

PREPUBLICATION DRAFT (revised March 2011). To appear as:
Jancewicz, Bill, & Junker, Marie-Odile. Cree syllabic fonts: development, compatibility and usage in the digital world. In Valentine, Randolph (ed.) *Papers of the 40th Algonquian Conference*.

Cree syllabic fonts: development, compatibility and usage in the digital world*

**Bill Jancewicz and Marie-Odile Junker
SIL International - Carleton University**

Abstract

The goal of this paper is to discuss and provide an overview of our experience in helping develop resources for typing in Cree syllabics in the area of information technology. Changes in the digital world have allowed the generalization of Unicode, but the question of what are the best tools—and whether they can be understood and accessed by the people who need them—remains. Aboriginal languages, being minority languages, remain vulnerable to the dominance of English in the domain of software development, hardware preferences for certain platforms, and ever-changing technologies.

We report on successful and less successful initiatives undertaken to serve the needs of the nine East Cree communities in Quebec: development and updates of typing tool packages, open source vs. proprietary-commercial software, distribution and accessibility of those tools, problems with obsolete practices, and training. We report on some resources made available via the eastcree.org website, including training videos in Cree, English and French. Like standardisation of orthography, standardisation of technology is essential. We demonstrate and emphasise the benefits stemming from collaboration and sharing between all people working in any capacity with a minority language.

* We wish to thank all the Cree writers, speakers and linguists who have participated in our dialogue about Cree fonts over the year, as well as the audience at the *40th Algonquian Conference*, Minneapolis, MN October, 2008. Special thanks to Timothy di Leo Browne for editorial comments, and to Delasie Torkornoo, for technical support. Research for this paper was partially funded by a SSHRC grant (# 856-2004-1028).

Introduction

Like other minority languages, but maybe even more so, Aboriginal languages are facing challenges in encountering information technology (henceforth IT). Our experience in helping develop resources for typing in Cree syllabics in the IT area (see also Jancewicz and Junker, 2002) has led us to explore the following questions:

- How do changes in IT affect languages like Cree?
- What are the best tools?
- Can these tools be understood and accessed by the people who need them?
- To what extent are minority languages vulnerable in IT?

We will start by discussing the recent history of character encodings and the effect it has had on Cree users. We then examine current issues pertaining to display of fonts, keyboarding, conversions, and distribution. Finally, we discuss some current collaborative IT applications and practices involving the Cree language within the East Cree community of speakers.

From Legacy (8-bit) Encodings to Unicode

The history of the development of computer technology in the 1980s and 1990s provides the reasons for some of the hurdles that needed to be overcome in order to provide efficient usability of Cree syllabics on computers. When access to personal computers was first within reach of Cree speakers, the physical limitations of the computer's machine language forced font designers to independently develop their own encoding systems. Within the 8-bit/1-byte framework there is a theoretical limit of 256 characters. But because certain characters were reserved by the operating systems and the application programs, users were left with a practical limitation of fewer than 200 characters.

The standard "English" (so-called "Latin") character set, with upper- and lower-case characters, numerals, and a few symbols used up most of the character codes available in the first 128 characters that a standard keyboard could type. An additional set of "extended" (ANSI) characters was made available when Microsoft Windows and the Macintosh Classic operating systems were introduced, but unfortunately Microsoft and Apple did not agree upon which

characters would be represented in their “extended” character sets. To make matters worse, Microsoft Windows and Macintosh also developed different approaches to the “code-page” (so-called “standard” re-coding of the extended character set to enable the use of some majority languages).

By the mid-1990's the only workable solution for Cree language communities was to restrict the choice of operating systems used by the community, that is, to require everyone to use either Windows or Macintosh, but not both. Moreover, a computer technician or font designer had to be retained to develop a "hacked" font that contained the characters that the language community needed. Having made the operating system choice, all the users then had to be trained in the use of this special local font. A computer technician or programmer was also tasked with converting previously keyboarded documents into the accepted local standard. Sharing documented across operating systems, or even across application programs, meant that elaborate conversion procedures had to be followed in order to maintain the integrity of even the simplest documents

In addition, many of the legacy (8-bit) fonts experienced interference problems with AutoCorrect features of word processors. For example, features that automatically capitalized letters after periods, corrected “two initial capitals”, or replaced straight quotes with “smart” quotes often caused the wrong syllabic character to be placed in context. Users routinely reported these problems without being able to identify their source. We have heard of people who were using the strategy of leaving blank spaces for the misbehaving character in their electronic document and filling it later by hand after printing. If such texts are retrieved from electronic archives in the future, they will have missing characters for this odd reason. For legacy fonts, all AutoCorrect features have to be disabled. These difficulties were overcome by using Unicode.

