27
Flexible IT resources for community language reclamation: using culturally appropriate contexts

Cat Kutay[1], George Fisher[2] and Richard Green[3]

Abstract

This paper describes work utilising information technology developed by Cat Kutay and computing students at the University of New South Wales to support two different Aboriginal language programs in Sydney under the guidance of community tutors – George Fisher who teaches Wiradjuri and Richard Green who teaches Dharug. These languages are in the process of being reclaimed from archival resources, supported by the remaining speakers. Each language presents unique challenges. While the New South Wales Department of Education, Department of Aboriginal Affairs and Board of Studies have supported these languages in schools, it is also important to develop programs where the parents and community are involved in reclaiming the languages to ensure the process is ongoing. This work is located in both contexts. We discuss our experience using computing resources to promote the sharing of language, situating it in the field of research into computer-mediated human interaction.

The aim of language reclamation is to provide the original speakers of a language with the opportunity to once again express themselves in their language. This requires more than knowledge of a word list or dictionary, or even the language structure. Without deep knowledge of their language it is difficult for Indigenous people today to express their culture and the related knowledge. This paper looks at a variety of fairly simple information technology (IT) resources that are being used or developed to increase the depth of language teaching and sharing within the Sydney community.

---

1 Computer Science and Engineering, UNSW.

2 Bankstown Elders Group.

3 Chifley College, Dunheved Campus.

Firstly, we discuss why it is important we reclaim languages in Sydney. Then we look at some of the different situations in which this is occurring using IT. In particular we look at the role and suitability of existing or proposed IT to support this work – the main criterion being the reduction of the impact of the computer in the process of sharing language resources, by providing seamless communication among speakers or between teacher and student.

We look at the difficulties encountered in developing software resources, many of which remain the same as for developing any Aboriginal language teaching material: the lack of historical data and authentic modern examples of a vibrant language. However there is, in addition, the nature of the technology, which is viewed as difficult by most of the language community.

Finally we end with a request for more such resources to be developed. This is a critical area where computing can support human communication in an under-resourced area. Dr Kutay is researching how the computer can mediate among the human users who provide the innovation and creativity within a framework afforded by applications on the web or on a local computer. In particular we look to the open source community as, while the resultant products tend to be designed by developers for developers, these can provide appropriate support for the process of language reclamation in a domain where funding for technology support is restricted.

Language for knowledge

Indigenous languages developed over centuries to enable the expression of a particular culture or worldview. In reclaiming Indigenous languages we seek to provide the original vocabulary and syntactic structure that is required for this expression. The significant difference between Indigenous languages and European languages is often the former's ability to describe detailed landscapes and interrelations as required for story telling. These descriptions are of interactions which are often spiritual in nature and which are human-to-human, and human-to-nature, with all relations presented as between equals. The language is highly contextualised in a continuum of time (Dreamtime stories) or place (songlines) (Groome 1995; Harkins 1994). European languages tend to be more focused on the expression of an individual negotiating with an external world, such as the use of dyads in teaching and the use of the impersonal form for many living beings, as well as a segmented view of time and place in describing events as located at one point (Christie 1985; Harris 1991).

Indigenous languages have developed to express Indigenous knowledge. Indigenous cultures in Australia have a very different focus on knowledge management to European cultures. The main differences are summarised in Nakata et al. (2008) and these apply similarly to the language used for the transmission of this knowledge. Much of the difference in requirements of a language stems from the fact that Indigenous knowledge is generally conveyed orally rather than by being written. The sharing of knowledge involves a thorough teaching of that knowledge, within a framework of poetry and singing, for ongoing recollection. Furthermore if information

was freely shared without the experience and background context required to enable understanding, the oral record would become jumbled and incoherent. This background context is the story in which the knowledge is conveyed. Thus when we teach Indigenous languages we need to retain the background context at all times.

