A LINGUISTIC ATLAS FOR ENDANGERED LANGUAGES: WWW.ATLAS-LING.CA

Marie-Odile Junker[1], Terry Stewart [2]

[1]*Carleton University (CANADA)*
[2]*University of Waterloo (CANADA)*
mojunker@connect.carleton.ca, terry.stewart@gmail.com

Abstract

The Cree-Innu linguistic atlas is a collaborative participatory action research project with partners involved in the preservation and the documentation of native american languages of the Algonquian language family. It was started in 2005 with a number of Cree, Innu and Naskapi communities. It is now being expanded to other languages of the same family: Western Cree, Metis, and Ojibwe dialects.

The linguistic atlas (www.atlas-ling.ca) offers interactive maps, downloadable texts, activities, and sound files. It is based on a conversation CD and manual, originally developed for second language acquisition of East Cree. The atlas currently contains 21 topics of conversation in over 15 languages and dialects of Cree-Innu.

Pedagogical applications include language acquisition via downloadable language lessons, and linguistic training via the discovery of the characteristics of this language family as well as the variation between languages and dialects at the phonological, morphological and lexicological levels. Our goal is to make some linguistic principles accessible to students in a playful way, facilitating the understanding of both the diversity and the deep unity of a family of native american languages.

In addition to documenting dialectal variation, the project also offers training and support for native speakers in language preservation and documentation, using information technology (IT). The central database is web-based, allowing collaboration at a distance. The atlas uses a Google map populated by an on-line relational database.

In this multimedia presentation, we show and discuss the tools, technology, and method used to implement this wide-reaching documentation effort, as well as its pedagogical applications.

Keywords: native american languages, Algonquian, Cree, Innu, linguistic atlas, endangered languages, google map, relational database

1   INTRODUCTION

The Cree-Innu linguistic atlas (www.atlas-ling.ca) is a collaborative participatory action research project with partners involved in the preservation and the documentation of native american languages of the Algonquian language family. It was started in 2005 with nine Innu dialects, seven Cree dialects, and the Naskapi dialects. It is now being expanded to other languages of the same family: Western Cree, Metis, and Ojibway dialects. The atlas offers interactive maps, downloadable texts, activities, and sound files, all based on conversation topics.

In this paper we discuss the tools, technology, and method used to implement this wide-reaching documentation effort, as well as its pedagogical applications whose goal is to facilitate the understanding of both the diversity and the deep unity of a family of native american languages.

