method for Preserving Oral Literature from Small Languages

Steven Bird

Dept of Computer Science and Software Engineering, University of Melbourne
Linguistic Data Consortium, University of Pennsylvania

Abstract. Can the speakers of small languages, which may be remote, unwritten, and endangered, be trained to create an archival record of their oral literature, with only limited external support? This paper describes the model of "Basic Oral Language Documentation", as adapted for use in remote village locations, far from digital archives but close to endangered languages and cultures. Speakers of a small Papuan language were trained and observed during a six week period. Linguistic performances were collected using digital voice recorders. Careful speech versions of selected items, together with spontaneous oral translations into a language of wider communication, were also recorded and curated. A smaller selection was transcribed. This paper describes the method, and shows how it is able to address linguistic, technological and sociological obstacles, and how it can be used to collect a sizeable corpus. We conclude that Basic Oral Language Documentation is a promising technique for expediting the task of preserving endangered linguistic heritage.

1 Introduction

Preserving the world's endangered linguistic heritage is a daunting task, far exceeding the capacity of existing programs that sponsor the typical 2-5 year "language documentation" projects. In recent years, digital voice recorders have reached a sufficient level of audio quality, storage capacity, and ease of use, to be used by local speakers who want to record their own languages. This paper investigates the possibility of putting the language preservation task into the hands of the speech community. With suitable training, they can be equipped to record a variety of oral discourse genres from a broad cross-section of the speech community, and then provide additional content to permit the recordings to be interpreted by others who do not speak the language. The result is an audio collection with time-aligned translations and transcriptions, a substantial archival resource.

This paper describes a method for preserving oral discourse, originating in field recordings made by native speakers, and generating a variety of products including digitally archived collections. It addresses the problem of unwritten languages being omitted from various ongoing efforts to collect language resources for ever larger subsets of the world's languages [1]. The starting point is Reiman's work [2], modified and refined so that it uses appropriate technology

for Papua New Guinea, and so that it can scale up easily. The method has been tested with Usarufa, a language of Papua New Guinea. Usarufa is spoken by about 1200 people, in a cluster of six villages in the Eastern Highlands Province, about 20km south of Kainantu (06°25'S, 145°39'E). There are probably no fluent speakers of Usarufa under the age of 25; only the oldest speakers retain the rich vocabulary for animal and plant species, and for a variety of cultural artefacts and traditional practices. Some texts including the New Testament and a grammar have been published in Usarufa [3]. However, only a handful of speakers are literate in the language.

2 Basic Oral Language Documentation

2.1 Audio Capture

The initial task in Basic Oral Language Documentation (BOLD) is audio capture. Collecting the primary text from individual speakers is straightforward. They press the record button, hold the voice recorder a few inches from their mouth, and begin by giving their name, the date, and location. The person operating the recorder may or may not be the speaker.

Fig. 1. Informal Recording of Dialogue and Personal Narrative

Collecting a dialogue involves two speakers plus someone to operate the voice recorder (who may be a dialogue participant). The operator can introduce the recording and hold the recorder in an appropriate position between the participants. The exchange shown in Fig. 1 involved a language worker (left), the author, and a village elder. The dialogue began with an extended monologue from the man on the left, explaining the purpose of the recording and asking the other man to recount a narrative, followed by some conversation for clarification, followed by an extended monologue from the man on the right. The voice recorder was moved closer to the speaker during these extended passages, but returned to the centre during conversational sections. In most cases, the person operating the recorder was a native speaker of Usarufa, and was also participating in the dialogue. The operator was instructed not to treat the recorder like a

hand-held microphone, moved deliberately between an interviewer and interviewee to signal turns in the conversation. Instead, the recorder was to be held still, and usual linguistic cues were to be used for marking conversational turns. A configuration which was not tried would be to have separate lapel microphones, one per speaker, connected to the digital voice recorder via a splitter jack. However, this would have involved four pieces of equipment (recorder, splitter, and two microphones), and increased risks of loss, incorrect use, and degraded signal quality.

2.2 Oral annotation and text selection

The oral literature collected in the first step above has several shortcomings as an archival resource. Most obviously, its content is only accessible to speakers of the language. If the language falls out of use, or if knowledge of the particular word meanings of the texts is lost, then the content becomes inaccessible. Thus, it is important to provide a translation. Fortunately, most speakers of minority languages also speak a language of wider communication, and so they can record oral translations of the original sources. This can be done by playing back the original recording, pausing it regularly, and recording a translation on a second recorder, a process which is found to take no more than five minutes for each minute of source material.

