

*Aidan Wilson*¹

Abstract

Owing to the disproportionately low level of literacy in remote Indigenous communities, especially in Indigenous languages, printed books are perhaps not the most appropriate form of delivering language-learning materials such as dictionaries. Electronic versions based on computers are more useful. However the availability of computers, and consequently computer literacy, in remote Australian communities is still very low. Mobile phones are a much more common form of technology. Unfortunately mobile phones generally only allow small applications, meaning that most content expected in a reasonable language learners' dictionary must be jettisoned. This paper proposes and documents a method of dictionary delivery that takes advantage of the flexibility and usability of computer-based dictionaries, as well as the portability of mobile phones. This process entails maintaining a single dictionary file that can be exported to dictionary visualisation programs and applications that can be installed on a mobile phone, as well as a number of other formats in various media. Computer-based resources may contain as much information as is necessary in a format that can be navigated easily, while a mobile phone-based version will contain only a reduced range of the original content, although it will be available to the user without the need of a computer.

Dictionaries are invaluable resources for language revitalisation; they aid linguists, language workers and teachers and, most importantly, provide critical access to information for the language learner. Several studies (Corris et al. 2000, 2004; Nesi 1999) have shown that electronic dictionaries can be much more accessible and engaging for users than traditional printed dictionaries, and suggest that they tend to be used much more frequently and for longer periods of time than paper dictionaries. Electronic dictionaries can offer ways of organising content and finding

¹Department of Linguistics and Applied Linguistics, University of Melbourne.

entries beyond the traditional method of searching through an alphabetically sorted list of headwords. They can also include multimedia content such as sound, images and video. With traditional printed dictionary materials however, it is only possible to include images.²

Not all electronic dictionaries, however, are so useful. An electronic dictionary consisting of formatted text is only as useful to the user as a printed dictionary, perhaps with the added benefits of being searchable and vastly more portable. Dictionaries consisting of marked-up text that is properly machine-readable allow for more electronic functionality than raw text; fields can be searched independently and content can be linked together through hyperlinking, allowing the user many more ways of navigating the content. Traditional dictionaries by contrast force the user into a reduced set of methods.

There are further considerations regarding the representation of data and the method by which dictionaries are delivered. Electronic dictionaries can be presented in a number of ways: online using a Hypertext Markup Language document; in a specialised electronic dictionary viewer such as Kirrkir (Manning et al. 2001); or using mobile phones (McElvenny & Wilson 2009). As discussed below the optimal way to compile and deliver dictionaries in the remote Australian context, and possibly in other areas of extreme language endangerment, is probably a combination of computer-based resources for use within the classroom and smaller mobile phone-based resources to which the user has continual access.

The purpose here is to recognise and take advantage of a technological niche to aid the potential reclamation of Indigenous languages alongside other language revitalisation efforts.

Master dictionary file

The key to delivering dictionaries in multiple formats without having to independently maintain a number of different versions is to preserve a master copy of the dictionary in a format that is completely machine-readable, from which the other versions can be derived as needed. It is important for the longevity of the content of the dictionary that the format chosen for this be stable and not become obsolete in the future.

This master dictionary file is virtually unlimited as to size; it can contain high-resolution images, high-quality audio recordings of individual words or example sentences, and perhaps even videos. It is also able to contain lexicographic and metalinguistic information well beyond the actual needs of most learners' dictionaries.

The purpose of the master dictionary file is in fact not to be a dictionary in itself; it is not intended to be used by anyone apart from the linguist or lexicographer. Instead its purpose is to serve as the centrally maintained file from which other purpose-

² The ideas discussed here are a result of an ongoing project in collaboration with James McElvenny to produce free electronic dictionaries for minority languages.

built dictionaries will be derived. A dictionary intended for linguists working on the language, for instance, may contain grammatical information, pronoun paradigms, scientific classificatory names for flora and fauna, and recording numbers and time-codes for example sentences so that the researcher can check the source data. A learners' dictionary will likely include none of these but will include sounds to aid the learners' pronunciation of new words and images to identify particular plants and animals. Each of the dictionaries will be exported from the master dictionary file retaining or ignoring specific content and formatting it as required.

