A Survey on the i18n Status of བོད་སྐད་

Elena Cabrio and Daniele Pighin
{cabrio,pighin}@fbk.eu

May 2008

1 Introduction

In this paper we will try to assess the current status of internationalization (i18n) and localization (l10n) of computer software with respect to the Tibetan language, or བོད་སྐད་ as a Tibetan user would write it.

We decided to focus on Tibetan for several reasons. First, recent political events in Tibet are drawing the public opinion attention towards this region and the difficult situation in which its political, economical and cultural identities are. As access to the consumption and production of information, and especially of *digital* information, may be a critical issue in this respect, we would like to understand how easy it is for an average Tibetan user to use a computer to produce culture or revenues and to share knowledge and information with the outside world.

Second, Tibetan has a complex script that makes it not-trivial to support its written form in a computer system, as special encodings, collation rules, fonts and input methods need to be available.

Finally, as Tibetan is spoken by about six million people, it is a language with a user base comparable to those of Norwegian or Romanian, which are generally quite well supported by computer systems. It is arguable that six million speakers may not be enough to justify a per-profit software house (such as Microsoft, or Sun or Apple, just to name a few) investing in the development of localized products, but are definitely enough to sustain a community of volunteers working on software localization. For this reason, in the discussion we will be mostly focusing on free or open-source software solutions, with eventual pointers to the proprietary software world.

During our research we have happened to notice that Dzongkha, which is a variety of Tibetan only spoken by a few hundred thousand people, is more widely supported than Tibetan itself. For this reason, it will often be considered in the rest of the discussion as a "special case" of support for Tibetan.

After a brief introduction on Tibetan politics and economy (Section 2) the description of the current internationalization status of the language has been made in terms of technical glossary (Section 3), locale (Section 4), writing system (Section 5), character encoding (Section 6.1), fonts (Section 6.2), keyboard layout and input methods (Section 6.3), spellchecker (Section 7) and availability of localized software (Section 8). Section 9 discusses our attempt at configuring a computer to be used to input and display Tibetan text. Finally, Section 10 summarizes the content of the paper and draws some conclusions.

2 A bit of history

Nowadays, Tibet as a whole is not self-sufficient; it has the lowest economic output of any region in China, and a million residents are still below the poverty line of $150 in annual income. China's critics and Tibetan exiles blame Tibet's poverty on Beijing for stripping Tibet of its resources and neglecting its people's welfare. Actually, the economy of Tibet is dominated by subsistence agriculture: there is very little arable land available and the main crops grown are barley, wheat, buckwheat, rye, potatoes and various fruits and vegetables. Much of the population is engaged in

This work is released under the terms of the Creative Commons Attribution-Share Alike 3.0 License.

a pastoral life (sheep, cattle, goats, camels, yaks, donkeys and horses), but the advances made by irrigation and the growing of forage crops is decreasing the amount of nomadism.

In recent years the economy has begun evolving basing on the development of both agriculture and tertiary industry: therefore, the existence of localised software could help improvements especially in these fields, allowing a broader diffusion of Information Technologies also towards people who only know their native language (87% of Tibetans live in rural areas where Chinese is rarely spoken).

Furthermore, primary education is conducted either primarily or entirely in the Tibetan language, and only during middle school and secondary school bilingual education (Tibetan, Chinese) is introduced. Within this frame, the use of localised software could offer the possibility to acquaint children and people with only primary education with the use of computers.

Tibetan language. The group of languages spoken in Tibet, Bhutan, Nepal, and in parts of northern Pakistan and India are generally known as Tibetan [1], which is classified by linguists as a member of the Tibeto-Burman subgroup of the Sino-Tibetan languages. It includes various regional dialects and sub-dialects, and the boundaries between Tibetan and other Himalayan languages are not always straightforward. In general, the dialects of central Tibet (including Lhasa), Kham, Amdo, and some smaller nearby areas are considered Tibetan dialects, while other varieties (e.g. Dzongkha, Sikkimese, Sherpa, and Ladakhi) are considered by their speakers as separate languages (for political reasons). As said before, this agglomerate of dialects is spoken by about 6 million people; Lhasa Tibetan, the most influential variety of the spoken language, is spoken by about 150,000 people.

