A Template to Use Existing Digitised Data for a CALL Program Shell for Indigenous Languages: A Community Centred Approach

Virginia Westwood
Faculty of Education
Monash University, Australia
vwwes1@student.monash.edu

Abstract: Many strategies have been employed to maintain or revitalise Australian Indigenous languages by linguists and communities. Despite this, the number of spoken languages continues to fall. There has been a call for sustainable Computer Assisted Language Learning courseware. This paper demonstrates a software template and shell application. The template makes use of digitised data compiled by linguists, reading the underlying files of a language dictionary from one community in Australia. The data is stored in XML files, and the interface allows the community to edit and import new data and multimedia resources. The data is then read by a pedagogically-sound program shell. Of importance to the success of this project is the Participatory Action Research approach with the researcher/educator, community and linguist. The template, the application and the process will be evaluated qualitatively. If successful, the transference of the project to other community languages will be tested.

Introduction

It is estimated that there were over 250 different languages spoken in Australia when the country was colonised by European settlers in the 18th Century (Dixon 2007). Of these, only about 145 are still spoken and the vast majority, about 110, are in the severely and critically endangered categories (AIATSIS 2005). The importance of language in the survival of culture is paramount and much has been done to preserve them. However, the policy and funding support in Australia has not been generous to Indigenous language revitalisation, frequently leading to short term projects that lapse when governments change or funding cycles cease. In some states of Australia, there are school programs for languages, and Australia wide there is a community movement to create Language Centres. For some languages, solutions have been developed from community input, with assistance from linguists or educators to support language maintenance initiatives (Hobson et al 2010). These have included books, dictionaries, posters, audio and video tapes, teaching by Elders and other speakers of the language, and some interactive digital resources. There are many factors affecting the success of language revitalisation programs (Walsh 2005), but too often current and motivating educational resources are not available, leaving the teaching to a declining number of speakers with competing demands on their time (Hattersley 2009).

Since the time of colonisation over 200 years ago, linguists have been documenting the Indigenous languages of Australia. More recently, audio recordings have been made, and some of the earlier tape recordings have been digitised and archived. Dictionaries have been created from linguists’ documentation and fieldwork with current speakers and much of this work is now in digital format, and available in centralised or local repositories. In many cases, languages are available through secondary dictionary documentation programs such as Lexique Pro or Toolbox (http://www.sil.org), Miromaa (http://www.miromaa.com.au) and KirrKirr (Manning et al 2001), but these products are designed as visualisation and research aids rather than learning materials, and are generally used by linguists rather than Indigenous community learners. While research and documentation is of vital importance to language maintenance, it does not directly help communities to learn their language. On the other hand, technology can be utilised to significantly increase the value of this language documentation work by repurposing the content and re-presenting it in the form of motivating, interactive multimedia CALL (Computer Assisted Language Learning) resources.

Allyson Eamer argues strongly for the use of technology in the revitalisation of languages which have “become endangered within a dominant cultural context” (Eamer 2010). Where CALL programs have been developed in Australia (see, for example, Auld 2002, Nathan 2007), they are often criticised as being expensive ‘one-off’ endeavours (Hobson 2010) not allowing for updating as the language grows, technology advances, community-specific information is desired, or for extensibility to other languages. However the landscape is changing.
Computers are now widely available in remote areas and, even if the Internet is not available yet, stand-alone browser-based applications can be quickly produced and easily distributed using Rapid Application Development tools such as Flex (http://www.adobe.com). Furthermore, it is the author's experience that even expensive, one-off products are successful in the more widely spoken languages, perhaps because there are numerous such products available, and the market is mature in their use. In a less widely spoken language, there are few resources so that there is no 'critical mass', and learners and teachers are not familiar with the media or the approach. Hence, one-off products are less successful. If such products could be produced from templates by the community that speaks the language, there would be more of them and this would not be such a problem. As noted in the National Indigenous Languages Survey Report 2005 (AIATSIS 2005), "[f]or some years, language maintenance programs and Language Centres have been calling for an open-source set of templates that can be used and adapted by them, but such a resource has not been made available yet."

This paper demonstrates a software template and CALL shell application which can access existing digital dictionary data prepared by linguists. The programs are adapted from the author's existing applications using Agile methodology (http://agilemethodology.org) within a Participatory Action Research (PAR) model (Kemmis & McTaggart 2000) to test the transference of the applications to the Indigenous context. The partnership for the PAR research comprises an Indigenous community, author/educator and linguist. The research proposed is the operationalisation of a doctoral study on CALL instructional design and pedagogy in different contexts.