In maintaining master dictionary files we have adopted a markup language that is commonly used, is very well documented and will remain, in principle, readable well into the future. Extensible Markup Language (XML) is essentially text that contains tags or codes that inform the reader, human or machine, of the content's structure, how specific content is related, and what each piece of content is; whether it is a headword, a gloss, an example and so on (World Wide Web Consortium 2008).

Field-oriented standard format

Currently the most common markup language used for creating electronic dictionaries is Field-Oriented Standard Format (FOSF), more commonly known as *backslash codes*.³ FOSF is the syntax used in programs such as Shoebox or Toolbox⁴ and Lexique Pro⁵ which remain the most common dictionary creation and display tools available to linguists and language workers. As a result many electronic lexical databases in existence – possibly the vast majority – are encoded in FOSF. Backslash codes are highly human readable as long as the alphanumeric codes are easy enough to interpret or are clearly documented, but the only programs that can computationally interpret FOSF are the programs mentioned above. Apart from this there are a number of disadvantages to FOSF that encouraged us to employ a more sophisticated and standard markup language.

The syntax of FOSF consists of a backslash `\`, an arbitrary alphanumeric code and a space, all followed by the actual content. For instance the headword content of a backslash-coded dictionary may look like `\lx headword` and a gloss will be `\ge gloss`. The content of the code, be it headword, example or gloss, is tacitly assumed to continue until the next carriage return; the start of the next line. Thus an example sentence in FOSF will look like `\xv this is an example`. The syntactic fact that a carriage return is the indicator for the end of one piece of content and the start of another has a serious corollary for the formatting of dictionaries: content cannot be embedded inside other pieces of content nor be grouped, which is important for distinguishing among different senses of a particular word or explicitly grouping a vernacular example and its gloss.

3 The acronym FOSF and the term backslash codes are used here interchangeably.

4 See www.sil.org/computing/toolbox/

5 See www.lexiquepro.com/

It must be said though, that FOSF had a number of benefits for which dictionary writers can be grateful. Firstly it is exceedingly easy to read and manipulate even without a program such as Toolbox, as the structure is entirely transparent and can be written with a mere text editor. Secondly using a program like Toolbox to create and manage dictionaries encourages a level of machine-readable consistency that other formats do not, although human error in selecting and using the correct codes can be common. Another benefit is that backslash codes enjoy a level of institutional support from the major source of field software, the Summer Institute of Linguistics, such that software is available that can quite easily convert backslash-coded text into formatted dictionaries ready for print. The Multi-Dictionary Formatter (MDF) for example, only requires that the alphanumeric code chosen be consistent with their specifications (Coward & Grimes 1995).⁶

Extensible markup language

XML differs significantly from FOSF in the way that data is structured. Rather than carriage returns marking the boundaries of content, information is delimited on either side by explicit tags. Data can also be embedded recursively by placing tags inside other tags, which is especially useful for lexical databases in that certain information can be explicitly grouped. An example and its gloss, for instance, can each be structured hierarchically within another dedicated tag to ensure that they stay together.

Despite the benefits of using such a structured and flexible markup language, XML has its distinct disadvantages. In particular it is nearly impossible for the untrained person to read, and editing XML without software that interprets the structure can have devastating consequences for the validity of the document. Although XML-editing software is readily available with varying degrees of quality, it generally amounts to highlighting – colour coding the machine-readable tags so the user can safely avoid them. Adding or removing tags or any other form of structural editing generally requires a more sophisticated XML editor or a knowledge of XML syntax sufficient to avoid any errors that would invalidate the document.

Although the disadvantages of XML would appear fatal to its use, the flexibilities of the more sophisticated structure make it a superior format and most suitable for our purposes. This however does not mean that linguists and language workers should stop using backslash codes; indeed, using FOSF is relatively easy as compared with XML and ensures a level of machine-readable consistency from which the lexicographer will benefit further downstream. In any case, any machine-readable format has enormous computation benefits over raw, untagged text.