The physical limitation of 256 characters for computers was overcome by the development of a standard referred to as Unicode. By assigning each character to two bytes, the theoretical limit was extended to 65,536 characters (256x256), and with the implementation of other strategies, the Unicode system is capable of handling millions of different characters on computer

systems. This is by far enough to accommodate every character of every known human language, modern and ancient.

The Unicode Consortium is an encoding committee that establishes agreed-upon character-to-code equivalencies. The Consortium publishes an exhaustive listing of all the characters used by all scripts, and the computer code (that is, the unique number) used to represent a given character.

For Canadian Syllabic script, a subset of Unicode characters referred to as Unified Canadian Aboriginal Syllabics (UCAS) was compiled with input from native speaker and community-based users to determine the characters required. All fonts developed that adhere to the Unicode encoding standard are interchangeable over different operating systems and application programs. Having access to multiple fonts using the same Unicode encoding enables users to select different typefaces according to utility or aesthetics.

BJCree UNI is one particular Unicode-based typeface for Cree that was developed in coordination and compliance with speakers and local communities. This means that the development of the glyphs (typeface shapes) and metrics (character spacing, leading, and kerning) was designed and optimized with the collaboration and approval of Cree-speaking readers.¹

While the introduction of Unicode has solved most of the previous compatibility problems, some Cree writers still have old texts in several obsolete standards, or sitting on old computers, which are not easily convertible. To remedy this situation we have endeavoured to maintain an archive of old fonts and conversion tools.

Unicode in Operating Systems and Applications

The first operating systems that were compliant with the technical constraints of Unicode were Windows 2000, followed by Windows XP and later versions of Windows (Vista, Windows 7, etc.) and versions of Macintosh OS X beginning with version 10.2.

¹ Some of the successful collaborative developments concerning fonts would not have been possible without the financial support of the Cree Regional Authority, the SSHRC project of Dr. Julie Brittain at Memorial University, as well as the partners of eastcree.org: SSHRC, Carleton University and Cree Programs (Cree School Board).

Application programs to run on these platforms that were Unicode-compliant came later. Currently, Microsoft Word (starting with Word 2000), OpenOffice 2.0, Nisus (for Mac), InDesign, (replaces PageMaker) and Quark Xpress (providing Unicode support only recently in versions 7-8 only) provide adequate Unicode support. Even though as this paper is written it is rare to find application programs that do not support Unicode, caution must be exercised to make certain of Unicode compliance when using early versions of software, or software that is not widely used. For example, SIL database software “Shoebox” is not Unicode-compliant, while the replacement version, “Toolbox”, is Unicode-compliant. Earlier versions of Macromedia Flash (now Adobe) or Notepad were not working properly with Unicode despite their claims. They now are Unicode-compliant, although special Unicode settings often have to be chosen before saving documents.

A website for East Cree language resources, including an on-line dictionary, has been under development since the year 2000; it has proven to be fertile ground for testing many applications. This testing, along with reporting of errors to software designers, has led to better compatibility over the years. The database-driven applications build with MySQL, PHP, and Python for the eastcree.org project, have proven to work well with Unicode.

Unicode on the World-Wide-Web

In the early 2000s, Unicode compliance on Internet browsing software was hit-or-miss. However, increasing levels of Unicode support have been implemented for most common browsers, such as Microsoft’s Internet Explorer, Mozilla Firefox, Macintosh Safari, and Google Chrome.

In a similar way, Web-based messaging and e-mail programs have also implemented Unicode support in recent years. Still, it is prudent to test various applications before deploying syllabic-script on the World-Wide Web.

Display issues

Until recently, users still had to ensure that target computers actually had the desired font installed when sharing documents between computers or across the Internet. However, newer computer operating systems come with a Unicode-compliant syllabic font in a standard

installation. Tiro Typeworks designed *Euphemia*, a Unicode font with a full set of Roman and Canadian Syllabic characters which newer operating systems will use by default (standard since Macintosh OS X.2 and Vista). Users who run older versions of the operating systems (such as Windows XP) can download and install Euphemia from the Tiro Typeworks website.