Language information

We are working to promote the sharing of Indigenous cultural knowledge through language while incorporating IT to provide flexible language learning resources. These resources are intended to assist language teachers and students in learning the reclaimed languages and are not developed specifically for a single language or for linguists, but designed to support community workers in many languages. Examples of such resources are Miromaa database developed by the Arwarbukarl Cultural Resource Association[4] in Newcastle and LanguageWiki developed by the Sydney Aboriginal Language and Computing Centre (SALC)[5] in Sydney.

To develop online shareable IT resources we first consider the issues relating to enabling Indigenous use of them and what may alleviate any problems. In relation to the language information from which we develop the computer resources, we need to consider security issues versus the need for public access. We have to ensure the security of data, both for storage and sharing, so that information can be updated only by those with the rights to access it, while maintaining as open as possible access to the data for knowledge sharing; and we need to look at the validity of the information that we present as authoritative language resources on the computer, or the *accountability* of the data (Bird & Simons 2003).

When collecting the first wordlist for teaching a language we have to verify that each word is actually from that language. In the case of Dharug we are working mainly with an oral record so this may include material common to neighbouring languages from people now living in Sydney. We have tried to overcome this difficulty by providing each language with a wordlist in online databases on the SALC website that are available for editing within a content management system, with the facility to upload waveform audio format (wav) files and images. For Wiradjuri, three speakers started to enter the information into the language database (Figure 1) and upload teaching resources developed at community workshops. This ensures that the resources are not static but able to be updated as our knowledge increases. To ensure that these databases are not corrupted they can only be altered by people who are allowed to register and create a password-protected account.

The important issue then was how to use the language data we collected. As Christie (2004, p. 1) noted, 'databases do not contain knowledge, they contain information. Education is not the transmission of information from one head to another … it is the negotiated production of knowledge in context'. Information must be presented

---

4 See www.miromaa.com.au

5 See www.salc.org.au

and shared in a manner that retains the context of that information. For instance in recording words from a language we want to link this to the pronunciation and an example of the use of the word. This is important, as any translation to English will not be an exact transfer of meaning, and much knowledge is lost by removing the words from their original context. Noting Nakata et al.'s (2008) concerns relating to the oral transmission of knowledge, we wanted to provide access for many users to upload their information including oral recordings, and enable these users to link this information to learning resources such as worksheets or games.

Figure 1: Web page view of data for a Wiradjuri word.

Learning environment

While much Indigenous knowledge is already shared online (Dyson et al. 2007; Kutay & Mooney 2008), the domain is still limited. In fact the need for the IT in this project arose from the language teaching environment. Most importantly the inclusion of technology was to enable a small number of language speakers to support a larger number of in-school tutors. We also saw the use of computers in language teaching as an opportunity for promoting computer literacy together with providing applications that go beyond word learning, to support grammatical literacy.

Aboriginal language teaching is rapidly expanding in Sydney schools, faster than tutors can be trained. These programs were originally seen as a way to promote Indigenous culture, however it is also believed to increase the participation of Indigenous students at school. The Dharug courses in Sydney have already gained

support for their perceived effect on attendance rates, although this has not been formally studied. In the schools teaching Wiradjuri there have also been reports from teachers that students have become more involved in English classes as a result of their own language studies.

The number of languages we would like to support is very large. In the schools where this work is based we are dealing with students who identify with a range of different languages, all of which are near dormant; most students' parents cannot speak their ancestral language, although they may occasionally use words from a variety of languages. This situation was invariably arrived at through involuntary relocation or removal from family, however commitment to and interest in language reclamation is still very strong.

Computer-mediated human interaction

The development of software support for teaching Aboriginal languages in school adopts the approach that the computer tools need to be as ubiquitous as possible, and assist the Aboriginal tutors and student users to work with the language knowledge that can be stored on computer in the form of a database, or basic grammar in a language parser. The initial applications were kept simple and, as the users became more ambitious with what they wanted to achieve, the software was developed to keep pace, where it could.

