

Toward the Interoperability of Language Resources

July 13-15, in conjunction with the 2007 LSA Summer Institute



Legacy Code, Web 2.0 & Preserving Indigenous Languages

Alexandre Sevigny
McMaster University

One of the greatest challenges that our group faces in designing our cultural and linguistic archive of Migmaq is not really technical, rather it concerns design and useability. In this mini-position paper, I'll raise a few questions about those challenges and how they relate to interoperability.

1. **How can an archive encourage literacy through the use of Web 2.0 technologies?** With the plethora of web services, most related to social networking, that have been associated with Web 2.0, how can a cultural archive remain relevant. What are effective ways to integrate our mostly in-house solutions with the various environment now offered by Google Video, Flickr, Facebook, etc. Many of the ingenious ideas concerning creating a language and culture-preserving “digital lifestyle” that have been developed artisanally in academic settings can now be achieved relatively cheaply and easily through the use of existing, more professionally done open source solutions. This is both a blessing and a challenge for applied computational linguistics.
2. **Text Processing for a Personal Docuverse?** The future of text processing probably lies in granting a wider audience the ability to manipulate, edit and “mashup” texts in multiple fashions. This is probably going to be a very large component of the effectiveness of linguistic and cultural archive as something more than just academic tools or museum-like repositories. How do we create tools that allow users, many of whom are not very tech-savvy to play with texts in their own languages? Are there emerging standards that will permit this sort of individualized text manipulating. Indeed there are: but which service will emerge as the “Flickr for texts” par excellence? How do academics, with limited resources, bet on which system to adopt?
3. **Geo-tagging: Google Earth or Open GIS?** Geotagging is an extraordinarily useful technology for creating a link between linguistic spread and place. We have been working on an Open GIS based interface for our Cultural and Linguistic Archive of Mi'kmaq. However, the current avail-ability of solutions such as Google Earth SDK have



made creating a quick and reliable geo-tagging system a lot easier. How to go about integrating these technologies into archives? What role should they play for end users who will be both academics and community members with varying computing comfort levels.

4. **Content Management Systems: Stay or Go?** As with most other early archive builders, our group has been using CMS that date from when our project first began to be implemented. However, recent advances in CMS, like Plone, are very tempting. They offer a great simplification in terms of back-end development. So how do academic researchers want to migrate to new, more streamlined CMS? How can it be done and what happens in terms of problems concerning legacy code (often not the neatest when it produced by RAs)? What do we do to pre-serve “neat tricks” that we may have developed in-house? Perhaps the real question here has to do with the format that archival material must take to remain easily portable to new systems and easily adaptable to new functionalities. This debate centers around XML vs other formats.

Most of the issues I raise in the preceding four provocations concern adapting our legacy and cherished in-house archives to more professional open-source standards that conform with the new world of Web 2.0. The possibilities are vast and what can now be achieved easily and cheaply means that academic text-processing and archiving can achieve a look and feel and offer services that give a “professional experience”. This can be achieved by arriving at interoperability with existing technologies. Thus the challenge becomes a design issue. It is the constellation of features that must be accounted for in the planning stages that will drive selecting what, from a menu of possibilities, the options we are going to offer users, and most importantly, the experience that will lead them to want to engage in learning their language and celebrating their culture on-line.