Another way to acquire a Unicode syllabic font is to use the ones distributed with keyboard utilities: BJCree UNI is installed automatically with the keyboarding resources CreeKeysDesktop3, CreeKeysPro3 and McCree3. Any document written in syllabics that is encoded in Unicode can be displayed in syllabics if a Unicode syllabic font is installed in the computer.

While Unicode was developed in order to assign unique character codes to every glyph (every character shape) in every language of the world, most Unicode fonts only contain a fraction of the total number of characters worldwide. Fonts are developed this way for a practical reason-- fonts that contain thousands of glyph-shapes are large computer files: Every font used in an application or document uses up some computer resources of memory and processing speed. Using any fonts with very large character inventories will significantly slow down or crash computer systems. For this reason, fonts that are developed to be used with Cree are normally limited to the standard (Roman, alphabetical) characters plus only those characters used in Cree.

With this in mind, there are a number of other Unicode syllabic fonts available that can be used to display Cree texts. Euphemia (already described above) contains the entire inventory of Canadian Syllabics. BJCree UNI only contains those characters necessary for Eastern and Western varieties of Cree, along with Naskapi, but does not contain those characters unique to Dene or Inuktitut. The appendix of this paper contains a current listing of Unicode fonts that have an adequate set of characters for varieties of Cree.

Syllabic characters can also be displayed on computers that do NOT have a syllabic font installed, when the fonts are embedded by the software program used to author the document. Fonts can be embedded in documents using Microsoft Word, Adobe Acrobat PDF, and OpenOffice writer, which can save as PDF. This way the document itself contains the font outlines, but the user need not have the actual font installed on their computer.

A new technology recently introduced is the embedding of fonts in websites, using software such as The Dojo Toolkit, a modular open source JavaScript toolkit or library. This way the web server contains the font outlines, but, again, the user need not have the actual font installed on their computer.

Keyboarding issues

Simply having a font installed only allows the display of syllabics. In order to input syllabics, that is, to use a keyboard to enter the syllabic characters desired, a separate software product must be installed. This provides an interpreter so that the keys pressed call the syllabic character to the screen. Operating System specific solutions include the following: for Windows, Tavultesoft Keyman and Keyman Desktop (CreeKeys) and for Macintosh OS X.3, the "International" dialog in System Preferences provides access to built-in and custom keyboard layouts (.keylayout) HTML files (McCree). Since Macintosh OS X.6 the dialog in System Preferences is referred to as "Language and Text" rather than "International".

Two different approaches to keyboarding syllabics have been employed: The most simple is to replace individual keystrokes on the standard keyboard with the syllabic character desired. One benefit of this approach is that the keyboard can be roughly laid out to approximate the syllabic chart (top row for the "e-vowel" series, second row for the "i-vowel", etc. (See the syllabic chart in the appendix.) An important drawback is the necessity to replace the keytop markings with marks corresponding to the syllabics keyboard layout, and also forcing the user to learn a new layout.

The other approach to keyboarding is to match the roman spelling of the word in syllabics to the keystrokes. That is, the typist presses the sequence of keys that correspond to the way the word "sounds", phonemically, using roman spelling, in Cree. The keyboard program takes care of converting the roman spelling to syllabics. The benefit of this approach is that there is no new keyboard to learn, the letters on the keys can stay the way they are, and new users can learn to type very quickly. The drawback of course is that roughly twice as many keystrokes are required to type any given word in syllabics.

Using a keyboard re-mapping program like Tavultesoft Keyman or the Macintosh custom keyboard layouts under the “International” menu provide the user with the option of choosing whichever keyboarding style is preferred.