2 WHAT IS AN INTERACTIVE LINGUISTIC ATLAS AND WHAT DOES IT OFFER?

2.1 An atlas from a language activist's perspective

While linguistic atlases are usually designed by linguists for linguists, our atlas started from a language activist's perspective. It is based on a conversation CD [1] and manual [2], originally developed for second language acquisition of East Cree. We started working in 2003 with a number of Cree, Innu and Naskapi communities who were adapting the East Cree Conversation CD and Manual to their own dialect. In 2004, a prototype was designed, using one category 'the family' (Fig. 1). The goal was to show aboriginal speakers and language specialists how their languages were related, and to generate an interest in sharing linguistic resources for language preservation and education. Because the education system in Canada is provincially administered, related linguistic communities do not fall under the same jurisdiction nor funding system, which creates a barrier to sharing resources and
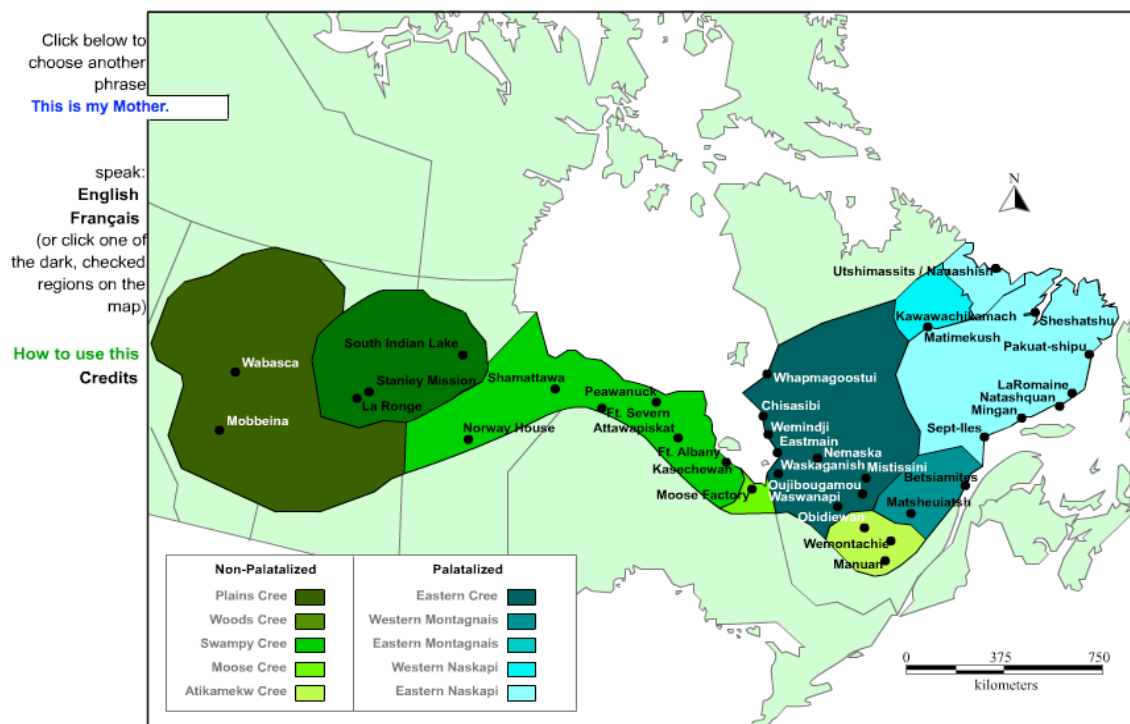
curriculum material.

Figure 1: interactive linguistic atlas prototype (www.eastcree.org) [3]

The framework used is a participatory research action process (henceforth PAR). PAR operates using the following parameters: the research process is more important than the research goal; The process aims at empowering participants: strengthen knowledge and self-esteem. The success is evaluated according to a positive impact on speakers and communities.

Following the format of the CD, the atlas today contains the following 21 topics of conversation, reflecting conversation topics of everyday life interactions in a Northern aboriginal community (Fig.2). This represents about 400 phrases or sentences. We are currently also collecting additional material, requested by our users and some linguists, to develop a better tool for dialectology studies.

1. Greetings	12. Location & Travel
2. Family	13. At Work
3. Numbers	14. At the Store
4. Days of the Week	15. At the School
5. Weather	16. Clothing
6. Seasons	17. Money
7. Physical Characteristics	18. Questions
8. Comforts, Discomforts, Needs	19. Orders & Requests
9. Feelings	20. Hunting, Trapping & Fishing
10. Social Events	21. Expressions of Time
11. Comings & Goings	NEW: Additional Grammatical features, a short song, a sample story

Figure 2: the topics of the linguistic atlas

2.2   Languages and dialects: Cree-Innu, Ojibwe

The atlas documents languages that belong to the Central Algonquian language family. The Cree-Innu languages: Plains Cree, Woodland Cree, Swampy Cree, Moose Cree, Naskapi, Atikamekw, and Innu (formerly called Montagnais) make up a linguistic continuum stretching from the Rocky Mountains to the Atlantic Coast (MacKenzie, 1980).[4]. Metis Cree and Michif are a form of Cree influenced by or mixed with French. The Ojibwe language family is a sister of Cree-Innu, belonging to the same larger Central Algonquian language family (Valentine, 1994) [5]. Dialects can be mutually intelligible if one understands the linguistic key to variation.