While such software resources will always provide only a partial knowledge base for a language, they can support the existing courses in schools that focus on wordlists with simple sentence construction exercises, sometimes with example songs and stories. Also by having speakers provide audio for the database, we now have suitable expressions for English terms, such as greetings, rather than direct translations.

In order to support the in-school tutors we intended to build up a body of IT resource templates and applications that provided a learning environment that could be used for similar languages. Then, by linking these applications to databases for the different languages and providing grammatical parsing rules for simple translation, we hoped to provide support for each language. The goal was to assist tutors to provide different practical exercises in a cultural context and link these to feedback to be provided to the students' activities within the learning software. One difficulty is that this feedback should preferably be in the spoken language.

The first step in this process was to assist the present speakers of the main languages represented in Sydney to store their language knowledge through a simple database structure. We then combined one language database with the resources, such as recordings presented by other speakers and sentences, to provide a rich learning environment. The steps in this process are described below. At the same time we needed to carry out this process with the many different languages, to verify if the applications could be generic enough to support the basic language structures and, where there were differences, to see whether these required specific grammar rules or a different construction of the database.

With our present work focusing on providing subject lists of words with simple link-ing words and some tense variation, the applications so far developed have proved useful. However we are wary of how far we will be able to support any in-depth teaching of languages. While the Sydney languages are all closely located (being from coastal and central inland NSW) there is much variation even within these languages (see Dixon & Blake 1979). In particular, we will need to collect more example sen-tences for the system to interpret these variations among languages.

Collecting the language data

In the reclamation of languages in Sydney the collection of oral resources has been hard. This is partly because most languages currently used here are not from this area originally. Many people have moved and lived here for generations. In Wiradjuri we are working with archival resources and a handful of language speakers who live far apart. Most speakers are not in the Sydney region where we are working. While people have now started to relearn their language it remains very time consuming to gather spoken examples.

To collect resources in the spoken language we also needed to go back to archival material held at the Australian Institute of Aboriginal and Torres Strait Islander Stud-ies (AIATSIS).[6] Much of the audio material is in the form of linguists' field recordings of wordlists or single sentences with some complete songs or stories, but little in the middle range suitable for beginning learners. Linguists are recorded asking, 'What is the word for … ' with a speaker's response. Long responses may not be transcribed or translated and longer recordings are often not described or analysed in any way. This results in long searches of archives to find material to support different lesson topics.

On the other hand, Dharug is a language of the Sydney area and many speakers live in western Sydney and the Blue Mountains to which we have ready access. To col-lect Dharug resources we work with the local speakers as well as including language structure and vocabulary described by the linguist Jakelyn Troy (1994). In Dharug we are also using song as the teaching medium (see Green, this volume). Song has always been recognised as a means of assisting retention of the words in a story and Aboriginal languages are ideal for singing, with a flexible phrase order that can per-mit re-arranging words to fit a rhythm or rhyme.

The next task was to store the wordlists electronically. Some short electronic word-lists have been published on CD by the Department of Aboriginal Affairs through the NSW Aboriginal Language Research and Resource Centre and more examples are available at SALC. The SALC wordlist database was designed to be used with differ-ent languages and built as a publicly accessible online site. The database has been modified in some cases, where language users wish to store alternative data (such as the archival source of a word definition) as part of their data. An image to portray

---

6 A large portion of text resources for Wiradjuri recorded by James Günther in the 1800s, and only recently discovered, has just been presented to the AIATSIS library for public use.

the meaning was added for words where possible. The Wiradjuri list was entered by scanning archival resources, then running text recognition software, which had to be checked by hand. The words were put into a table. Unfortunately it was only later that the words were linked back to their source sentence, kept in a separate database. The use of a computer made it easier to develop English to Wiradjuri and Wiradjuri to English lists, check for repetition, and so on. Then a sound file of a modern speaker saying the word and a text example of the word in a sentence were added. The sound was retrieved from different sources. We had some Central Australian Aboriginal Media Association recordings of recent speakers, as well as George Fisher who was equipped with good quality recording instruments and motivated to do this work for minimal funds. All resources could be uploaded at home by the language users, gradually in their own time.