The morphology of Tibetan can generally be described as *agglutinative*, i.e. most words are formed by joining morphemes together, while in Classical Tibetan syntax and meaning are shaped more by use of particles and word order rather than by inflection (*analytic* language). Tibetan is an *ergative* language and grammatical constituents usually maintain head-final word order. Unlike Classical Tibetan, some dialects have developed tones.

3 Technical glossary

As necessary condition, the dissemination of modern science and technology in a country needs the language to be an intermediary: therefore, a big challenge for the Tibetan language is how to follow up on the trends of the times, covering the linguistic gaps caused by the introduction of objects and concepts unknown until now. Over the past decades, a large number of new terms have appeared in the Tibetan vocabulary concerning scientific and technological fields, created by lay people, academics and translation specialists. Anyway, although a great number of new terms have come into existence, a huge gap due to the speed of progress in world science and technology is still present. Since the development of science and technology is continuous, it is impossible to work out all the technical terms once and for all.

Online lexical resources principally consist of Tibetan to English dictionaries (see [15] or [16] or [17]). Consulting the only one English to Tibetan online dictionary [18] almost no terms concerning computer science are present (only *computer* and *mouse* - probably the animal - but not *keyboard*, *PC* or *software*). A complex process of standardization of technical terms is needed: it would require all technical terms to be examined and approved in accordance with the academic theories of each sub-branch of discipline, to avoid confusion that might lead to misunderstanding [14].

4 Locale

The Common Locale Data Repository (CLDR) [11] is the Unicode initiative for the standardization and collection of locale data. At the end of July 2007 the version 1.5 of the repository data was released, including for the first time Tibetan locale files (bo) and support for the Chinese (bo_CN) and Indian (bo_IN) variants. Figure 1 shows an excerpt of the CLDR Tibetan locale description.

Some software projects, like OpenOffice.org, maintain independent locale formats and collections. The "The Locale Generator (LocaleGEN)" [12] web tool simplifies the creation of locale files

```
<dayWidth type="wide">
  <day type="sun">གཟའ་ཉི་མ</day>
  <day type="mon">གཟའ་ཟླ་བ</day>
  <day type="tue">གཟའ་མིག་དམར</day>
  <day type="wed">གཟའ་ལྷག་པ</day>
  <day type="thu">གཟའ་ཕུར་བུ</day>
  <day type="fri">གཟའ་པասངས</day>
  <day type="sat">གཟའ་སྤེན་པ</day>
</dayWidth>
```

Figure 1: An excerpt of the CLDR Tibetan Locale file.

for OpenOffice.org which can eventually be converted into CLDR format. The localeGEN project lists three contributions for Tibetan (Chinese variant, bo_CN) but, to date, the locale has not been included in an OpenOffice.org release.

As already mentioned, the Tibetan variety Dzongkha is more widely supported than Tibetan itself: Dzongkha Development Authority (DDA), in consultation with computer vendors and language experts has indeed compiled a standard locale for Dzongkha for Bhutan (dz_BT). Dzongkha (also known as Bhutani) is now included as a distinct language within ISO 639 (language codes "dz" and "dzo") (see [10]), and in 2004 it appeared in CLDR 1.3 (it includes date, time, calendar formats, name of days, months and numbers, etc.).

Through the research efforts of Bhutan team of PAN Localization project, Dzongkha locale is supported on the Linux platform ([13], see Figure 2). Locale for GNU C library has been created and implemented in Linux operating system; support for locale and collation rules has been included in OpenOffice.org since version 2.0.