At the time of writing 17 different languages and dialects of the Cree-Innu subfamily have been documented and published on the web (Fig.3).
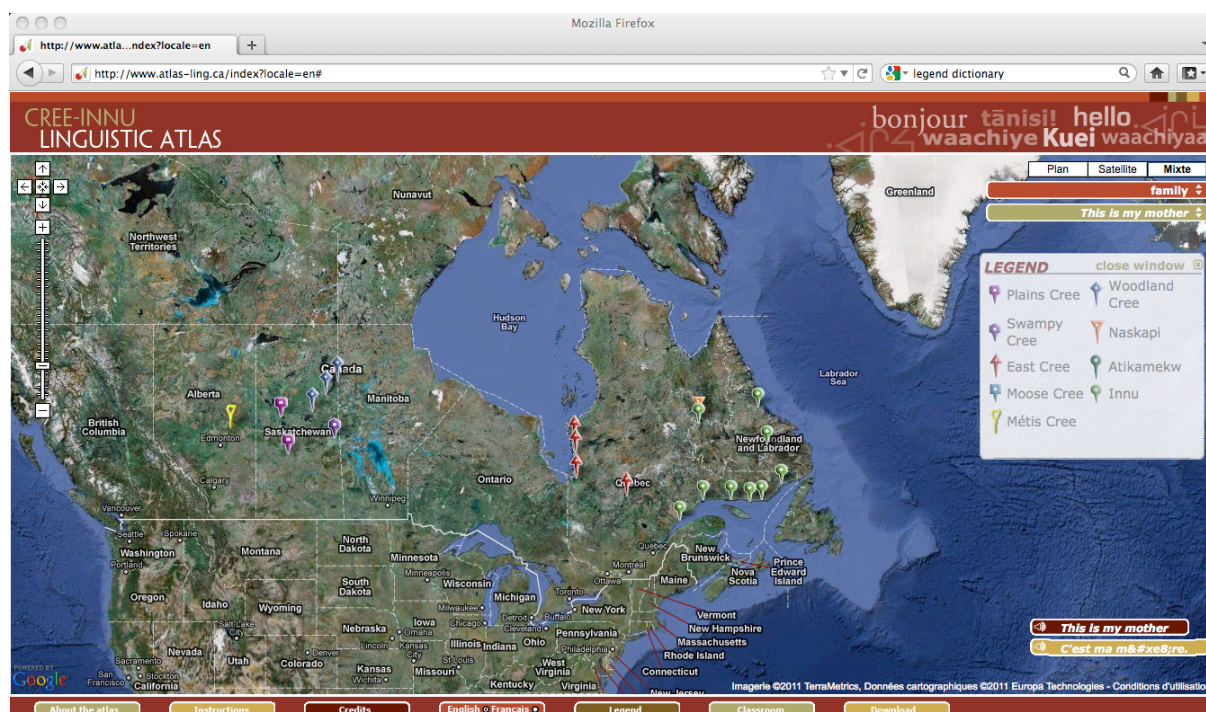

Figure 3: The Cree-Innu linguistc atlas (www.atlas-ling.ca)

For some dialects, there are several speakers, bringing up the total number of speakers to 25, and the total number of sound files to close to 10 000.

With interest growing amongst speakers, new partnerships being established, and funding made available last year, the project is now in its second phase and is being expanded to include two other

dialects of East Cree, 8 dialects of Ojibwe, 6 Swampy Cree dialects, 4 Plains Cree dialects, 3 dialects of Michif and 2 of Cree-Metis, for a total of 28 other languages or dialects.