Computer presentation

While there are a number of programs that have been developed to interactively display dictionaries, Kirrkir (Manning 2003; Manning et al. 2001) utilises an intuitive

⁶ The FOSF codes given as examples are all MDF compliant.

and engaging user interface and is most suitable for children. Kirrkirr was originally developed as a means of electronically displaying the Warlpiri dictionary (Laughren et al., in preparation; Laughren & Nash 1983). It is open source, cross-platform and free. Kirrkirr allows the user to navigate content using a variety of methods and supports multimedia content such as images, audio files and video files, although the latter has yet to be explored as an option within the scope of this project. Users can search for words using the target or the source language, travel among words in a network by their links to one another, or move through a collection of semantic domains to find related words. Most importantly with respect to language revitalisation projects in Australia and elsewhere, Kirrkirr is designed specifically to be accessible to dictionary novices.

Using a program such as Kirrkirr takes full advantage of the hierarchical structure of XML. For instance elements (contiguous chunks of information) can be hidden, meaning that while they are still present in the master dictionary file they are not shown on the display. This key feature is very important for the ideology that informs much of the project described here. An important principle in creating digital versions of information is to preserve everything, lest that version be the last record in existence at some point in the future. So all information (internal comments, tape references of example sentences, scientific names of flora and fauna and so on) is retained in the underlying dictionary structure: the master dictionary file. Using XML stylesheets to render surface realisations enables the lexicographer to decide which elements are displayed. A potentially limitless number of stylesheets can be specified for different versions of the dictionary and users can easily switch among the various stylesheets within the Kirrkirr user interface.

The Kurna electronic dictionary

In 2008 a team at the University of Sydney was commissioned⁷ to create an electronic dictionary of Kurna based on two original documents from the 19th century (Teichelmann 1857; Teichelmann & Schürmann 1840) that had been typed into backslash-coded text. An important concern for this project was that the text from the original documents be displayed alongside any modern interpretation, both for the inquisitive user and for the digital preservation of the originals. In effect the electronic version was to serve as a digital archival copy of both Teichelmann (1857) and Teichelmann and Schürmann (1840).

We decided then that Kirrkirr would be a suitable application as it allows for the display of multiple versions at once and for the option to hide everything apart from the modern interpretation showing only the lexical information. Furthermore it allows for the insertion of sound files so that learner users can access information regarding pronunciation.

⁷ The Department of Communications, Information Technology and the Arts is owed a debt of gratitude for funding the Kurna electronic dictionary project.

During the Kurna electronic dictionary project McElvenny (2008) devised a method of displaying the content of the dictionary on a mobile phone, as it would be more accessible to the younger Kurna community members. After reducing the size of the sound files that had been recorded by Kurna learners for the electronic dictionary we were able to include them in the mobile phone version, thus enabling the learner immediate access to pronunciation. Exploring the possibilities of mobile phone dictionaries has since become an important aspect of our larger project.

Mobile phone presentation

While delivering dictionaries electronically using computers is more intuitive than traditional printed materials in the remote and Indigenous Australian context, using mobile phones as a method of presentation is the most appropriate. Computers are still rare in remote communities and only schools are adequately equipped with them.

Several recent studies have shown that mobile phones are very common among Indigenous people in various regions around Australia. In Cape York for instance, mobile phones are the dominant form of information and communications technology (Brady, Dyson & Asela 2008; Dyson & Brady 2009). The rate of mobile phone ownership among Indigenous people in Central Australia is around half, and is highest among younger people (Tangentyere Council & Central Land Council 2007). Mobile phone ownership is moderate even in communities that still lack coverage (Australian Communications and Media Authority 2008).

Furthermore an informal survey of researchers active in remote communities around Australia suggests that mobile phones are far more common than computers and that many people either own mobile phones or can access one without difficulty. Consequently, young adults are generally more phone-literate than they are computer-literate. With all this in mind, mobile phones should be carefully considered for the effective delivery of language learning materials such as dictionaries.⁸

Naturally there are a number of drawbacks to mobile phones as a means of dictionary delivery. Most obviously there are tight restrictions as to the amount of data they can contain, and any further information – which may include example sentences, grammar and usage information, comments and notes – is unfortunately jettisoned. However the purpose of mobile phone dictionaries as proposed here is not to compete with or usurp the status of computer-based electronic dictionaries, but instead to complement them; to provide continued access to users even when the computer with the full version of the dictionary is no longer available for use. Computer dictionaries and mobile phone dictionaries are intended to work together to reinforce language learners' efforts.