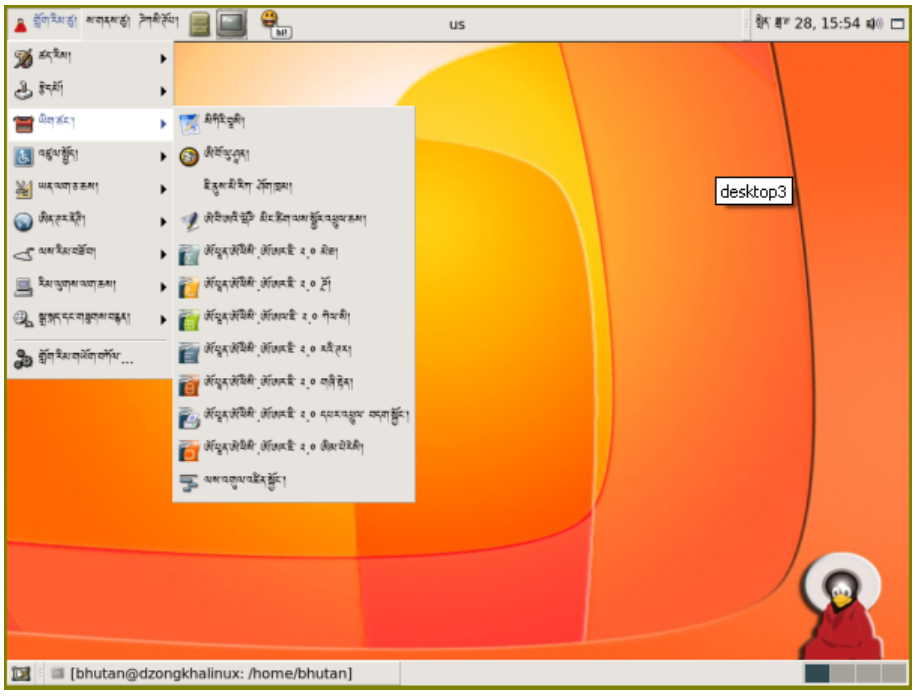

Figure 2: Screenshot of Dzongkha Linux desktop [13].

Concerning proprietary software, all versions until Microsoft Windows XP did not include the

locale definition for Dzongkha-Bhutan, while Vista does (it has the display name of "Dzongkha (Kingdom of Bhutan)").

5 Writing system

Throughout the centuries, Tibetan has been written using different types of scripts, among which the most common are *U-chen* and *U-me*. The first one has been used for centuries to print religious books and today it is still used for books, newspapers and other media, while *U-me* is the hand-written cursive form used in everyday writing, although it is sometimes used in books [2].

The Tibetan script is an *abugida* of Indic origin, i.e. a segmental writing system in which each letter (basic character) represents a consonant accompanied by an inherent vowel; other vowels are indicated by modification of the consonant sign, either by means of diacritics or through changes in the form of the consonant itself. The writing system derived from the pronunciation of the language as it was in about the 7th century, and varies in many ways from colloquial Tibetan as it is spoken today. Besides Tibet, the writing system is also used in Bhutan and in parts of India and Nepal.

The script is transliterated in a variety of ways. For instance, the *Wylie transliteration scheme* (defined by Turrell Wylie in 1959) uses the keys on a typical English language typewriter; it has become a standard transliteration scheme in Tibetan studies, especially in the United States. *Tibetan Pinyin* is instead the official transcription system for Tibetan in China: it is based on the Lhasa dialect and reflects the pronunciation very accurately, but it does not mark tones.

Similarly to European languages, Tibetan is written left to right and from top down. In writing it by hand, horizontal strokes are written from left to right and vertical strokes from top down; the top stroke of Tibetan letters (always drawn first) should align with the others; then, symbols are altogether adjusted in size and shape. Little dots (called *tsheg* ཚེག་) separates syllables. Vertical lines at the end of the phrase are called a *shey* (ཤེད་) and they are the closest form of punctuation to a period; when Tibetan is transcribed, a *tsheg* most frequently simply becomes a space. Sometimes a dash is used in order to say how the words are separated.

6 Technological requirements for Tibetan script

Using a script in a computer system requires that:

- the operating system knows the glyphs (i.e. the shapes and meanings) that make up a script;
- there is some standard to assign each distinct glyph a binary representation;
- the hardware and software of the computer system allow the user to input and visualize the glyphs of a script, according to the rules it is subject to.