2.3   Interactivity for the users

2.3.1   The interactive map

The bilingual (English-French) user interface allows the user to select and display one phrase or sentence at a time for all dialects by first selecting a topic, then one of its phrases or sentences. Clicking on a marker on the map loads up and plays the sound file and brings up a smaller window containing the associated text in standard orthography, as well as information about the dialect, the speaker and its location, and relevant comments. If the language uses a syllabic writing system, the equivalent conversion in roman orthography is provided above (Fig.4). If there are several speakers for that particular location and dialect, they appear in numbered tabs in the small window. The sound can be replayed as many times as necessary by clicking on the sound icon in the small window.
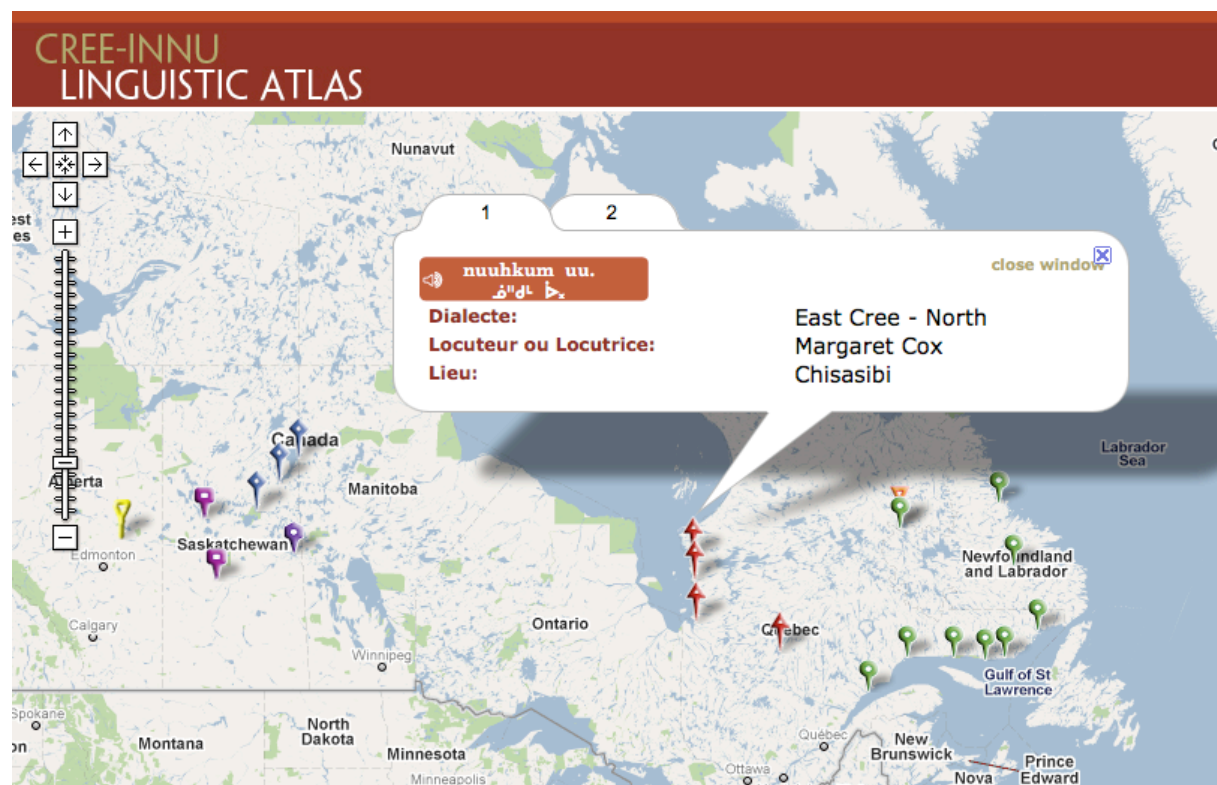

Figure 4: main user interface

2.3.2   Downloading Sound Files

The sound files can also be downloaded in mp3 format by selecting categories/topics and speakers/dialects (Fig. 5). The user decides which category, language or dialect to download as well as their download order. The system generates a custom mp3 file with all of the chosen sounds with pauses between phrases to allow repetition for oral practice. For people using this to learn an aboriginal language as a second language, we recommend following our topic order, and downloading the English or French before the Aboriginal language or dialect of their choice. For example, for a francophone moving to Ekuanitshit (Mingan) who wants to learn how to greet people in Innu, s/he could select the *Greetings* and *Family* categories, and the *French* and *Innu*: *Ekuanitshit* languages. The downloaded sound file will consist of each phrase in the chosen categories repeated in each of the chosen languages. For people who already speak an aboriginal language from the Cree-Innu family, but who want to learn a different dialect, we recommend that they place their dialect first and then the new one they are trying to learn. If people are using this to compare dialects, they can download as many dialects as they want, in the order they want. The download can be saved onto the user's computer, and played from there, a CD can be burned, or it can be imported into any mobile device such as the IPod.

Figure 5. Downloading sound files: selecting the categories

2.3.3 Downloading Text
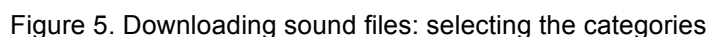
Text can also be downloaded. We currently provide the user with texts in the form of edited conversation manuals, in pdf, written to accompany CDs containing the entire 21 topics of conversation for a selection of languages including English and French. This is mainly designed for second language learners of those languages.

Figure 6: Downloaded text from a conversation manual (East Cree)

We also have an ability to download the text entered in the relational database, but as of now, this is reserved to the editors (see maintainers' interface). Other downloads include classroom tools (see section 4).