**Background**

CALL has been used worldwide for many years to learn and teach languages, yet there has been little application of CALL in the maintenance of Indigenous languages. Generic CALL authoring programs usually are simply quiz makers; those that produce more contextualised learning require specialist language and teaching knowledge. The prototype for the project will be based on the existing instructional design and pedagogy demonstrated in the CDROM and online software programs developed by Protea Textware (http://www.proteatextware.com). These have proven successful over many years in teaching English and literacy to diverse learners (Brown, I. 2006, De Fazio 1997, Siennicki, B. 2006, Turnbull 2004, Waterhouse 2005). They incorporate five principles that Reyhner and Tennant (1995) assert need to be addressed in any language-teaching programs: - putting primary emphasis on communication, not grammar (although grammar is not dismissed); - using context that is real or at least realistic; - processing content of high interest to the learner; - adjusting the pace of instruction to the students' progress, including moving from simple to complex (generally speaking), emphasizing speaking over speaking correctly, and putting comprehension before completion; - and finally - correcting students through modelling. While individual learning styles may differ with a group, it is contended that learning styles do not differ significantly between groups. Thus the instructional design developed to be used by both children and adults worldwide will be used in the prototype CALL application for Indigenous learners.

There are several projects in Australia and internationally which have or are addressing aspects of the need for template-based CALL resources for Indigenous or other endangered languages. For example, Ward in her thesis (Ward 2001) devised a generic template which could be used to generate CALL resources for any language, using an El Salvadorian language, Nawat, as an example. The template is based on a reverse-engineered version of Hot Potatoes, an application designed primarily for more commonly spoken (Western) languages, utilising lessons and a limited set of textual quizzes. Ward’s application stores data in XML(eXtensible Markup Language) files, using XSL (eXtensible Stylesheet Language) for transformational logic and to create HTML files for the interface. The template requires the introduction of language from scratch in creating the exercises, and requires an experienced linguist or speaker-educator to construct a course. Ward (2001) comments on the steep learning curve for her in becoming proficient in the required computer languages, and logistic difficulties in recruiting speakers and managing the technology in a technologically-poor community. This approach combines the knowledge of the linguist and the technologist. In Taiwan, Yang & Rau developed a framework for archiving, documenting and creating learning materials for a Batanic language, Yami (Yang & Rau 2005) which they proposed could be used as a model for other languages. Audio and image data were collected and stored as XML files and in a relational database, while work proceeded on the orthography of the language. The learning materials used the annotated image database simply as a tool to help the learners understand the meaning of words and expressions rather than creating a specific CALL program. This approach combines the knowledge of the educator, the linguist and the technologist. Cat Kutay et al (Kutay, Fisher & Green 2010) describe a project which collected existing data from
various sources for two Australian languages, Dhurag and Wiradjuri, currently being taught at college level. The project initially collated dictionary entries and audio samples into a system which they intended to become “a body of IT resource templates and applications that provided a learning environment that could be used for similar languages” (Kutay, Fisher & Green 2010). In this project, the emphasis is on the technology, extending to research on speech recognition and Natural Language Processing. This approach combines the knowledge of the community, the linguist and the technologist. While some of these projects have worked with the community whose language survival is of concern, none have based resources on established CALL methodologies, considering instructional design and pedagogy in a non-Western context. Furthermore, most of these projects rely on collecting new digitised data for their content, rather than making use of the extensive existing digital language documentation.

Based on 17 years of CALL program development for teaching English, and information from the literature, the author has identified a gap of knowledge in the revitalisation of endangered languages where the four aspects of instructional design, pedagogy, technology and community partnership overlap. This leads to the testing of four aspects which are seen as being critical to the successful creation of templates which can be reused for creating sustainable, educationally-sound CALL resources for Indigenous or other endangered languages:

1. Cross-sector partnership with community, linguist, educator and programmer.
   The community participants include Elders, Language and Culture Hub employees, community representatives, school and resource centre workers. It is their language and their knowledge which makes the content of the CALL program. Their collaboration is required to authorise, create and contextualise the content through cultural, language and digital representations as well as participate in the design and carry-through of the process. The linguist will provide the technical language skills. The educator dictates the pedagogy of the CALL shell program. The programmer (author) incorporates instructional design and structures the programming and digital resources for the template and the CALL shell. All partners are seen as co-researchers and integral to the research process, impact and outcomes.

