

Language Documentation & Linguistic Theory 2

Towards a model of maximal accessibility in linguistic documentation work

Conor McDonough Quinn, *University of Nizwa*

Towards a Model of Maximal Accessibility in Linguistic Documentation Work

The goal of the present effort is to facilitate heritage-community access not just to the products of documentary linguistic work, but also to their means of production.

In this presentation we show how to share with a non-technical audience the means to easily link together a rich and complicated range of linguistic source material in a form that is both stably archivable and broadly accessible/disseminable. The approach suggested is a simple one: the skills and technology involved are minimal and readily learnable, even to the most computer-illiterate/computer-unconfident.

As an example case, we demonstrate how to handle a core problem in community language work, namely, how to effectively collect, organize, and distribute multiple versions of the same source material: e.g. the original text, plus a phonemicized and/or alternate spelling system version of that text, plus notes and comments, plus cross-references to dictionaries, plus links to related video and sound recordings, and so on.

To the non-specialist, this would seem to require either extensive technical computer knowhow, or at least an expensive and/or complex database tool. In this talk we showcase how one can, in a wholly non-technical, broadly accessible manner, share the basic notion(s) and application(s) of XML-based data management---with all the benefits of being free, open-source, platform-neutral, fundamentally human-legible, flexible, and simple to use and troubleshoot---in the form of a very brief but effective introduction to the small set of skills needed to create documents of this type.

This introduction shows nonspecialist language workers how to easily create basic annotated interlinear glossed texts, forming master documents which are not only searchable, but can also readily link in related sound/video files, or image files of original documents, as well as any number of alternate or earlier versions of the same core material, e.g. direct transcriptions, and phonemicized versions thereof, plus corrections. Particularly highlighted for the non-technical audience is this approach's flexibility: that alongside this capacity for extensive annotation and cross-linking of documentary materials comes a capacity for exceedingly fluid presentation, allowing users to create from one single master document

- interlinear glossed texts
- facing-page (synoptic) bilingual texts
- monolingual texts or translations
- selective presentations thereof (e.g. school-appropriate dictionary entries)

at essentially the press of a button. This even as the master document can be printed out as a transparently human-legible paper backup, just in case all digital data systems become defunct.

For concreteness, we draw the actual content from a collection of pre-modern Penobscot (E. Algonquian) texts, including the original text and phonemic retranscription of a pre-phonemic collection of traditional literature (Speck 1918), alongside a set of similarly redacted rare native-speaker writings from the 19th and 20th centuries. Having demonstrated this approach's use for individual documents, we then show how this and related Penobscot/E. Abenaki materials can be equally simply integrated into one massive digital document set linking all known documentation for these languages---a document set that nonetheless is not bound to a specific computer platform or indeed to computer technology at all.

In short, the technology here is not new. We show here that it can be detechnicalized and presented in a form that makes it usable to those with limited computing resources---as one step in the ongoing process to democratize access not just to the products of documentary linguistic work, but also to the tools of its creation.