2.4 Training and support

The project also offers training and support for native speakers in language preservation and documentation, using information technology (IT) and communication technology (CT). We offer training sessions for sound recording and editing, and support the dialog between youth and elders for language documentation and preservation. Training sessions on the atlas tools, including the maintainers' interfaces are offered at Carleton University, on site or through on-line video-conferencing. Training of aboriginal speakers as graduate students is also made possible through the atlas project. For example, one aboriginal MA graduate who worked on the first phase of the atlas and subsequently got an instructor position at a Northern aboriginal college is now back with us as a research collaborator, training her own students in the second phase of the project.

3   TECHNICAL CONSIDERATIONS

3.1   Choosing Open source

The Linguistic Atlas is based on Open Source non-proprietary software. One major reason for this choice was to save on development costs, as all components are freely available to both users and developers. Furthermore, since we chose standard widely-used tools, it was simpler to hire developers familiar with these tools and to make use of highly knowledgeable online communities to help with debugging and development problems. Most importantly, this reduces the barrier to entry for other linguistic research groups who may wish to make use of our software, since there is no need to purchase expensive proprietary tools.

3.2   Google maps

To present the map at the core of the Linguistic Atlas, we use the Google Maps API (Application Programmer Interface). While it is not completely open-source, it is free for non-commercial use, and Google has publicly committed to keeping it available long-term. It is currently used by over 350,000 web sites.

The Google Maps API is responsible for drawing the map images, meaning that we do not have to invest effort in maintaining and updating the aerial and satellite photography used by the atlas. Google also ensures that the map is supported by all major web browsers (Firefox, Safari, Internet Explorer, Opera), and provides JavaScript functionality for drawing lines and showing markers on the map.

3.3   Required Software

To run the Linguistic Atlas, we require a single standard Internet server. To store the linguistic information, we use the open-source relational database system MySQL <http://mysql.com>. The actual web pages, including the administration interface for adding and modifying the data, are served using a program written in the Python programming language <http://python.org> using the free TurboGears library <http://turbogears.org>. This is a system designed for rapidly creating interactive web sites. We had previously used these tools to develop a search engine for East Cree [6] and to develop a set of interactive language lessons for East Cree <http://eastcree.org/lessons/>.

The overall system is organized using a Model-View-Controller paradigm. The raw information as to the phrases, their associated glosses, sound files, and so on (including identifying the speaker and their location) are stored in the database. This is known as a model. The view is how users see this information, and is defined by a set of web pages written with Kid templates <http://kid-templating.org>. Every web page has its own Kid template specifying the look of the page (using standard XHTML tags) and placeholders to fill in the specific information from the database. This allows the web site design to be easily changed without modifying the underlying information stored in the database.

Finally, the controller consists of custom Python code that connects the model to the view. That is, it sifts through all the information in the database, extracting out whatever data is relevant for the current web page. For example, if a web page needs to show all the sounds for a particular speaker that are in the category "family", then the controller is responsible for pulling those items out of the database, while the view determines exactly how to present that information to the user. By separating tasks in this manner, we enable a rapid and robust development environment, and by using these standardized tools we can make use of the extensive documentation and large community of other developers following this approach.

4   DESIGN CONSIDERATIONS

4.1   Database Organization

The most fundamental decision in the design of the Linguistic Atlas is how to organize the information used. This database organization forms constraints on the type of information that is most easily represented in our system. Our overall relational structure is shown in figure 7. Our chosen emphasis is on conversational meanings. Every sound entered into the database is associated with both a

speaker and a particular meaning. Every gloss (the written form of the phrase) is associated with a
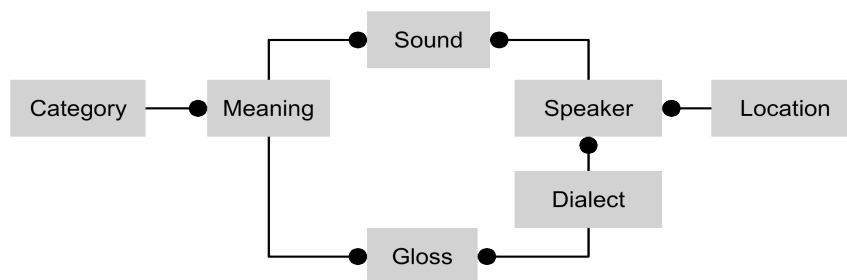
dialect and a meaning.