A second shortcoming is that the original speech may be difficult to make out clearly by a language learner or non-speaker. The speech may be too fast, the recording level may be too low, and background noise may obscure the content. Often the most authentic linguistic events take place in the least controlled recording situations. In the context of recording traditional narratives, elderly speakers are often required; they may have a weak voice or few teeth, compromising the clarity of the recording. These problems are addressed by having another person "respeak" the original recording, to make a second version [2, 4]. This is done at a slower pace, in a quiet location, with the recorder positioned close to the speaker. This process has also been found to take no more than five minutes for each minute of source material.

A third shortcoming is that the original collection will usually be unbalanced, having a bias towards the kinds of oral literature that were the easiest to collect. While it is possible to aim for balance during the collection process, one often cannot predict which events will produce the best recordings. Thus, it is best to capture much more material than necessary, and only later create a balanced collection. Given that the respeaking and oral translation take ten times real time, we suggest that only 10% of the original recordings are selected. This may be enough for a would-be interpreter in the distant future to get a sufficient handle on the materials to be able to detect structure and meaning in the remaining 90% of the collection. The texts are identified according to the following criteria:

1. cultural and linguistic value: idiomatic use of language, culturally significant content, rich vocabulary, minimal code-switching

2. diversity: folklore, personal narrative, public address, dialogue (greeting, discussion, instruction, parent-child), song
3. recording quality: clear source recording, minimal background noise

2.3 Recommended protocol for oral annotation

The task of capturing oral transcriptions and translations onto a second recorder offers an array of possibilities. After trying several protocols, we settled on the one described here. The process requires two native speakers with specialised roles, the *operator* and the *talker*, and is illustrated in Fig. 2.

Fig. 2. Protocol for Respeaking and Oral Translation: the operator (left) controls playback and audio segmentation; the talker (right) provides oral annotations using a second recorder

Once a text has been selected, it is played back in its entirety, and the two language workers discuss any aspect of the recording which is problematic. For instance, older people captured in the recordings may have used near-obsolete vocabulary unknown to younger language workers. This preview step is also important as the opportunity to experience the context of the text; sometimes a text is so enthralling or amusing that the oral annotators are distracted from their work. When they are ready to begin recording, the operator holds the voice recorder close to the talker, with the playback speaker (rear of recorder) facing the talker. The talker holds the other recorder about 10cm from his/her mouth, turns it on, checks that the recording light came on, and then introduces the recording, giving the names of the two language workers, the date and location, and the identifier of the original recording.

For the respeaking task, the operator pauses playback every 5-10 words (2-3 seconds), with a preference for phrase boundaries. For the translation task, the operator pauses playback every sentence or major clause (5-10 seconds), trying to include complete sense units which can be translated into full sentences. The

talker leaves the second recorder running the whole time, and does not touch the controls. This recorder captures playback of the original recording, along with the respoken version or the translation. The operator monitors the talker's speech, ensuring that it is slow, loud, and accurate. The operator uses agreed hand signals to control the talker's speed and volume, and to ask for the phrase to be repeated. When necessary, the talker is prompted verbally with corrections or clarifications, and any interactions about the correct pronunciation or translation are captured on the second recorder. Once the work is complete, recording is halted, and the logbooks for both recorders are updated.

2.4 Logbooks

For each primary text, the language workers note the date, location, participant names, topic, and genre, using the logbook provided with each recorder. Genre is coded using the OLAC Discourse Type vocabulary [5]. If there is any major problem during the original recording, a fresh recording is started right away. Pausing to delete files is a distraction, draws attention to the device, and is prone to error. The recorder has substantial capacity and extraneous recordings can easily be filtered out later during the selection process.

Folder	Number	Names	Topic	Genre
C	1	Steve		
C	2	Jema Mark	Orphan	N
C	3	Panayoi . E	Older People	N
C	4	Panayai . E	shooting a snake	S
C	5			
C	6	Wake . Mark	Testimony	N
C	7	Moi . E	Father beating him	N
C	8	Moi E		S
C	9	Moi . E	Biography	
C	10	Mark Negi	Biography	

Fig. 3. Metadata Capture in Village: (a) creating metadata by listening to the opening of each recording; (b) scanned page showing file identifier, participants, topic and genre (date and location were already known in this case)

2.5 Summary

Fig. 4 summarises the process, assuming 10 hours (100k words) of primary recordings are collected. It includes a third stage – not discussed above – involving pen and paper transcription using any orthography or notation known to the participants, such as the orthography of the language of wider communication. Such transcripts, while imperfect, serve as a finding aid and as a clue to linguistically salient features such as sound contrasts and word breaks. A separate archiving process involves occasional backup of recorders onto a portable mass storage device, keyboarding texts and metadata, and converting audio files to a non-proprietary format, steps that typically require outside support.