2. A pedagogically-sound CALL shell program that reads content from underlying data files.
   By separating the content from the instructional design, we can create a template that is transferable to other languages. The pedagogy of the shell program should reflect Indigenous ways of knowing (Yunkaporta 2010) as well as established CALL methodologies.

3. Access to existing digitised language records, including sufficient language and grammar documentation.
   The project should make use of existing digitised data as far as possible. This reduces cost, increases authenticity and speeds up the development process. It also maximises the use of resources contributed by linguists in the field.

4. An editing template that can be managed by the community.
   The template must be able to be used and managed by the community, without extensive training or additional support.

These four aspects encapsulate most of the 14 points advanced in Walsh’s ‘wishlist’ for a successful language revitalisation project (Walsh 2010). A partnership was formed between the Indigenous community’s Language and Culture Hub management, the linguist and the author. The Indigenous community in North-west Australia has worked actively to maintain their language, of which there are about seven fluent speakers remaining, and has established a Language and Culture Hub for this purpose. The linguist has been working with the community for over 20 years, following original documentation of the language by a PhD researcher linguist (Stokes 1982), and has compiled a dictionary in Lexique Pro including audio recordings and images. The community has made this dictionary available for use in this project. The author is a co-developer of award winning CALL materials which are based on sound educational principles (see, for example, Westwood & Kaufmann 1995a; Westwood & Kaufmann 1995b, Kaufmann & Westwood 2004, Kaufmann et al 2007). The programs use the communicative approach, presenting content in context (Bangs & Cantos 2004).

The Research Project
The project was first conceived when the author was approached in late 2009 by the linguist working with the Indigenous community in North-west Australia. It will test whether established instructional design and pedagogy can be successfully transferred to a new Indigenous language and cultural context, and elucidate changes necessary to achieve the transference. Sub-questions ask if Participatory Action Research (PAR), following the model of Kemmis & McTaggart (2000), is a suitable vehicle for this; if effective CALL programs can be made from existing digital language documentation; if the making of such programs is sustainable in the community; and can technology accommodate Indigenous ways of knowing (Yunkaporta, 2010). The data collected (field notes, changes in instructional design, monitored use of CALL programs in community, video evidence of users of CALL programs, think aloud recordings, etc.) will be assessed qualitatively and will lead to a new approach to instructional design and user interface/user experience (UI/UX) respecting indigenous peoples’ ways of knowing and learning. If the initial language resource created within the community is successful, it will be used by the community to build CALL materials of specific cultural significance, as well as contemporary language as it evolves. It will also lead to further work in the transference to other communities wishing to revitalise or maintain their language.

Methodology

Four-way cross-sectoral collaboration is proposed as being vital to the success of this project, and is a factor most commonly missing from other projects of this nature (Kutay, Fisher & Green 2010, Ward 2001, Yang & Rau 2005). The contribution of the community participants and the linguist as co-researchers to the design of the process and the development of the programs is also unique. The linguist’s contribution is vital in the creation of the underlying dictionary and providing grammatical and orthographical support. Using PAR, the author will build the template and CALL shell program based on models of her existing commercially successful applications that the community can later use to create their own learning materials.

The project was first conceived in September 2009. The partnership was formed in June 2010 and the work formally commenced in January 2011 with the author beginning the development of the template. At the same time, a plan was devised by the researchers to map out the progress of the project. The prototypes of both the template and CALL interfaces will be completed by November 2011. Meanwhile, content collection by the researchers will commence in July 2011. An important part of the process involves capacity-building within the community through semi-formal training in digital processes so that they are able to continue to create new CALL resources as they see fit. This includes capturing and managing images, animation, audio and video resources, as well as specific training in the use of the template to import and edit text and multimedia resources. This training with the Community Language and Culture Hub in digital resource collection and processing and, with the linguist, in the use of the template (to perform as ‘editor’) will start in November 2011. The appraisal by the author, community participants and linguist will begin in November 2011. Community training and media collection will continue through to June 2012, with continuous cycles of template and CALL interface development. During this time, the participants will be involved in posing and answering questions as to how the template and program may better interface with Indigenous ways of knowing. This will overlap the community use of the template, collection of media and appraisal of the two applications and the process by the researchers. Data will also be collected by the author throughout the 18 month project for full reporting and evaluation of the process, and of the programs by qualitative (Hubbard 1996) and quantitative (Khodabandeh, Afshari & Manian 2010) methods for a PhD dissertation. The data will include the author’s notes and reflections, field notes and records of comments, feedback, think-aloud protocols and interviews with the researchers and observers.