Figure 7: Database structure. Circles indicate a one-to-many relationship. (For example, a single Speaker is associated with multiple Sounds, but every Sound is associated with a single Speaker.)

Choosing this conversational meaning approach leads to some interesting and non-intuitive features of our database. For example, *akup* means 'coat' in Innu, and *akuhp* means 'coat' in East Cree. Most dialects follow this pattern, but in Plains Cree, *akohp* means 'blanket' while the word for 'coat' is *miskotākay*. Since 'blanket' is not one of the meanings in our database, we do not currently include the Plains Cree *akohp*, even though there is a clear etymological relationship between it and *akup* and *akuhp*, which we do include. For a historical linguist, these words would all fall under the same etymology, from the proto-Algonquian *akohm*. Comparing variants of the original form is what an historical linguist, phonologist, dialectologist, or semanticist might be interested in, as coats and blankets both keep you warm, and therefore in some dialects 'blanket' came to mean 'coat'. However, this is not the intent of this Linguistic Atlas. We do not establish relationships based on comparable forms, nor on etymology, but rather on equivalent meanings as used in conversation.

Sometimes the equivalent meaning is not available. Innu does not have words for the four cardinal directions (North, South, East, West) but uses a system made of words describing the directions in which the main winds blow and the rivers flow ('upstream', 'downstream'). Furthermore, depending on their environment, people do not engage in the same activities. While Northern people living by the coast go seal hunting, people living in the Plains do not, but may go buffalo hunting instead. This implies that meanings will not necessarily exist across dialects. The same problem arises with our location and travels section, where people have distinct words and expressions to talk about their common destinations around their communities. To deal with this, we add community-specific meanings to the database that only have data for the particular relevant dialects. We include this data in the conversation CDs and manuals, but exclude them from the Atlas since there are no relevant comparisons.

4.2   Maintainer's Interface

To modify and maintain the linguistic atlas, we include a web-based interface as part of the system. This allows work at a distance and collaborative interventions. Each screen organizes the information in a different way, allowing us to view, enter, and modify the data by phrase, dialect variant, speaker, and location. (Fig. 6)

Phrase	Dialect Variant	Speaker	Location	Category
There's a dance.	Please Select An Option	Please Select An Option	Please Select An Option	Other

Categories: social events   **Exclude from Map**   **Exclude from Download**

Speaker	Sound file		Location	Gloss
Marguerite MacKenzie (no location)	en/social events/There is a dance-en.mp3	note	English	There's a dance.
Marie-Odile Junker (no location)	fr/social events/There is a dance-fr.mp3	note	French	Il y a une danse.
Leda Corrigal (Beauval)	Beauval-LCorrigal/social events/There is a dance-Beauval-LCorrigal.mp3	note	Plains Cree	nimihitonāniwin.
Martha Michell (Kinoosao)	Kinoosao-MMichell/social events/there is a dance-Kinoosao-MMichell.mp3	note	Woodland Cree - North	nimihitonāniwin.
Minnie McKenzie (Stanley Mission)	StanleyM-MMcKenzie/social events/There's a dance-StanleyM-MMcKenzie.mp3	note	Woodland Cree - South	nimihitonāniwin.
			Swampy Cree	
Marion Cox (Chisasibi)	Chisasibi-MarionCox/social events/There is a dance-Chisasibi-MarionCox.mp3	note		
Margaret Cox (Chisasibi)	Chisasibi-MargaretCox/social events/There is a dance-Chisasibi-MargaretCox.mp3	note	East Cree - North	niiminaanuu. / ᓃᒥᓈᓅ
Frances Visitor (Wemindji)	Wemindji-FVisitor/social events/There's a dance-Wemindji-FVisitor.mp3	note		
Louise Blacksmith (Mistissini)	Mistissini-LBlacksmith/social events/There is a dance-Mistissini-LBlacksmith.mp3	note	East Cree - South - Inland	niiminaanuu. / ᓃᒥᓈᓅ

Figure 6: the maintainers' interface (text and sound, with selection by Phrase: *There is a dance*)

The text elements can be viewed by category and exported as text file as needed, for inclusion in another database or to create manuals (Fig. 7). Integrated conversion tools convert roman orthography into syllabics (for East Cree, Naskapi and other languages which use a syllabic writing system). We also found that particular tools were useful for keeping the database as a whole organized. When new sounds are added to the system, we have a page that will guess what meaning they should be associated with based on the file name of the sounds. This guess is presented to the user for confirmation, or they may select another existing meaning from the database. We also found it useful to be able to browse through the existing sound files with an indication of which speaker they have been associated with, and to be able to find any sounds or glosses which have not yet been associated with a phrase.