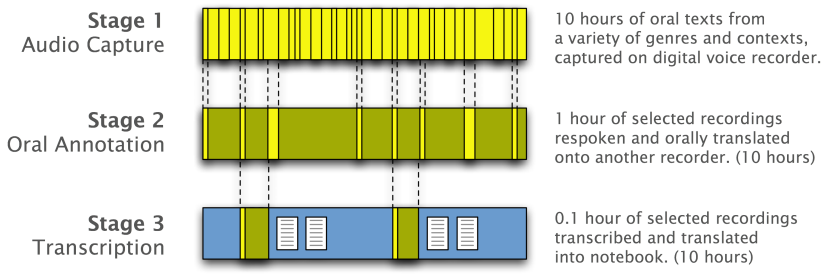

Fig. 4. Overview of Basic Oral Language Documentation

3 Pilot Study in Papua New Guinea

The above protocol was developed during a pilot study during April-June 2009. Bird and Willems trained a group of language workers in the village for one week, then left them to do oral literature collection and oral annotation for a month, then brought them into an office environment to work on further oral annotation and textual transcription. In this section, the activities are briefly described and the key findings are reported.

3.1 Activities

Village-based training. Teachers, literacy workers, and other literate community members were gathered for a half-day training session. It took place in the literacy classroom in Moife village, with everyone sitting on the floor in a circle. We explained the value of preserving linguistic heritage and demonstrated the operation of the voice recorders. Participants practiced using the recorders and were soon comfortable with the controls and with hearing their own voices. Next, participants took turns to record a narrative while the rest of the group observed. Later, we demonstrated the oral annotation methods and the participants practiced respeaking and oral translation. The four recorders were loaned out, and participants were asked to collect oral literature during the evening and the next day, and to return the following day to review what they collected and to continue practicing oral annotation. A further five days were spent doing collection and annotation under the supervision of Bird and Willems.

Village-based collection and oral annotation. In the second stage, we sent the digital voice recorders and logbooks back to the village for two 2-week periods. This would assess whether the training we provided was retained. Could the participants find time each day for recording activities? Could they meet with an assigned partner to do the oral annotation work using a pair of recorders? Could they maintain the logbooks? Apart from reproducing the activities from the first stage, they were asked to broaden the scope of the work in three ways. First, they were to collect audio in a greater range of contexts (e.g. home, market, garden,

church, village court) and a greater range of genres (e.g. instructional dialogue, oratory, child-directed speech). They were to include a wider cross-section of the community, including elderly speakers and children, and to go to the other villages where Usarufa is spoken, up to two hours walk away. Finally, they were asked to train another person in collecting oral discourse and maintaining the logbook, then entrust the recorder to that person.

Town-based oral annotation and transcription. In the third stage, we asked the language workers come to Ukarumpa, a centralized Western setting 20km away, near Kainantu, with office space and mains electricity. This provided a clean and quiet environment for text selection and oral annotation, plus the final step of the BOLD protocol: writing out the transcriptions and translations for a selection of the materials. The town context also permitted us to explore the issue of informed consent. Four speakers saw how it was possible to access materials for other languages over the Internet (see Fig. 5), and even listen to recordings of dead languages. As community leaders, they gave their written consent for the recorded materials to be placed in a digital archive with open access.

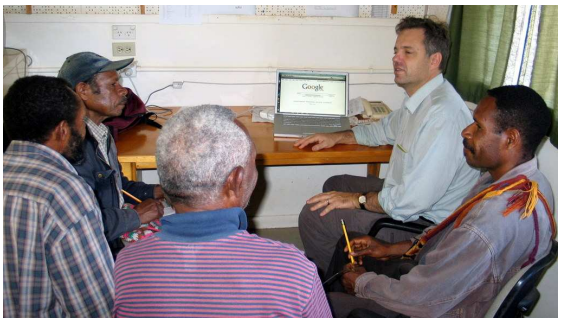

Fig. 5. Experiencing the Web and Online Access to Archived Language Data

3.2 Findings

The findings summarized here include many issues that were encountered early on in the pilot study but resolved in time for the town-based stage, leading to the protocol described in Section 2 above.