Tavultesoft Keyman is a commercial product that allows the creation of custom keyboards. Earlier versions of the software (version 5.0) allowed a registered user to create keyboard packages that could install multiple keyboards and fonts on end-users computers, without cost to non-governmental and non-commercial users. Tavultesoft has released subsequent versions (version 6.0 and above) that do have certain licensing restrictions on end-users. For example, to all users of current versions of the software are required to purchase a licence from Tavultesoft. The Windows package that has been distributed on the eastcree.org website, named *CreeKeysPro3*, was based on Keyman version 5.0. So far, most Cree users have found it unnecessary to upgrade from Keyman version 5.0. Google Chrome (up to current version 9) and Opera (up to current version 11) have full support on Windows 7, Vista and XP. However, the current versions of certain browsers do not work with Keyman version 5.0. Specifically, Internet Explorer 8 on Windows 7, Vista, and XP supports Keyman 5.0 when you type in the address bar and search boxes but does not work when you type into a field in a page on the web. For Firefox (all Windows versions), Keyman removes the character just typed when inserting multiple-keystroke syllabics. Tavultesoft has fixed this behaviour in newer versions of Keyman.

In 2009, Tavultesoft has offered to update the *CreeKeysPro3* package to correspond with the current version of Keyman Desktop 7.1. This new package requires users to purchase a licence from Tavultesoft, but addresses several Windows language and input-method compatibility issues. This new package, called *CreeKeysDesktop3*, is available for general distribution on Tavultesoft’s website, and includes technical support from Tavultesoft. See the Tavultesoft website in the References section below for more information.

For the Macintosh, software from SIL International (Ukelele and KeyLayoutMaker--both free of charge) allows users to create their own open-source keyboard layouts. These programs were used to create the Cree keyboard layouts for Macintosh that correspond to the keyboards available for Windows.

The Macintosh package distributed on the eastcree.org website, named *McCree3*, contains the Unicode fonts and keyboard layouts.

Web-based keyboarding solutions are also available: a homemade Flash applet was developed for the on-line dictionary on the eastcree.org website. This applet interprets the typed sequence of Roman characters and converts them into East Cree syllabic characters.

Although the Canadian syllabic inventory is represented relatively consistently as an orthography across the country, there are some important differences in usage that must be taken into account. For example, communities on the western side of Hudson's Bay tend to use a different set of syllabic "finals" than communities on the eastern side use. Also, the placement of diacritics (the so-called "w-dots") is different depending on the speech community or dialect. Normally, a local authority, such as a school board or education department has established which syllabic standard is adhered to by a community or group of communities. It is not within the scope of this paper to provide a full accounting of these differences, but it should be understood that keyboard entry-methods are necessarily different, depending upon which syllabic orthographic tradition a given region has adopted. The examples and resources in this paper are mainly for "East Cree", the communities of Cree speakers across northern Quebec. For other varieties such as Woods Cree or Swampy Cree, the syllabic fonts will be adequate to display the language on-screen. However, the East Cree keyboard entry methods will not place diacritics properly nor output the appropriate syllabic "finals" for Western Cree.

It is recommended that users who are interested in keyboard entry methods for syllabics used in languages other than East Cree refer to the website resources listed at the end of this paper.

Conversion issues

The standards and dominance of English make database encoding preferable in Roman orthography. There is thus a need for efficient conversion tools for database activity, for example, in the eastcree.org dictionary developments, we use conversions for keyboarding and for search engines.

In addition, older fonts need to be converted to newer (Unicode) fonts: valuable pedagogical material or texts are often encoded in an older font and it would take too long to retype.

The development of “simplified Roman” on text messaging or for other communicative purposes is out of sync with “convertible” standard Roman orthography. There is an increasing need for developing spelling checks and relaxed search tools that allow users to produce consistent results even though they may make common spelling or orthographic errors while keyboarding. (See the dictionary search for eastcree.org, described in Junker & Stewart, 2008)

The East Cree keyboarding packages CreeKeys and McCree contain Visual Basic (VBA) encoding conversion programs that are accessible to average users. These programs make it possible to convert texts keyboarded in the older 8-bit “legacy” fonts to be converted to Unicode for archiving, standardization, and placement on the Internet. Unfortunately, these programs currently only function on MS Word, because VBA is a proprietary program. At this time, these VBA conversion programs do not work on open-source software such as OpenOffice. See the training videos noted in the References section below for tutorials in how to use these conversion programs.

Distribution issues

Distribution is also an issue. Often computer stores and companies do not know what is available, or what is best for minority languages. A Cree speaker will usually not find advice or help about Cree fonts from a retailer. One solution designed to meet this need was to develop a resource page on www.eatcree.org with downloadable packages, training videos, and Frequently Asked Questions (FAQ), which address as many of the issues mentioned above.