The database was formatted as a flat table that uses web references to files for accessing non-text data. A single English word translation is used as one search field to enable the database to be more usable in parsing, and an alternative more explanatory meaning is included, however any word could be entered many times for its different meanings. Also the more complex meaning was included as an alternative translation. The web system is coded in PHP, a server-side scripting language, to read from the database using MySQL (Structured Query Language). This enables the database to be transmitted as a CSV (comma separated values) text file compatible with Microsoft Excel, including descriptors for each column. When users wish to find a word, variations can be provided based on all search fields, but for simplicity we used the single English word search only. This word can also be used to search the database of sentences for relevant examples.

This process has also been incremental. While the database was initially fairly simple, as we learnt of the growing needs both for tutors and students, more information has been linked to the table. Also while the table was designed for providing words lists, we wanted to join this to example sentences and include more grammatical constructions, which requires steadily more information be inserted. This approach was used as the initial users were very unclear as to the potential of computers to help their work, so the project developed as this understanding increased with further requirements. For example, we recoded the parser to search the example sentences first, then the alternative, then the single word search fields to provide a more reliable translator based on the extra information in the sentences and particular cultural expressions entered in the examples database.

The final task that we have completed on the word data at this stage was to collect grammatical information to teach how the words can be used correctly in sentences. However our understanding of the grammar is still developing as we continue teaching, so we are only learning a few steps ahead of the students. We initially worked in lessons with single words as the students were still at this level. Then we started to edit the example sentence we already had, replacing one or two words to change the context of the story and provide alternate examples. While some errors will be

introduced due to the significant linguistic differences among Aboriginal languages and between them and English, this formulaic method is an approach to reclamation promoted by some Australian linguists, such as Rob Amery (2000) and Christina Eira (2008). It was only later that we added in the language parser that is discussed below, based on the grammatical structure as we learnt it.

Structure of the learning environment

To design the learning environment, we collected existing text-based resources that had already been published in various New South Wales Aboriginal languages, including Gamilaraay and Gumbaynggirr, and we are grateful to the authors of these resources for sharing them. These worksheets have exercises involving placing words on drawings, finding words, and answering questions with words. These exercises gradually move from wordlists into writing sentences. Much of this structure can be repeated in the software environment (Figure 2).

Figure 2: Example exercise with wordlist to assist students.

There are various ways in which IT can assist teachers to develop such worksheets in their language. An interface was developed in Runtime Revolution[7] that uses the language database as a back end, and provides tools for teachers to develop documents that are displayed as worksheets for the students to complete. The first step is to create a wordlists for each lesson or set of lessons. This wordlist window is retained through the set of lessons in each file. We also went back and updated the online database to enable all words to be sorted into topics. Tutors can then select a topic list for each exercise set. This ensured that students were dealing with a smaller, more manageable vocabulary list, yet using these words in a wider range of exercises, such as use of location and verb tense.

The worksheets that are written by tutors are interactive and provide feedback on the sound of words, as well as a translation of the sentences students enter in answer to a question. If a group of words is selected by the tutor for more information, each sound file is concatenated to form the sentence to be played, which provides a simple example of the spoken phrase, although lacking expression. Thus it was important that we supply a simple sound file for every word. This information on words and sentences can be selected with links, which are created in a similar manner to Wiki pages. That is, when a tutor wishes to create a new exercise, they enter the title as a link, and a new page is created as a worksheet with the ability to add text, image and audio resources. Also the idea of linking words to a page providing their translation and sound file is similar to links on Wiki pages. Users can select different pages or worksheets in the exercise file, and do the exercises on each one. Hence the software is called LanguageWiki.