Recording. The Usarufa speakers had no difficulty in operating the recorders and collecting a wide variety of material. The built-in microphone and speaker avoided the need for any auxiliary equipment. The clear display and large controls were ideal, and the small size of the device meant it could be hidden in clothing and carried safely in crowded places. We gave out four recorders for periods of up to two weeks, and some were lent on to others, but none were

lost or damaged. Many members of the speech community were willing to be recorded, though some speakers spoke in a stilted manner once the recorder was turned on, and others declined to be recorded unless they were paid a share of what they assumed the language workers were being paid per recording.

Respeaking. Talkers usually adopted the fast tempo of the original recording, in spite of requests to produce careful speech. When the audio segment was long, they sometimes omitted words or gave a paraphrase. Texts from older people presented difficulties for younger speakers who did not always know all the vocabulary items. These problems were resolved, by having a second person control playback and monitor speed and accuracy of the respoken version, and by having both people listen through the recording first, to discuss any problematic terms or concepts.

Oral Translation. A key issue was the difficulty in translating specialised vocabulary into the language of wider communication (Tok Pisin). For example, the name of a tree species might be translated simply as *diwai* (tree), or *sampela kain diwai* (some kind of tree). They were asked to mention any salient physical or cultural attributes of the term the first time it was encountered in a text. Another problem arose as a consequence of using the transcriber to control playback. The translator sometimes paused mid translation, in order to compose the rest of the translation before speaking. This pause was sometimes mistaken for the end of the translation, and the transcriber would resume playback. Occasionally, the resumed translation and resumed playback overlapped (just like when two people might start speaking simultaneously after a brief pause in conversation). This problem is solved by having the translator nod to the operator when s/he is finished translating a segment.

Segmentation. Fundamental to respeaking and translation is the decision about where to pause playback of the original recording. While listening to playback, one needed to anticipate phrase boundaries in order to press the pause button. Older participants, or those with less manual dexterity, tended to wait until they heard silence before deciding to pause playback, by which time the next sentence had started. These problems were largely resolved once we adopted the practice of having participants review and discuss recordings before starting oral annotation, and simply through practice (e.g. about an hour of doing oral annotation).

Metadata. Each participant was able to document their recordings in the supplied logbook. There was some variability in how the participants interpreted the instructions, resolved in later work by attaching a fold-out flyer with examples in the back of the logbooks. It was easy for anyone to check the state of completeness of the metadata by pressing the folder button to cycle through the five folders, and checking the current file number against the corresponding page of the exercise book. At the end of the pilot study, the logbooks were scanned and converted to PDF format for archiving. These scans are the basis for creating OLAC metadata records [6, 7].

Archiving. The contents of the recorders were transferred to a computer via a USB cable. We had engraved unique identifiers on the recorders, but the filenames inside each recorder were identical, and care had to be taken to keep them separate on disk. A more pernicious problem was that the file names displayed on the recorder (e.g. folder `C`, file `01`) did not corresponded to the name inside the device, where file numbers were in time order and not relative to folder. For example, `C01` could have filename `VN52017` (which means that it is the 17th file on the recorder, even though it is the first file in Folder `C`). Thus, the identifier for the audio file (machine id, folder letter, file number) should be spoken at the start of each recording. (Care must be taken to only read out the file number once recording has started.) A selection of the audio files were burnt on audio CD for use back in the village, and the complete set of recordings are being prepared for archiving with PARADISEC [8], and with the Institute of PNG Studies in Port Moresby.

4 Conclusions and Further Work

This paper has described a method for preserving oral literature that has been shown to work effectively for a minority language in the highlands of Papua New Guinea. Using appropriate technology and a simple workflow, people with no previous technical training were able to collect a significant body of oral literature (30 hours), and provide oral annotations and textual transcriptions for a small selection. Much of the collection and annotation work could happen in the evenings, when people were sitting around their kitchen houses lit only by the embers of a fire and possibly a kerosene lantern. At US $50 per recorder, it was easy to acquire multiple recorders, and little was risked when they were given out to people to take away for days at a time. Note that the important matters of access and use have not been addressed here (cf. [9, 10]).