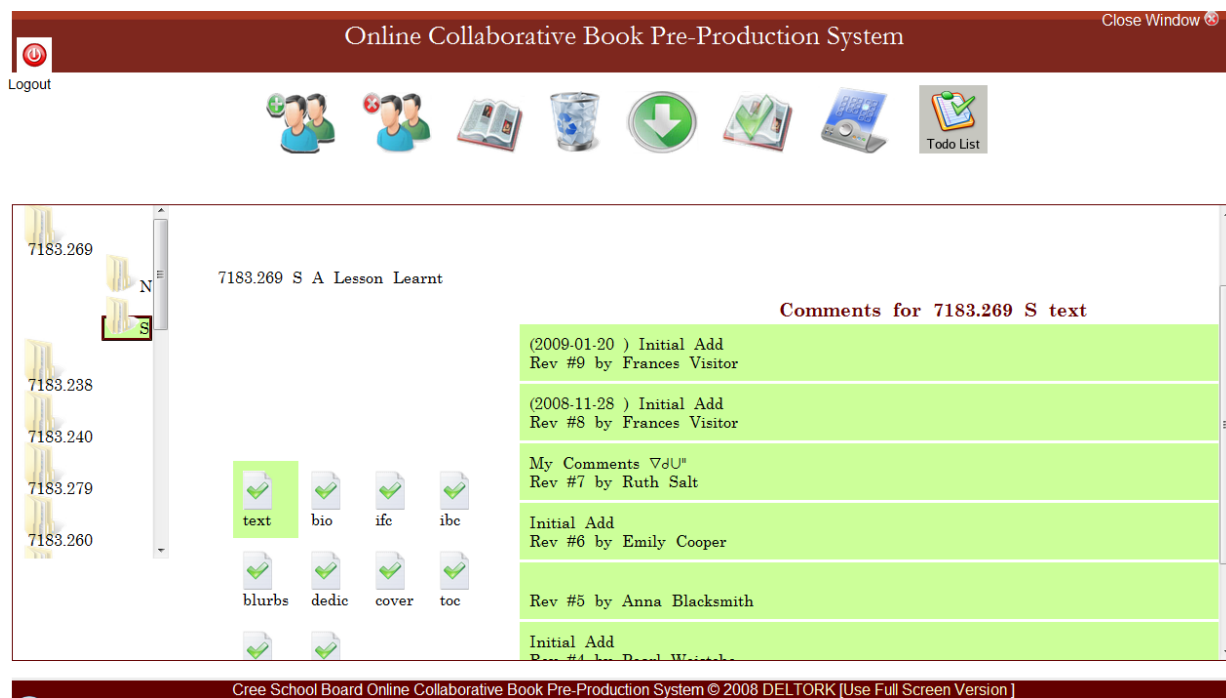
These pages also contain links to font websites, where users may download all available and free Unicode fonts that work for the East Cree language. However, it remains the user’s task to identify their particular problem, find the solution, and then apply the solutions to their problem.

Speakers have reported to us that they sometimes have encountered problems when using other websites with obsolete information or material. In general, although well intended, such initiatives that offer resources need to indicate dates of last updates. Any resource website is faced with the problem of long-term maintenance: who is and who will be responsible for regular updates?

