

The field of computational linguistics is exciting inasmuch as it permits linguists of different stripes to model language behaviour and also to model connections between different sorts of data. Having worked in both cognitive science and Indigenous linguistics, I have had occasion to create tools that are useful in each area and also see how fruitful the results can be when we apply cognitive scientific ideas to indigenous dictionary construction, linguistic mapping and human interface design for learning portals. Here is a list of challenges I have encountered and could speak to at the Summit:

1. **Lexicon/Knowledge Base Design** – our attempts at creating an on-line Migmaq Living Archive brought to the fore the needs for a different ontological organization of categories to properly account for Indigenous world-view, inferences and pragmatic knowledge. Even if this is at a basic level, that is, the difference between the lexical semantics of animacy and presence/non-presence or the ontological implications of how the scale of salience of persons is organized (i.e. in the Migmaq world-view, the order of persons is 2-3-1) demand a rethink of how to organize a KB.
2. **Links to other technologies.** When taking on a project of both sociolinguistic and pragmatic-cultural nature, the tools of scientific visualization used to represent the interlinked properties of info in the KB are important. Linguistic tools must be designed to be compatible with Geographic Information Systems (i.e. ArcView 9.0), conceptual mapping software (i.e. Leximancer), site search engines, parsers.
3. **Presentation and Storage of Information.** Once information has been gathered and collated in the KB, it is necessary to find ways of making that information accessible and useful to the various Indigenous populations that will be using it. The linguist and her team have the ethical obligation to present information in as accessible and pedagogical a fashion possible. To satisfy the Canadian Tri-Council Ethics requirements for research partnerships with Indigenous Peoples, the data gathered must be used to further community interests first. As well, if one were to follow the tenets of critical language awareness, then the “Digital Archive” produced must be to preserve culture or encourage literacy. This necessitates certain human-computer interface design choices for using the tools contained in the database.
4. **Usefulness of KB to Scientists.** A final point is that the information must be organized in such a way as to serve two masters. First and foremost, as I have already mentioned, it must serve the cause of Indigenous People’s literacy. Secondly, it must be organized and encoded in such a fashion as to be of use to the diverse community of scientists and social scientists interested in Indigenous Peoples research. These include linguists, applied psychologists, anthropologists, cognitive scientists, education theorists, among many others. This means adhering to open-source themes, attempting to generate platform-independent code, using XML, etc.

To serve all of these different masters, the computational linguist must make do with a suite of tools available in the open-source world each of which are often cleverly conceived and very well-designed, but not designed for “inter-operability”. Neither are they designed to be easily adopted for use in an online portal environment.