This approach to language documentation has several benefits. It harnesses the voluntary labour of interested community members who already have access to a wide range of natural contexts where the language is used, and who decide what subjects and genres to record, cf. [11], and who are in an excellent position to train others. They are also able to move around the country to visit other language groups far more easily than a foreign linguist could. As owners of the project they may be expected to show a higher level of commitment to the task, enhancing the quality and quantity of the collected materials. The activities easily fit alongside language development activities, adding status and substance to those activities, and potentially drawing a wider cross-section of the community into language development. Limited supervision by a trained linguist/archivist is required between the initial training and the final archiving. Metadata can be collected alongside the recording activities in a simple logbook which accompanies the voice recorder, and then captured for the later creation of electronic metadata records. The whole process is able to sit alongside ongoing language documentation and development activities (and there is no suggestion that it supplant these activities).

Building on the success of the pilot study, a much larger effort is underway in 2010, involving 100 digital voice recorders donated by Olympus Imaging Corporation, in collaboration with the University of Goroka, the University of PNG, Divine Word University, the Institute of PNG Studies, and the Summer Institute of Linguistics, with sponsorship from the Firebird Foundation for Anthropological Research (http://boldpng.info/).

Acknowledgments

I am indebted to staff of the Summer Institute of Linguistics at Ukarumpa, especially Aaron Willems, for substantial logistical and technical support.

References

1. Maxwell, M., Hughes, B.: Frontiers in linguistic annotation for lower-density languages. In: Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006, Association for Computational Linguistics (2006) 29–37 http://www.aclweb.org/anthology/W06-0605.
2. Reiman, W.: Basic oral language documentation (2009) Presentation at the First International Conference on Language Documentation and Conservation.
3. Bee, D.: Usarufa: a descriptive grammar. In McKaughan, H., ed.: The Languages of the Eastern Family of the East New Guinea Highland Stock. University of Washington Press (1973) 324–400
4. Woodbury, A.C.: Defining documentary linguistics. In Austin, P., ed.: Language Documentation and Description. Volume 1. London: SOAS (2003) 35–51
5. Johnson, H., Aristar Dry, H.: OLAC discourse type vocabulary (2002) http://www.language-archives.org/REC/discourse.html.
6. Bird, S., Simons, G.: Extending Dublin Core metadata to support the description and discovery of language resources. Computers and the Humanities **37** (2003) 375–388 http://arxiv.org/abs/cs.CL/0308022.
7. Bird, S., Simons, G.: Building an Open Language Archives Community on the DC foundation. In Hillmann, D., Westbrooks, E., eds.: Metadata in Practice: a work in progress. Chicago: ALA Editions (2004)
8. Barwick, L.: Networking digital data on endangered languages of the Asia Pacific region. International Journal of Indigenous Research **1** (2005) 11–16
9. Duncker, E.: Cross-cultural usability of the library metaphor. In: JCDL '02: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, Association for Computing Machinery (2002) 223–230
10. Jones, M., Harwood, W., Buchanan, G., Lalmas, M.: StoryBank: an Indian village community digital library. In: JCDL '07: Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, Association for Computing Machinery (2007) 257–258
11. Downie, J.S.: Realization of four important principles in cross-cultural digital library development. In: JCDL Workshop on Cross-Cultural Usability for Digital Libraries. (2003)

University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
BIRD, STEVEN

Title:
A scalable method for preserving oral literature from small languages

Date:
2010

Citation:
Bird, S. (2010). A scalable method for preserving oral literature from small languages. In Proceedings of the 12th International Conference on Asia-Pacific Digital Libraries, Gold Coast, Australia.

Publication Status:
In Press

Persistent Link:
http://hdl.handle.net/11343/27682

File Description:
A Scalable Method for Preserving Oral Literature from Small Languages

University Library

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Bird, S

Title:
A Scalable Method for Preserving Oral Literature from Small Languages

Date:
2010-08-16

Citation:
Bird, S. (2010). A Scalable Method for Preserving Oral Literature from Small Languages. Chowdhury, G (Ed.) Koo, C (Ed.) Hunter, J (Ed.) ROLE OF DIGITAL LIBRARIES IN A TIME OF GLOBAL CHANGE, 6102, pp.5-14. SPRINGER-VERLAG BERLIN. https://doi.org/10.1007/978-3-642-13654-2_2.

Persistent Link:
http://hdl.handle.net/11343/32227