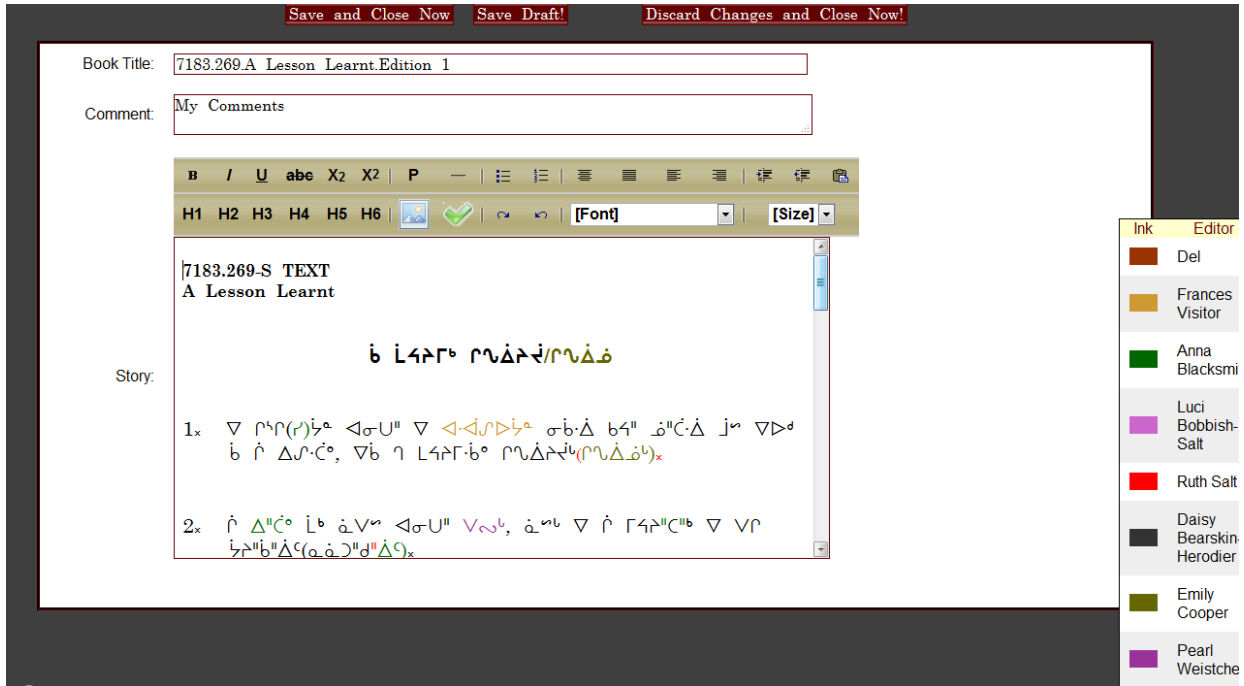
Collaborative IT Applications

Within the East Cree speakers' community and especially in the wake of the collaborative project eastcree.org², several interesting IT collaborative applications have proven successful. They all allow educators and language specialists to collaboratively produce resources and material for supporting the transmission of the Cree language and culture, while working at a distance, in communities that are sometimes thousands of kilometres apart. They all involve web-based editing, and typing in Cree syllabics in browsers. They include:

- Image and sound database for the encyclopaedic dictionary, developed in 2007, this allows for uploading of images and sounds to the on-line Cree dictionary.
- Editing for book production: This application is the most recent (developed on eastcree.org in 2008 by Delasie Torkornoo): staff members at Cree programs are now using an on-line editing tool (Google docs style) for creating and producing their school books in Cree syllabics.



² The eastcree.org project started in 2000 as an effort to develop language resources for the East Cree teachers and speakers exploring how IT could help language documentation and preservation. The nine East Cree communities in Northern Quebec have their own school board, the Cree School Board, whose Cree Programs department became a partner to the project in 2001. The maintenance of the site is progressively being transferred to the School Board, while the development of the resources continues to involve Cree language specialists and linguists.