Phrase	Dialect Variant	Speaker	Location	Category
Please Select An Option	Please Select An Option	Please Select An Option	Please Select An Option	Other

Category: family

Export to file

English	French	East Cree - North	East Cree - South - Coastal	East Cree - South - Inland	Innu - Ekuan
Family	La famille				
This is my mother.	C'est ma mère.	nikaawii uu / ᓂᑳᐅᐄ ᐤ	nikaawii uu / ᓂᑳᐅᐄ ᐤ	nikaawii uu / ᓂᑳᐅᐄ ᐤ	Nikaui ue. / Nikaui an.
This is my father.	C'est mon père.	nuuhtaawii uu / ᓅᐦᑖᐅᐄ ᐤ	nuuhtaawii uu / ᓅᐦᑖᐅᐄ ᐤ	nuuhtaawii uu / ᓅᐦᑖᐅᐄ ᐤ	Nutaui ue. / Nutaui an.
This is my younger brother or sister.	C'est mon frère ou ma soeur.	nishiim uu. / ᓂᔒᒻ ᐤ	nishiim uu. / ᓂᔒᒻ ᐤ	nishiim uu / ᓂᔒᒻ ᐤ	Nishim ue. / Nishim an.
This is my older sister.	C'est ma grande soeur.	nimis uu. / ᓂᒥᔅ ᐤ	nimis uu. / ᓂᒥᔅ ᐤ	nimis uu. / ᓂᒥᔅ ᐤ	Nimish an. / Nimish ue.
This is my older brother.	C'est mon grand-frère.	nistaas uu. / ᓂᔅᑖᔅ ᐤ	nistes uu. / ᓂᔅᑌᔅ ᐤ	nistes uu. / ᓂᔅᑌᔅ ᐤ	Nishtesh an. / Nishtesh ue
This is my child.	C'est mon enfant.	nitiwaashishiim uu. / ᓂᑎᐙᔑᔒᒻ ᐤ	nitawaashishiim uu. / ᓂᑕᐙᔑᔒᒻ ᐤ	nitawaashishiim uu. / ᓂᑕᐙᔑᔒᒻ ᐤ	Nitauassim an. / Nitauass

Figure 7: the maintainer's interface (text by Category, with *Export to file* button)

Finally, we added a method for merging together two different meanings. This was necessary since we found that it was sometimes unclear whether we should split a meaning into two separate meanings due to regional variation in activities, as discussed in section 4.1. Being able to easily merge these back together (rather than manually changing the meaning for each sound and gloss separately) allowed us to easily make changes to the organization of the database.

5   PEDAGOGICAL APPLICATIONS

5.1   Goals

Pedagogical applications include language acquisition via downloadable language lessons (sound and text), and linguistic training via the discovery of the characteristics of this language family as well as the variation between languages and dialects at the phonological, morphological and lexicological levels. Our goal is to make some linguistic principles accessible to students in a playful way, facilitating the understanding of both the diversity and the deep unity of a family of native american languages.

5.2   Language learning

For language learning people usually download the sound files and manuals and use the material off site. (see downloading sound files and text in section 3). For the most frequently used learning combinations we provide high quality audio CDs, to accompany the edited manuals, that can be purchased from our aboriginal partners. One special feature of our CDs is that the colonial languages (English and French) can be turned off, to allow only playing the target aboriginal language. There is a long enough pause between phrases to allow repetition. A special headphone enhancement technique has been used to improve sound quality when listening with headphones. Profits from CD sales go toward language preservation and linguistic documentation training for Cree youth.

5.3   Dialectal variation, literacy, standardization of orthographies

Teachers who have an internet connection in the classroom have reported that they use the interactive atlas for teaching about dialectal variation. Teacher training programs also make use of the atlas to bring awareness of both the diversity and the unity of the language families. Teacher training often takes place bringing together speakers from many dialects, and the atlas is used to break the ice.