The template functionality will be determined by the needs of the CALL program, and so the two applications will be developed in tandem. An Agile development methodology, following the work of the author and Heather Kaufmann (Westwood & Kaufmann 1999) and recent definitions of the methodology (http://agilemethodology.org), is being used for both the template and CALL shell.

The CALL interface will also be built in Flex/FlashBuilder, but will be compiled as SWF (ShockWave Flash) modules, allowing for cross-platform distribution, as well as online use, should this become a viable solution in the remote areas of Australia where the Indigenous communities are largely located. This medium has proven to be stable over many years of CALL development. The CALL shell will be an interactive multimedia program consisting of elements of several of the author’s previous CALL programs. For example, there will be an alphabet module with presentation, matching, construction, listening cues and keyboard activities (see Fig. 1); beginner
spelling activities (look, listen, speak, write) (see Fig. 2); intermediate words presented in context at the sentence level with hotwords indicating grammatical structures, definitions and gloss (see Fig. 4), look, listen, speak, write activities and usage notes; help (see Fig. 3) and feedback. There will be a module for stories with extended passages of natural speech with associated activities such as comprehension questions, cloze, spelling and dictation (see Fig. 5). Additional elements such as listening discrimination, voice recording, vocabulary exercises and learning and testing modes will be included.

Figure 1: Examples of activities from The Alphabet (Westwood & Kaufmann 1995a).

Figure 2: Example of ‘look, listen, speak, write’ activities from The Alphabet (Westwood & Kaufmann 1995a).

Figure 3: Example of visual help from The Alphabet (Westwood & Kaufmann 1995a).

It is expected that the style of interactions will be redesigned on the basis of community appraisal of the CALL resources and the availability of technology and other factors which may favour more group work than individual. In addition, features of the interface design will cater for low computer and language literacy at the beginning level, using visually demonstrated help, large graphical targets and consistent screen design. Following the PAR model,
the community’s knowledge of Indigenous pedagogy will determine the use of representational iconographies and images as well as structure of the learning resource.

The template is being built in Flex/FlashBuilder (http://www.adobe.com) as an AIR (Adobe Integrated Runtime) application which can be deployed to PC or Mac platforms. The underlying data files are built as XML files by the template, containing textual data including paths to externally held multimedia resources. The template will be modelled on input/editing templates designed by the author for previously localising English language CALL programs in Toolbook (http://www.sumtotalsystems.com) (Westwood, 2005). However, it will be modified to manage XML data (as opposed to database data in previous usage) and greatly extended to cater for the variety of modules in the CALL shell. The linguist will be involved in working with the community to create new language texts to be included in the content where required. The editor is the person from the community who has been trained by the author to use the template. This may be the same person in the community who is trained in digital processing to collect and manage the digital resources for the project. The template includes translation of media resource names to a common set that can be used by the CALL application, but the actual media will have to be converted by the editor to the appropriate formats using standard image, audio and video conversion utilities outside of the template. It is possible that some of these tasks may be managed by the template.

Figure 4: Example of screens from The Interactive Picture Dictionary (Westwood & Kaufmann 1995b).

Figure 5: Example of screens from Issues in English 2 (Westwood & Kaufmann 2004).

The template makes use of digitised data compiled by linguists, reading the underlying files of a Lexique Pro language dictionary from the community in Australia (see fig. 6).

In the prototype, the data to be included is predetermined by the author and includes entries for each dictionary headword in language, alternatives, gloss, context sentence, topic, audio file names for word and sentence, image
names, and additional notes. Later iterations may expose these decisions to the editor. The template writes the selected data into XML files and initially displays the words and their gloss by topic (see Fig. 7).

Figure 6: Example dictionary file data.

Figure 7: Initial input screen for importing dictionary data and organising at word and topic level.