The exercises the software supports include diagram labelling and word selection exercises as well as audio practice and sentence construction. Tutors are supported in making the exercise sheets with a database search that is automatically activated when they select a word for translation. All the versions of the word from the database are displayed for tutors to select one per example and tutors can add new information to the data including sound files, alternative meanings, or a new word. At present the school version of the database is localised to each school, although regularly updated, initially due to firewall issues in accessing the database outside the Department of Education network, but also due to the need to maintain integrity of the original online database.

There was an added advantage of having the tutors develop the worksheets rather than using published material and possibly translating this material to a different language. The tutors appreciate that the work done by linguists provides a grammatical structure for their language, but they wanted to have a role in contributing the cultural context to their lessons. Also, where there was interesting material available illustrating cultural differences (such as stories recorded and stored at AIATSIS), the stories are someone else's and the tutor felt they would not necessarily do them justice by using them in their course. LanguageWiki is designed for each tutor or group of tutors to author their own exercises using the database as support for their work.

---

7 See www.runrev.com/home/product-family/

Electronic recording of words

While some tutors are fluent we have many who are still learning their language. Because the process of recording speakers usually requires some technical assistance it was initially difficult to encourage speakers to do this on their own. As we needed a sound file for every word we initially looked at speech synthesis technologies, in particular the Festival[8] system but realised this could not readily be utilised for these languages.

Many of the resources used in teaching are literacy based, and hence rely on a written representation of the language. We were using a phonetic script (ITRANS-3) to assist students to learn the new sounds of the language. Some sounds are unfamiliar to the ears and tongues of the learners, such as the palatal sound written as dj or j in some Aboriginal languages (often anglicised to approximate the first and final sounds in judge). We used this unfamiliar script as a way to encourage learners to try to approximate as closely as possible the way sounds like this were pronounced at the time when there was a large community of active speakers. This script has already been used in speech synthesis projects for Hindi (Kishore & Black 2003).

Also Wiradjuri literacy was being taught in Sydney at the time by analysing words into their component two- or three-sound syllables. We considered the use of syllables as units to be concatenated in speech synthesis. However the available software was designed for the smaller unit of language diphones, not syllables, and the dynamic nature of spoken language is much more complex than such a system would allow. Schultz and Black (2007) identified the need for large amounts of spoken data for analysis as one of the major difficulties in building up computer support for language processing. While tools such as Shoebox[9] can provide support for collection of such a corpus of language resources, we were working with the mostly text-based resources we had. We did not have such a large body of spoken, let alone translated, language so relied on collecting a series of simple word recordings from three speakers. As mentioned, these words were combined to form phrases using the software, which is simpler than trying to combine syllables with varying stress to form words. However we still had the problem that the voices used often varied across the group of words.

The concern we have with the re-created form of speech, is similar to the experience of language speakers supplying the pronunciation of words for the phonetic description used in text-based dictionaries. Attempts by non-speakers to repeat the pronunciation have often been incorrect and the speakers who supplied the pronunciation have come to distrust dictionaries as a result (Simpson 2003). Audio recordings of text resources are still necessary to develop these kinds of learning resources, and human speech should always be preferred over computer output.

Given the above problems with the phonetically generated sounds, and since another teaching program was being devised at the same time by Wiradjuri speakers on country, we changed to their phonemically designed orthography for consistency.

---

8 See www.cstr.ed.ac.uk/projects/festival/

9 See www.ethnologue.com/tools_docs/shoebox.asp

Grammatical parser

When students are translating from English to language in class exercises they need accurate feedback on their translation or expression attempts. For this purpose we built a text parser using the existing grammatical information to try and translate their entries in the answer box provided on worksheets. While the software Shoebox is used by linguists to gloss transcribed text, we were working from a wordlist without grammatical information so had to attempt to derive the grammar ourselves from the sentence examples and other documents. The parser was written in Python using the Natural Language Toolkit,[10] which extends the principles used in Shoebox parsing.