Due to the fragile situation of these languages, teaching of literacy to speakers and future elementary teachers is often a priority. For that, the sound-text interface of the atlas is very much in use. The question of standardization of orthography is still much being discussed. For example, the Innu language, spoken both in the Canadian province of Quebec and Labrador has many different dialects but has decided on one standard orthography. It is common for speakers who are still learning to read and write in their language to find it difficult to understand the concept of a standard orthography. The atlas is being used to illustrate how one spelling can cover many pronunciations. For dialects still developing their standard orthographies, the atlas gives sound/text examples of what other related languages have done.

5.4   Classroom activities

With the same goal of furthering an understanding of both linguistic unity and diversity, a series of guidelines for discovery activities are provided, as downloadable pdf sheets.(Fig. 8)

Figure 8: downloadable classroom activities web page

The activities cover phonological, morphological, and lexicological differences between languages and dialects. They are designed for a mixed and multi-level audience of Linguistics students, Aboriginal language teachers, students, as well as First or Second language learners. In preparing these activities [7], the challenges were numerous: to reduce linguistic and grammatical terminology to a minimum without sacrificing precision, to harmonize writing systems across dialects (some use syllabics, but different ones; others use roman orthographies, but with different symbols to indicate vowel length: macrons or hats on vowels, or double vowels), to use or not to use the International phonetic alphabet (IPA) notation when describing phonetic characteristics, as well as to take into

account the level of technological literacy of our learners when directing them to on-line activities.

Figure 9 : a downloadable activity sheet to understand dialectal variation (also available in French)

6 CONCLUSION: NEXT STEPS

Since this is work in progress, let us mention our next steps. In addition to documenting more languages and extending our community of users, we are currently preparing new interfaces for dialectological studies. A complete inventory of existing linguistic features as well as supplementary data will augment our database in order to generate maps with isoglosses (the geographical boundary of a certain linguistic feature, such as the pronunciation of a vowel, the meaning of a word, or use of some syntactic feature). Due to participatory dimension of our project, we expect new developments driven by the research process, especially in the area of communication technologies or web 2 (Facebook pages of our projects are quite popular, and useful to engage a larger community of users). The planned addition of songs and stories has already led to the development of new oral stories databases, based on the eastcree.org model [8], for Swampy, Moose and Innu-Cree languages [9].

REFERENCES

[1] Junker, M-O. ed. (2002) Cree conversation crie - iinuu/iyiyuu ayimitaau. CD (Audio). 70mn. ISBN 894843-24-X. Download order form and description from: http://www.eastcree.org/cree/en /resources/cds

[2] Blacksmith, L., Junker, M-O, MacKenzie, M., Bobbish-Salt, L., Whiskeychan, A. (2002). Cree conversation Manual/ Manuel de conversation crie. Edited by M.-O. Junker. 43 pages. ISBN 1-894843-25-8. Download from http://www.eastcree.org/cree/en/resources/cds

[3] The Cree-Innu linguistic Atlas- L'atlas linguistique cri-innu:

    Prototype (2004): www.eastcree.org/cree/en/grammar/#

    Current development (2005-2011): www.atlas-ling.ca/

[4] MacKenzie, M. (1980) *Towards a Dialectology of Cree-Montagnais-Naskapi*. Ph.d thesis, University of Toronto.

[5] Valentine, J. R. (1994). *Ojibwe dialect relationships.* Ph.d thesis, University of Texas, Austin.

[6] Junker, M.-O., Stewart, T.C. (2008) Building Search Engines for Algonquian languages. In Hele, K. & Darnell, R. (eds), *Papers of the 39th Algonquian Conference*. London: University of Western Ontario, 378-411.

[7] Junker, M.-O., Logan, Kelly & MacKenzie, M. (2009) Discovering the Cree-Innu languages: Pedagogial applications for the Linguistic Atlas. Presented at the *41e Congrès des Algonquinistes*. Concordia University, Montreal, Canada- October 2009

[8] Junker, M.-O. & Luchian, R. (2007) Developing web databases for Aboriginal language preservation. *Literary and Linguistic Computing* 22(2),187- 206.

[9] Ellis, C.D. & Junker, M.-O (ed.) *Cree Legends and Narratives.* On-line audio collection. www.spokencree.org.

[10] *The gift of language and culture*: www.giftoflanguageandculture.ca