⁸ For a more full discussion of individual mobile phone dictionary projects, or for information about the software and how to produce mobile phone dictionaries, please see the website for the Project for Free Electronic Dictionaries pfed.info/ .

Given the observable trajectory with respect to technological development it is entirely plausible that mobile phones in several years will be closer to hand-held computers, with higher capacity for memory and the ability to run software designed for a computer, such as Kirrkirr or other dictionary visualisation programs. It will then be possible to create dictionaries for mobile phones that do not sacrifice any content. Until then it is more important to make use of the multitude of electronic dictionaries of Indigenous languages by delivering them in a form that people can utilise effectively.

The Wagiman dictionary

Kybrook Farm, about 90 kilometres north of Katherine, is the home of around 100 people, roughly half of whom are ethnically Wagiman (S. Wilson 1999a). The Wagiman language is now only natively spoken by less than five individuals all of whom are aged in their sixties. Without a concerted effort to revitalise it Wagiman is expected to disappear within ten years (A. Wilson 2006).

Kybrook Farm is a typical remote Aboriginal community in that computers are rare; while the community office has a small number of computers they are generally not available for community members to use freely. Furthermore individuals do not have their own computers. Mobile phones though, are ubiquitous; almost all community members have mobile phones and all are technologically proficient in using them. For these reasons Wagiman was one language chosen for a trial run of an early incarnation of a mobile phone dictionary.

While an electronic dictionary had been created for Wagiman by S. Wilson (1999b), a revision of the dictionary contents was necessary. This provided an opportunity to port the Wagiman dictionary into Kirrkirr and moreover, to produce a mobile phone version. A demonstration version using the content from the online dictionary was produced and shown to the Wagiman community during a fieldtrip in February 2009. The response to the dictionaries was very positive, both from the younger members of the community and the adults and Elders. The consensus was that the portability of the mobile phone meant that the children, and indeed the adults, could always keep the dictionary with them. After subsequent work to complete the dictionary, a first edition was released in September 2009.

Conclusion

Computers are still rare in areas that are enduring language endangerment although mobile phone ownership is relatively high; most people high school age and above either own, or are in close proximity to, a mobile phone at all times. Mobile phones, though continually evolving closer to miniaturised computers, are still unable to contain a large amount of information. As a result dictionaries developed for mobile phones must sacrifice a large amount of content that is usually critical for language learners' dictionaries. Resources based on computers on the other hand are not subject to the same space constraints as today's mobile phones; they are able to contain huge

amounts of data including images, sounds and movies. The disadvantage of computers is that they are not portable and their price still restricts their availability, meaning they are relatively rare in remote Australia and Aboriginal Australia. However these constraints on computers and mobiles phones may soon diminish, as recent history shows that computers are becoming smaller and less expensive, while mobile phones are becoming more powerful and, in fact, closer to computers in their capacity for multimedia content and functionality.

One potentially effective way to take advantage of the technological infrastructure of remote Australia is to create and disseminate both computer- and mobile phone-based language materials such as dictionaries. The computer-based resources would be of considerable use in classrooms – which are in fact well equipped with computers – and the mobile phone-based resources would be available to everyone at any time.

This is not to suggest that mobile phone- and computer-based dictionaries are in themselves sufficient to stave off language endangerment; they are merely tools and should be utilised in conjunction with other initiatives, such as bilingual education and Indigenous language education, in an attempt to strengthen Indigenous languages in Australia.

Acknowledgements

The project owes much to many people, most notably Jane Simpson for her unceasing mentorship, but also Peter Austin, Steven Bird, Sarah Cutfield, David Nash and others who attended the 2009 Australian Languages Workshop at Kioloa in March for their helpful comments, discussions and willingness to test and show off some of our dictionaries. We would also like to thank the linguists of Katherine: Lauren Campbell, Greg Dickson, Salome Harris and Colleen McQuay for their enthusiasm and support.

Most of this project would not be possible without the financial support of the Hoffman Foundation, whose donation has already enabled us to produce a number of dictionaries for minority languages.