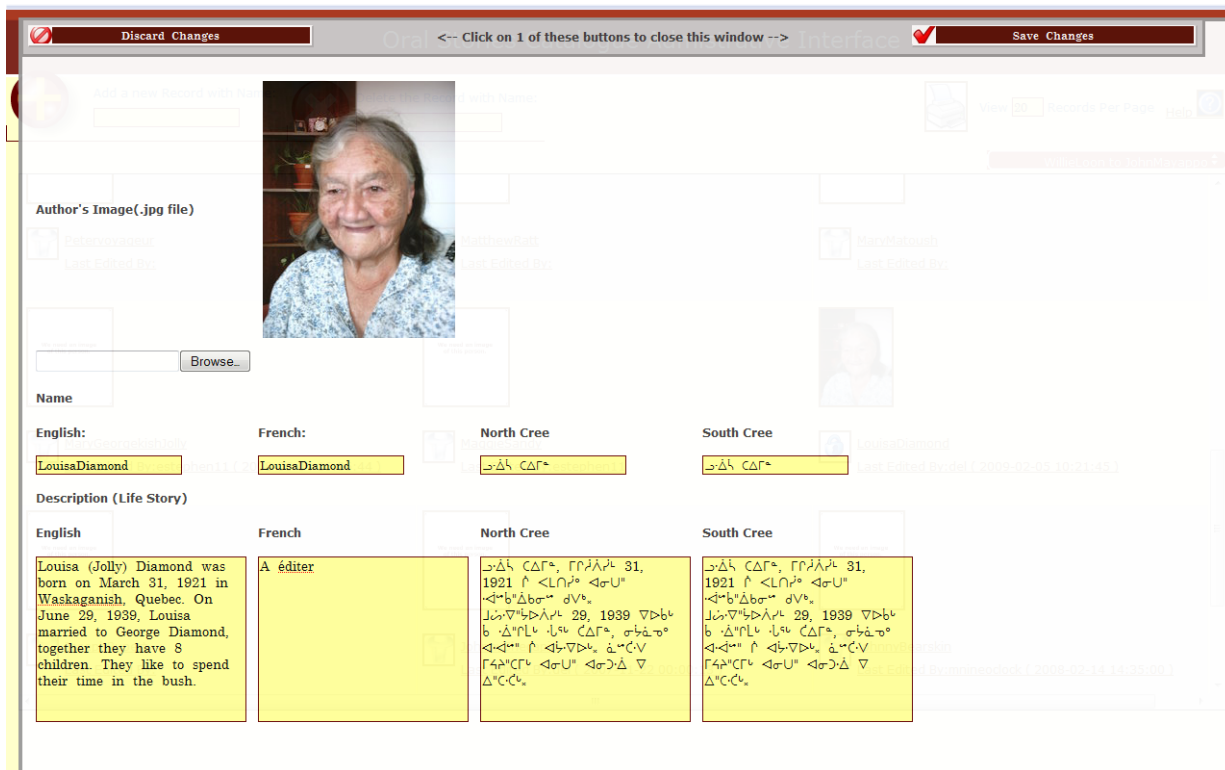
- Editing and maintenance of an on-line Educational Resources catalogue. This application was the first one, developed in 2002, and served as a successful model for later versions. It is widely used by educators in the Cree communities.

PREPUBLICATION DRAFT (revised March 2011). To appear as:
 Jancewicz, Bill, & Junker, Marie-Odile. Cree syllabic fonts: development, compatibility and usage in the digital world. In Valentine, Randolph (ed.) *Papers of the 40th Algonquian Conference*.

The screenshot shows a web-based form for entering book metadata. The form is organized into several sections:

- Metadata:** Includes fields for ISBN (1-55036-599-1), TYPE (Book), Number of Copies (18), Publication Year (1998), Price (\$3 CDN), and Level (6).
- Thumbnail and Cover:** Fields for Thumbnail (.gif file) and Book Cover (.jpg file), with an option to "Automatically Generate thumbnail (GIF)".
- Topic:** A grid of checkboxes for various categories:
 - Family (checked), Traditions, Legends, Spring, Winter (checked), Places, Clothing, Trees, Transportation, Stories
 - Boy (checked), Celebrations, Memoirs (checked), Summer, Animals, Artifacts, Man, Water, Birds, Song
 - Girl, Past, Seasons, Fall, Plants, Food, Woman, Shelters, Fish, Other
- English fields:** Title (John Ross remembers), Author (Linda Visitor), Edited by (Linda Visitor, Luci Salt), Illustrated by (Morley Stewart), Photographer, and Description (The little boy packs lunch to check rabbit snares with mom and aunt).
- French fields:** Titre (John Ross se rappelle), Auteur (Linda Visitor), Édité par (Linda Visitor, Luci Salt), Illustré par (Morley Stewart), Photographe, and Description (Un petit garçon prépare son diner pour aller lever des collets de lièvre avec sa mère et sa tante).
- North Cree fields:** Fields for title, author, and description in North Cree syllabics.
- South Cree fields:** Fields for title, author, and description in South Cree syllabics.
- Source:** Checkboxes for Inuit Broad. Corp., Cree Programs, English, Nemaska, Inuktitut, French, Cree Literacy, Plain, Northern Dialect, and Southern Dialect.
- Also Available in:** Checkboxes for English, Northern Dialect, French, and Southern Dialect.
- MISC:** A checkbox for "Recently Published".