In developing a language parser for Wiradjuri we had hoped to create a system that was fairly generic and could also be applied, then adjusted, to the local Sydney language which is even less well resourced. At the same time we wanted to be careful that, if a student entered alternative but correct forms, this would be translated correctly. The wordlist we started with was in the form of a simple corpus with words tagged for eight parts of speech: verb, noun, pronoun, adjective, adverb, preposition, conjunction, and interjection. This was a classification understood by the Aboriginal speakers setting up the data. However, while some flexibility was built into the parser, because it was based in the grammar of English it was not ideal for the purposes of automatically parsing Aboriginal languages.

Of course all our goals could not be achieved. In particular there are limits to the number of words in the wordlist, and so the parser may not 'know' a word even though it may have been introduced by the teacher in class. Thus a student may be given feedback that they are wrong when they are actually correct in their attempt. The software learning tool was therefore designed so that when tutors create worksheet resources, if they enter a language word not already in the database, they are invited to update it.

Aboriginal languages make extensive use of suffixes to show tense in verbs and case in nouns, or as shortened forms of pronouns. The same suffix form can also have different meanings depending on whether it is affixed to a noun or a verb and the kind of relationships among nouns in a sentence. As in English the same form can be the root of both a noun or adjective and a verb. There is also a high degree of systematic variation of form in suffixes in most languages. To add to these difficulties, when we go back to the archives to verify the wordlist, there are a number of words, or parts of words, which are not translated in the simple sentence examples. These may be repetitions for emphasis, colloquial use or simply undocumented uses of words.

Another issue for translation that came up was the absence of an equivalent to the English verb 'to be' in Wiradjuri, as in most Australian languages. This meant that when the parser attempted translation from English it would have to ignore this, and to translate into English it would have to insert it at required points. But this would usually not match the original sentence the students were translating – an impossible

10 See www.nltk.org

task. Again we returned to the people working on the project to assist by providing more examples for students of known correct sentences. Different users can edit these examples online and can upload audio recordings of these examples.

Sound comparison

The final software application we wanted to include was to support students improving their sound to match the existing language speakers. However there is much more to learning good pronunciation than simply hearing a word or sentence spoken. An example of the sort of open source support which can be provided for sound development is the software Fluency.[11] However this system uses a text-based interface relying on knowledge of linguistic terms and is currently available only for English. Hence it would not be suitable for either school students or the adult tutors in this context.

The approach we have taken is to link the words which have a recording included in the database with an option for the student to record themselves, and then compare the two versions using Audacity sound editing software.[12] The result is very simplistic, in that it only displays a graph that corresponds to variations in volume and frequency, allowing users to infer syllable divisions. Audacity offers students the option of playing back the sounds separately by muting each channel alternatively so they can compare the movement of pitch and intonation in their speech to the saved form, as recommended by Anderson-Hsieh (1992).

There were a lot of issues with such a crude method. When comparing sentences, rather than single words, the pauses between words by the speaker differed from the concatenated file provided by the software. If the student spoke softly or slowly, it was hard for them to match this with the original recording. Also both significant and insignificant differences in the sound were equally visually noticeable to the students.

However, this does provide visual and audio feedback to the student that they can practise with in their own time. Generally the students tend to mimic well, so the speed of the word tends to match that of the recording for single words. The initial response has been positive. The main limitation to enabling this feature is that it requires that the school computers possess a microphone for students to record themselves.

Speech and song-writing

The need for more extended texts for learning suggested we use existing technology for innovative purposes. One of the most used resources in Indigenous language teaching has been Microsoft PowerPoint slides with embedded sound files to present stories for the classroom. In providing material for the Bankstown Elders' Group, George Fisher used Microsoft Word files with embedded sound files. In Dharug, linguist Amanda Oppliger and Richard Green worked together with researchers at the University of

---

11 See www.lti.cs.cmu.edu/Research/Fluency

12 See audacity.sourceforge.net

Sydney to establish a dictionary accessible using mobile phones (see Wilson, this volume), and now some speakers share knowledge about Dharug words and language through SMS (see Green, this volume).