References

- Australian Communications & Media Authority (2008). *Telecommunications in remote Indigenous communities*. Canberra: Australian Communications & Media Authority.
- Brady F, Dyson LE & Asela T (2008), Indigenous adoption of mobile phones and oral culture. In F Sudweeks, H Hrachovec & C Ess (Eds). *Proceedings: Cultural attitudes towards communication and technology 2008*, (pp. 384–98). Perth: Murdoch University.
- Corris M, Manning C, Poetsch S & Simpson J (2000). Bilingual dictionaries for Australian Aboriginal languages: user studies on the place of paper and electronic dictionaries. In U Heid, S Evert, E Lehmann & C Rohrer (Eds). *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000* (pp. 169–81). Stuttgart: Universität Stuttgart.

Corris M, Manning C, Poetsch S & Simpson J (2004). How useful and usable are dictionaries for speakers of Australian Indigenous languages? *International Journal of Lexicography*, 17(1): 33–68.

Coward DF & Grimes CE (1995). *Making dictionaries: a guide to lexicography and the Multi-Dictionary Formatter*. Waxhaw, NC: Summer Institute of Linguistics.

Dyson LE & Brady F (2009). Mobile phone adoption and use in Lockhart River Aboriginal community. In X Hu, E Scornavacca & Q Hu (Eds). *2009 International Conference on Mobile Business* (pp. 170–75). Dalian: Dalian University of Technology.

Laughren M, Hale K & Hoogenraad R (forthcoming). Warlpiri dictionary. Unpublished electronic datafiles. Brisbane: University of Queensland.

Laughren M & Nash D (1983). Warlpiri dictionary project: aims, method, organization and problems of definition. In P Austin (Ed). *Papers in Australian linguistics No. 15: Australian Aboriginal lexicography*. Series A–66 (pp. 109–33). Canberra: Pacific Linguistics.

Manning CD (2003). *Kirrkirr: software for the exploration of indigenous language dictionaries*. [Online]. Available: nlp.stanford.edu/kirrkirr/ [Accessed 26 March 2009].

Manning CD, Jansz K & Indurkha N (2001). Kirrkirr: software for browsing and visual exploration of a structured Warlpiri dictionary. *Literary and Linguistic Computing*, 16(2): 135–51.

McElvenny J (2008). *Mobile phone dictionaries* [Online]. Available: blogs.usyd.edu.au/elac/2008/07/mobile_phone_dictionaries.html [Accessed ?].

McElvenny J & Wilson A (2009). *Electronic dictionaries for language reclamation*. Paper presented at Supporting Small Languages Together: The First International Conference on Language Documentation and Conservation, University of Hawai'i at Manoa, 16–17 March 2009.

Nesi H (1999). A user's guide to electronic dictionaries for language learners. *International Journal of Lexicography*, 12(1): 55–66.

Tangentyere Council & Central Land Council (2007). *Ingerrekenhe antirrkweme: mobile phone use among low income Aboriginal people – a Central Australian snapshot*. Alice Springs: Tangentyere Council & Central Land Council.

Teichelmann CG (1857). Dictionary of the Adelaide dialect. ms. No. 59 Bleek's catalogue of Sir George Grey's library dealing with Australian languages, South African Public Library.

Teichelmann CG & Schürmann CW (1840/1962). *Outlines of a grammar, vocabulary, and phraseology, of the Aboriginal language of South Australia, spoken by the natives in and for some distance around Adelaide*. Adelaide: Published by the authors at the native location. Facsimile edition 1962 State Library of South Australia. Facsimile edition 1982, Adelaide: Tjintu Books.

Wilson A (2006). Negative evidence in linguistics: the case of Wagiman complex predicates. Unpublished honours thesis, Department of Linguistics, University of Sydney, Sydney, NSW.

Wilson S (1999a). *Coverbs and complex predicates in Wagiman*. Stanford, CA: Center for the Study of Language & Information.

Wilson S (1999b). *The Wagiman online dictionary* [Online]. Available: www.arts.usyd.edu.au/departs/linguistics/research/wagiman/ [Accessed 26 March 2009].

World Wide Web Consortium (2008). *Extensible Markup Language (XML) 1.0*. 5th Ed. [Online]. Available: www.w3.org/TR/2008/REC-xml-20081126/ [Accessed 26 March 2009].