- Editing and maintenance of an Oral stories database with an Elders' biographies database. Developed in 2003, this database now contains over 500 oral stories and is still being populated (see Junker and Luchian (2007), for a discussion). It is also widely used.



Conclusion

Like standardisation of orthography, standardisation of technology is essential. Changes in the digital world have allowed the widespread use of Unicode, but the question of what are the best tools and whether they can be understood and accessed by the people who need them, remains. As minority languages, Cree and other aboriginal languages remain vulnerable to the dominance of English in the domain of software development, hardware preferences for certain platforms, and ever-changing technologies. As is the case for the development of practical orthographies, the appropriate and successful development of language-specific information technology tends to occur at the grassroots level, and is informed by specialized (often non-commercial and non-mainstream) knowledge. Standardization of technology can only be achieved if there is cooperation between all the stakeholders and a routine dialogue and collaboration between developers and users. It is our hope that this paper will at least in part continue this dialogue.

Eastern James Bay Cree Syllabic Chart

▽ e		△ i	△̇ ii	▷ u	▷̇ uu	◁ a	◁̇ aa			o	u	h
	·▽ we	·△ wi	·△̇ wii	·▷ wu	·▷̇ wuu	·◁ wa		·◁̇ waa				
∇ pe	·∇ pwe	∧ pi	∧̇ pii	∩ pu	∩̇ puu	◁ pa	◁̇ paa	·◁̇ pwaa	<	p		
U te	·U twe	∩ ti	∩̇ tii	∩ tu	∩̇ tuu	∩ ta	∩̇ taa	·∩̇ twaa	c			
q ke	·q kwe	∩ ki	∩̇ kii	∩ ku	∩̇ kuu	∩ ka	∩̇ kaa	·∩̇ kwaa	b	d	kw	
∩ che	·∩ chwe	∩ chi	∩̇ chii	∩ chu	∩̇ chuu	∩ cha	∩̇ chaa	·∩̇ chwaa	∩	ch		
∩ me	·∩ mwe	∩ mi	∩̇ mii	∩ mu	∩̇ muu	∩ ma	∩̇ maa	·∩̇ mwaa	∩	m	mw	
∩ ne	·∩ nwe	∩ ni	∩̇ nii	∩ nu	∩̇ nuu	∩ na	∩̇ naa	·∩̇ nwaa	e	n		
∩ le	·∩ lwe	∩ li	∩̇ lii	∩ lu	∩̇ luu	∩ la	∩̇ laa	·∩̇ lwaa	∩			
∩ se	·∩ swe	∩ si	∩̇ sii	∩ su	∩̇ suu	∩ sa	∩̇ saa	·∩̇ swaa	∩	s		
∩ she	·∩ shwe	∩ shi	∩̇ shii	∩ shu	∩̇ shuu	∩ sha	∩̇ shaa	·∩̇ shwaa	∩	sh		
∩ ye	·∩ ywe	∩ yi	∩̇ yii	∩ yu	∩̇ yuu	∩ ya	∩̇ yaa	·∩̇ ywaa	∩	y		
∩ re	·∩ rwe	∩ ri	∩̇ rii	∩ ru	∩̇ ruu	∩ ra	∩̇ raa	·∩̇ rwaa	∩	r		
∩ ve	·∩ vwe	∩ vi	∩̇ vii	∩ vu	∩̇ vuu	∩ va	∩̇ vaa	·∩̇ vwaa	e	v,f,ph		
∩ the	·∩ thwe	∩ thi	∩̇ thii	∩ thu	∩̇ thuu	∩ tha	∩̇ thaa	·∩̇ thwaa	∩	th		

References

- Jancewicz, Bill and Marie-Odile Junker. 2002. Cree on the Internet: How to Integrate Syllabics with Information Technology and the Web. Presented at the *34th Algonquian Conference*, Queens University. <http://www.eastcree.org/pdf/FAQ.pdf>
- Junker, Marie-Odile and Radu Luchian. 2007. Developing web databases for Aboriginal language preservation. *Literary and Linguistic Computing* 22(2), June 2007:187- 206.
- Junker, Marie-Odile and Terry Stewart. 2008. Building Search Engines for Algonquian Languages. In Karl S. Hele and Regna Darnell (eds). *Papers of the 39th Algonquian Conference*. London: University of Western Ontario Press: 378-411.