We also needed a robust song learning tool, so we purchased Finale Songwriter.[13] This has been used to provide Dharug songs for children to sing in schools. This work was possible as Richard Green is fluent in the language, and has an understanding of the poetic nature of the constructions. Students learning Aboriginal languages as part of their IT work were inspired by witnessing Richard develop the translation for the song 'O Come All Ye Faithful' in a day, scanning the new song correctly to the original tune.

A recent performance in Dharug involved students from schools around Sydney, including children from the Dharawal program in south-east Sydney. They responded to the new language with great enthusiasm, showing that in the process of teaching one Indigenous language we open our children to the knowledge and skills to communicate in many related languages. This also suggests opportunities to share resources among tutors of these languages. This similarity between the languages needs to be further researched, as students in Sydney schools will often be learning a language that is not that of their parents or grandparents, but is similar enough for them to use this knowledge later to reclaim their own language.

Extensions

The focus of the next stage of development is to enable the functionality of such programs directly on the web, rather than simply using local versions on single computers. This will enable wordlist updates and new features to be installed without having to distribute new versions to the users, or to bypass a firewall for remote access.

In developing media-rich environments for learning language we need access to more resources that are relevant to particular themes and topics. Developing standard tagging of sound and video files in the future would help this process. The next feature needed for the website is to enable users to upload and share detailed material in language. Audio resources can be recorded using internet telephony software such as Skype, or locally on a home computer. There is a need for video and sound annotation applications such as an online version of Elan[14] to encourage the speaker to provide some text information or translation while uploading, even if just a description of the topic. Speech-to-text tools exist, but require training to the individual speaker and are only available for major world languages, hence are not realistic at present. Then programs such as LanguageWiki or games software with a database back end, could integrate these resources into a learning environment based on a series of topics.

---

13 See www.finalemusic.com/SongWriter/

14 See www.lat-mpi.eu/tools/elan/

The transcriptions will need to come from language users, as well as linguists,[15] and hence the interface for such tools will need to consider various knowledge and cultural aspects, such as who has the authority to hear the information, and in what context it may be repeated. For instance some explanations may relate to images of kinship, or specific locations on land, so need to be presented in that context only. In particular there is hesitance by language speakers to share their material in an open context as this may risk losing control of the information residing in the language. What we are proposing is the development of easy-to-run software that provides a suitable learning environment, like the *Mayalambala: Let's Move It*[16] posters and cut-outs developed by Muurrbay Aboriginal Language and Culture Co-operative. Within such a framework, exercises from different languages, using a small selected vocabulary, could all be linked to relevant pictures or videos.

While language resource sharing sites such as that developed by Ngapartji Ngapartji[17] have been around for a few years (see Sometimes & Kelly, this volume), these could be enhanced by becoming more conversational in form (perhaps like Facebook or MySpace) and using audio more than text. This would allow the creation of interactive sites to share language resources and will enable learners to keep up to date with what has been created. The cultural and technical aspects to be considered in the interface requirements include: a focus on sound rather than text; the need to reduce complexity particularly for people to upload; and providing a rich environment where users feel they are contributing to a large body of work, rather than only adding a small drop to an diminishing pool of resources.

Conclusion

The market for computing support of language recording and teaching is growing to include members of Indigenous communities. There is a desire by community members to have direct access to recording and storage equipment which ensures they control how it is presented to the public, and how it can be used in learning. This will bypass, at least for the present, the challenge of developing sophisticated algorithms to support partially documented languages with few speakers and a small electronic corpus, and where few linguists have the time to provide full annotation for existing or future resources.

At the same time there is a growing interest in Aboriginal communities for IT products which support their use of their own language. It has been encouraging to see tutors who are often wary of computers become interested in taking up the opportunity to use these resources and tailor them to their teaching needs. And, while many

---

15 Thieberger discussed the difficulty of getting linguists to create reusable records of the languages they record in 'Does language technology offer anything to small languages?' (Australasian Language Technology Association, 2007).

