As we may link:
Time -aligned concordances of field recordings, a working model.

Computing Arts 2001, Digital Resources for Research in the Humanities

Nick Thieberger
n.thieberger@linguistics.unimelb.edu.au
http://www.linguistics.unimelb.edu.au/people/postgrads/thieberger/ntcv.html

> One can now picture a future investigator in his laboratory. His hands are free, and he is not anchored. As he moves about and observes, he photographs and comments. Time is automatically recorded to tie the two records together. If he goes into the field, he may be connected by radio to his recorder. As he ponders over his notes in the evening, he again talks his comments into the record.
>
> Vannevar Bush (1945)

It has taken some time, but we are now able to create a system like the one envisaged by Vannevar Bush over 50 years ago. And despite the obvious leaps and bounds in technologies there are still areas in which much needs to be done. Linguists working on small languages (those typically spoken by indigenous people) with limited research grants typically patch together tools that will do what we want. Our research involves recording stories, sentences and so on, and then analysing that material to write a grammatical description. What we have done is record on to cassette, then transcribe the cassette and store it safely somewhere (like in our garage, or a cupboard). However a growing awareness that the products of our work need to be preserved in perpetuity means that we are also actively seeking principled approaches to language documentation.

We need tools that will allow us to do our recording and analysis, and in fact will enhance these tasks, but will also give us archivable products. In this paper I discuss a process I have undergone as part of my research for a grammar of a language of Vanuatu. I wanted a tool that met my current need to work interactively with digitised audio and its transcript (Desideratum 1). I had established links like this before by chunking audio files and naming them with a textual rendition of their contents (Thieberger 1994), or by building explicit links between a line of text and an audio file. But this was far too labour intensive an approach for the size of data I envisaged linking (say 30 fieldtapes).

In addition the tool has to be interactive as a transcript can always be improved as one's understanding of the language improves. So we need to be able to change the transcript.

In fact as the key desiderata were achieved they soon spawned other obvious functions:

Desiderata:

1. a textual index of digitised audio
2. instant access

3 unlimited amounts of unsegmented audio files
4 ability to change the transcript
5 conformant text mark-up
6 a concordance point of entry to the text
7 ability to create citations fromthe data for use as examples in my thesis
8 ability to create a playlist of selected sentences for use in presentations

At the beginning of the 21st century we can link digital data in exciting ways, both for representation of the material and for analysis.This seems to be a fairly simple task. We routinely use web browsers in which a link to an audio grab is accessed by a click. This works well for short audio files, but (in general) requires the audio file to be segmented to the required chunk size. In addition, the text is viewed through a browser which is not an editor, so the text is fixed. At the time that I was making decisions about how to deal with my transcripts I was unable to find a satisfactory way to transcribe, link the transcript and audio at the same time as I was transcribing, and then have access to all of my linked data.

Futureware

It has been clear for some time that we should use standard coding formats like SGML or XML partly because that was where corporate software development was headed. In the Humanities there has been a great deal of effort expended by the Text Encoding Initiative (TEI), a consortium of industry and academia, to establish standard forms for humanities computing. Their focus was on an international standard for text mark-up, known as SGML. This is a superset of a more simple but nevertheless powerful mark-up language called XML, and it is XML that looks like being the appropriate direction for the interchange of linguistic data.

Using this standard is known as being confomant. In the meantime the sort of software that was available for tagging audio (like SoundEdit) was quite non-conformant. It happily allows you to tag an audio file but does not make that explicit linkage available in a standard format. So a few years ago we had to have faith that being conformant would reap results for us in the future. Demonstrations of futureware entice us down this path. But futureware is like having a leaking roof and being told that the optimal solution ('best practice') is to use the state of the art sealant being developed overseas, currently available as a demo that is able to cover a maximum of 2 square centimetres, but with the promise that it will, VERY SOON, be able to cover the roof. The alternative is to use existing materials to stop the leak and get on with your life. An example of futureware is EUDICO, which has the following blurb on its website: " EUDICO is seen by the Max Planck Institute as the linguistic tool of the future. It is considered to be a "universal" work bench for linguists dealing with corpora as they are used at the MPI and elsewhere. EUDICO's main purpose is to offer a set of general tools for browsing, viewing, creating, editing, searching and analyzing collections of annotations on digitized video and audio recordings of linguistically interesting phenomena." (http://www.mpi.nl/world/tg/lapp/eudico/eudico.html)

And another is Lingualinks from SIL, which makes no mention of audio data:

"Lingualinks
Integrated tools for data management in your fieldwork.
Provides tools to:
analyze and document the relationships among semantic concepts in the language; collect, manage, analyze, and publish interlinearized texts; manage lexical data and publish a dictionary; perform a phonemic analysis on a corpus of phonetic data (Microsoft Windows only)" (http://www.ethnologue.com/LL_docs/ll_intro.asp)

Best practice? Or as good as it can be given other constaints?
At issue with 'best practice' is the time taken and effort involved. Some practitioners in a discipline may be intrigued by the arcane nooks and crannies of mark-up languages and the software tools required to run them, others may be completely disinterested and just want immediate results. Postgraduate linguistics students currently have no incentive to do any more than the minumum. If their thesis is a grammar of the language there are no extra points for having a richer language document than a grammar provides. If we want to encourage good or best practice then we should be able to build techniques for dealing with data that do not require huge additional effort, and preferably are part of our normal operating procedure.