The initial screen allows the researchers to choose words from the underlying dictionary to be used in the final CALL product, to rearrange the words in topics, and to add, edit or delete topics and words (see Fig. 7). Buttons and hypertext are used for navigation to the additional screens of the template. The template will extract an alphabet from the data, which can be edited through a further screen, with audio files for sounds and names attached, and the appropriate XML files will be created and populated. Through the alphabet screen, keyboard mapping can also be chosen for characters not displayed on the standard keyboard. The editor can choose a topic and work on the content for the vocabulary activities through the template interface. Existing resources for each vocabulary item, or headword, can be immediately seen, and what needs to be added can be determined (see Fig. 8). Audio, images, animations and videos are captured by the editor or other participants from the community Language Hub as the editing progresses. They form a library and those used in similar contexts are identified by the template and displayed to the editor for attachment to the headwords. Hotwords (to illustrate sentence text) can be chosen and associated resources allocated from the library in a similar manner by the editor. New sentences for spelling and dictation activities can be added and syllabification and stress determined by the editor or linguist. Lists of words commonly used in the topic sentences are available for reference to aid in recycling of vocabulary. Notes on word usage drawn from the original dictionary file can be added or expanded by the editor or linguist. Through different screens, the editor can choose words for the beginner spelling activities. These will already have resources attached, as described above, and the words marked in the XML files for that component of the CALL program. The template maintains the underlying XML data files as editing changes are made.
At any time the CALL program can be run to view the results of the editing, as it simply reads the content from the XML files produced by the template. In this way, the CALL program can be bundled for distribution and exposed to more community participants for appraisal and evaluation, resulting in further development cycles by the author and the editor. The CALL resource meta-data (language name, ownership, permitted usage, contributors, acknowledgements, etc) can be added through the template.

Results

As this project is being undertaken using a Participatory Action Research methodology, the project will be evaluated over many iterations by the participants. The process itself will be evaluated against the initial questions. As both the CALL shell and the editing template will be developed using an agile methodology, technical evaluation and testing of the software is a continuous process by the author and editor. It is expected that the participants will raise questions and make suggestions regarding the process and the technology that will be incorporated into the PAR process. However, an overall evaluation based on operational fit, learner and teacher fit, as described by Hubbard (1996) will be followed. An empirical evaluation of user satisfaction with the CALL materials will be undertaken in the manner of Khodabandeh et al (Khodabandeh, Afshari & Manian 2010). The evaluation of the template as an application will cover useability functions and will be qualitative, focussing on the appraisal and use by the editor and linguist, taking into account field notes, programming notes, interviews, suggestions and feedback, usage and useability. The research project will be reported using a narrative presentation so that the voices of the participants (the community workers, linguist, author, program) are all heard. The results will also be written up in a manner that is accessible to the participants and members of this and other interested communities.

Project risks

The success of this project requires commitment and participation by the community, the linguist and the author. The work with this specific community also depends on the availability of government funding for the ongoing maintenance of the Language and Culture Hub, which cannot be guaranteed. It requires skills training in the community in computer literacy and digital media capture and processing which will be provided by the author. It also requires that the community are actively involved in the PAR cycles to create the CALL resources so that they feel ownership of the resultant learning materials, and are empowered and motivated to continue to use the skills and template to continue to create new material. Partners from the community will also be trained in language skills by the linguist. The availability of the linguist to work on the project is ultimately outside of the control of the author. Community and family matters can often take precedence over work which is still essentially a Western construction and this can pose risks to the continuity of the project and adherence to the agreed timeline. The involvement of the community in planning the work tasks and timelines within a cultural context should obviate problems of this nature. The project may be too ambitious for the time frame allowed, and some elements of the CALL program may have to be sacrificed. The passing of speakers of the language and the willingness of Elders to continue to contribute to projects such as this may also constitute a risk to the gathering of authentic materials. There may be problems with representations (photos, videos, voice recordings) of persons who may pass away during the course of the project.
of the project. There will need to be an adequate quantity of materials, or permissions for, or avoidance of, pictorial representations where required, to mitigate this risk. Walsh (2010) highlights the necessity for community cohesion in regard to linguistic decisions. This has been demonstrated by the community participating in this research through successful workshops to resolve orthographic and spelling issues of the written language (Hattersley 2009). However, differences may still arise in the community which prevent work proceeding. Finally, the work depends on the author gaining University and State government ethics approval to undertake the research aspects of the project with the Indigenous community.

**Future work**

The final CALL program and use of the input template will remain with the community at the end of the project. It could be extended to include higher level or more complex language elements. It could also be modified to import a wider range of dictionary or linguistic data. It will be used by the community for creating further local CALL programs. If successful in this one community, the project will be continued, looking at transference to other Indigenous languages. If such transference is successful, the model would have application for Indigenous, minority and endangered languages worldwide.

**References**