16 See www.muurrbay.org.au/publications.html

17 See ninti.ngapartji.org

students are already familiar with computers and use them regularly, it is good to see the response of Indigenous students to discovering their culture and language is also being supported by IT.

Finally, this work has provided the opportunity to look at the similar needs across Australian Aboriginal languages undergoing revitalisation, which suggests that work on technology developed to support one language may in fact provide support for others, so the gains could be substantial.

Acknowledgements

References

Amery R (2000). *Warrabarna Kaurna! reclaiming an Australian language*. Lisse: Swets & Zeitlinger.

Anderson-Hsieh J (1992). Using electronic visual feedback to teach suprasegmentals. *System*, 20(1): 51–62.

Australasian Language Technology Association (2007). *Australasian Language Technology Workshop 2007: invited Speakers* [Online]. Available: www.alta.asn.au/events/altw2007/alta-2007-speaker.html [Accessed 12 September 2008].

Bird S & Simons G (2003). Seven dimensions of portability for language documentation and description. *Language*, 79: 557–82.

Christie M (1985). *Aboriginal perspectives on experience and learning: the role of language in Aboriginal education*. Waurn Ponds, Victoria: Deakin University Press.

Christie M (2004). Computer databases and Aboriginal knowledge. *Learning Communities: International Journal of Learning in Social Contexts*, 1: 4–12.

Dixon RMW & Blake BJ (1979). *Handbook of Australian languages*. Canberra: Australian National University Press.

Dyson LE, Hendriks M & Grant S (2007). *Information technology and indigenous people*. Hershey, PA: Information Science Publishing.

Eira C (2008). *Language development on the ground: three practical sessions.* Paper presented at the Indigenous Languages Institute '08: Bayala ngarala. 10 July 2008, University of Sydney, NSW.

Groome H (1995). *Working purposefully with Aboriginal students.* Wentworth, NSW: Social Science Press.

Harkins J (1994). *Bridging two worlds: Aboriginal English and cross cultural understanding.* St Lucia, Qld: University of Queensland Press.

Harris P (1991). *Mathematics in a cultural context: Aboriginal perspectives on space, time and money.* Geelong, Victoria: Deakin University Press.

Kishore SP & Black AW (2003). Unit size in unit selection speech synthesis. In *Eurospeech 2003* (pp. 1317–20). The 8th European Conference on Speech Communication & Technology, 1–4 September 2003, Geneva, Switzerland [Online]. Available: www.isca-speech.org/archive/eurospeech_2003 [Accessed 9 June 2007].

Kutay C & Mooney J (2008). Linking learning to community for Indigenous computing courses. *Australian Journal of Indigenous Education,* 37S: 90–95.

Nakata M, Nakata V, Byrne A, McKeough J, Gardiner G & Gibson J (2008). *Australian Indigenous digital collections: first generation issues.* Sydney, NSW: University of Technology Sydney ePress [Online]. Available: epress.lib.uts.edu.au/dspace/bitstream/2100/809/1/Aug%2023%20Final%20Report.pdf [Accessed 10 January 2009].

Schultz T & Black AW (2006). *Challenges with rapid adaptation of speech translation systems to new language pairs.* Paper presented at the International Conference on Acoustics, Speech, and Signal Processing. 14–19 May 2006, Toulouse, France [Online]. Available: www.cs.cmu.edu/~awb/papers/ICASSP2006/0501213.pdf [Accessed 11 June 2008].

Simpson J (2003). Representing information about words digitally. In L Barwick, A Marett, J Simpson & A Harris (Eds). *Researchers, communities, institutions, sound recordings.* Proceedings of the Digital Audio Archiving Workshop, Sydney, October 2003. Sydney: University of Sydney [Online]. Available: www.paradisec.org.au/Simpson_paper_rev1.html#Abstract [Accessed 10 July 2008].

Troy J (1994). *The Sydney language.* Flynn, ACT: J. Troy.