When we record information in the field, be it stories, songs, or whatever, we immediately decontextualise the material and abstract it into an object of study. We further isolate parts of what we now call the data into example texts or sentences for our analysis. But ultimately we will need to recontextualise as much of this as we can for referencing and for long-term archiving. This could mean a fairly painful effort of locating where material came from, or alternatively it could mean we just don't bother and end up leaving the various bits in their dis-integrated parts for posterity. We need a methodology for dealing with our field recordings that allows us to maintain the links between contextual data and the examples we abstract. If we have a good way of doing this then it should prove no more onerous than the actual task in hand, with a great payback in terms of accessibility for ourselves and for future users of the data.

When looking at methods for digitising fieldtapes I was concerend that my work would have long-term benefits, beyond the thesis that is my immediate goal. However I also needed results now and in a form that I could operate. None of the current solutions allows you to simply amass your field tapes and produce a text-based interface to them so that you can click on a sentence anywhere in your transcript and hear it.

I have used an old and no longer supported tool, HyperCard, to instanitate links between objects, in this case digitised audio and its textual representation. HyperCard is orphaned software first produced for Apple computers in 1987 and best described as a mechano set for Macs. The latest version, 2.4.1 came out some 3 years ago and unfortunately Apple has not continued to support it. What I am presenting here is a kind of plasticene (or mechano) model of what someone with more skills may do in a more principled way Real Soon Now. As will be seen in this demonstration, each of the desiderata discussed above is met by this working model.

The links between utterances (whatever size one requires) and audio were created using SoundIndex (http://www.multimania.com/jacobson/SI/Index.htm.), from LACITO (http://195.83.92.32/presentation/index.html.en) in Paris. I chose SoundIndex because it was free, it was simple to learn and the authors answered my queries. I have looked at Transcriber (http://www.etca.fr/ CTA/gip/Projets/Transcriber/) but it involved installing Tcl/Tk, and Snack and there were issues of version incompatibility so I gave up.

Fieldtapes can be transcribed within SoundIndex, or existing transcripts can be imported as text and used as the basis for establishing links. The transcript remains in a text file, and an additional text file is created specifying the start / end times of the audio and the corresponding text start / end points.

The authors of SoundIndex kindly provided me with a Perl script for combining these text files into a single XML encoded file, in a process that illustrates the importance of working with standard formats. The tool built in HyperCard similarly allows the data to be re-exported as conformant XML.

Sample XML output combining index and transcript via SoundIndex:

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<?xml-stylesheet type="text/xsl" href="default.xsl"?>
<TEXT>
<HEADER>
<SOUNDFILE href="98003b.aif" />
</HEADER>
<BODY>
<S id="s1"><TRANSCR>Ipiatlak nmatu iskei,</TRANSCR><AUDIO start="0.0000"
end="2.3000"></AUDIO></S>
…….
<S id="sn"><TRANSCR>Go Ririal imer nrik Ririel kina, &amp;quot;Tik, ag =pa fag.&amp;quot;
</TRANSCR><AUDIO start="29.9799" end="33.2001"></AUDIO></S>
</BODY>
</TEXT>
```

Despite HyperCard being old, it nevertheless provides a wonderful means of presenting and relating data. In addition, there are existing HyperCard tools for one of the other desiderata in this context: concordancing, or providing an index of each item in the texts with an interactive link to its context. Mark Zimmerman's Free Text concordancer (http://www.his.com/~z/c/index.html) (provided under the GNU General Public License philosophy) makes it possible to simply index the transcripts and acccess them via a concordance, which then plays the audio of the selected text.

While there may be other ways of implementing these links, I need a tool I can use right now, and so have pursued a course with which I was already familiar.

As I have reiterated, this is not a product that I am promoting. Rather I am suggesting that the process of thinking about data as having a life beyond our immediate use of it, together with our need as practitioners to access audio and textual data should conspire to ensure that our work practices include issues such as explicit audio-text links. Finally I

make a plea for appropriate software tools for practitioners who just want to get their work done in a principled way and who don't have time to wait for futureware.

References

Bush, Vannevar 1945. As we may think. *Atlantic Monthly*, July 1945

Himmelmann, Nikolaus 1998. Documentary and descriptive linguistics. *Linguistics* (36). 161-195.

Thieberger, Nicholas 1994. Australia's Indigenous languages Information Stacks, Canberra:AIATSIS