This last point requires that the system has proper fonts installed, that the applications can support the script requirements in terms of spacing, ligatures and so on, and that input devices for the script language exist. The remainder of this section will try to assess the status of Tibetan i18n with respect to these issues.

6.1 Character encoding

The Unicode initiative is involved in an incremental effort to assign a unique binary representation to every glyph used in human scripts. Languages represented in the Unicode standard can be safely interchanged between computer systems as the standard itself defines a one-to-one correspondence between a glyph and its binary representation.

Tibetan script glyphs are encoded in the Unicode standard in the character range U+0F00-U+0FFF. Included are letters, digits and various punctuation marks, as well as special symbols used in religious texts, as shown in Figure 3. Religious texts also include substantial portions of transliterated Sanskrit, which when represented with Latin characters require a wide range of

Figure 3: Unicode 5.1 character code chart for the Tibetan language [4].

diacritic marks to be properly displayed. All the necessary diacritics are encoded in the following sections of the Unicode standard [19]:

- Basic Latin: U+0000 – U+007F;
- Latin-1 Supplement: U+0080 – U+00FF;
- Latin Extended-A: U+0100 – U+017F;
- Latin Extended-B: U+0180 – U+024F;
- Latin Extended Additional: U+1E00 – U+1EFF.

Full Support for Unicode Tibetan encoding was introduced in Windows in 2007 with the Vista release of the OS. Support in Mac OS X was introduced in 2008 with Leopard 10.5 [27]. For Linux systems, support for Tibetan Unicode dates back to December 2004 with the release of Pango v1.8.0 [23] and Qt v3.2 [22]. These open source frameworks, which are able to display complex scripts and render Unicode glyphs, form the backbone of the internationalization facilities on modern Linux desktops.

6.2 Fonts

The rendering of Tibetan script requires fonts providing glyphs for all the needed characters and diacritics, and the possibility of combining glyphs in ways far more sophisticated than needed for Latin fonts. There are several font technologies that can handle the degree of complexity required to render non-Latin languages correctly, such as Apple's Apple Advanced Typography (AAT) [30] and Microsoft's Open Type [31].

OpenType is an extension of the widely spread TrueType font technology, upon which it adds several options to enhance the font's typographic and language support capabilities. OpenType

Figure 4: The standardized Tibetan keyboard layout [42].

has several distinctive features that allow it to cope with complex scripts, most notably it is based on Unicode and can support any script (or multiple scripts at once), each font can have up to 65,536 glyphs and the font files are intended to be cross-platform.

On the Web it is possible to find a small collection of OpenType fonts supporting Tibetan script glyphs and the required diacritics. Among these, the "Tibetan Machine Uni" and "Jomolhari" fonts [9] appear to be the most complete sets, they are widely employed and are available under the GPL and the SIL Open Font License, respectively.

For the typesetting of scientific documents, at least two packages can be used to include Tibetan script in a LaTeX document, namely *ctib* [34] (used for this document) and *otibet* [34], both coming with Tibetan fonts in MetaFont [34] format.

6.3 Input devices

The inclusion of a character set in the Unicode standard and the availability of good quality fonts allow users to access contents in their own script language. On the other hand, in order to be able to *produce* content in a given script form the computer system must provide facilities that make writing a script language ergonomic and convenient.

Traditional input systems have been designed around the needs of Western users: our keyboards only have to cope with (generally less than) thirty alphabet symbols plus ten digits, and a few key combinations are enough to include in a document all the needed diacritics and character variants. This is not the case with complex scripts in general, and Eastern languages in particular where the number of individual symbols can be in the order of the hundreds or the thousands.

This section presents an overview of the alternative methods currently available to input Tibetan script into computer documents.

6.3.1 Keyboard layout

In 2005, the China Electronics Standardization Institution (CESI) finally defined a standard for the layout of the Tibetan keyboard. Figure 4 shows how the standard mapped Tibetan characters onto the keys.

As a keyboard has 47 keys that can be assigned to individual characters, the standard defines five keyboards to cover all the 193 Unicode Tibetan characters. The first (main) keyboard allows direct access to the most frequent characters. The keys of the four secondary keyboards can be accessed using a combination of meta-character keys and keys from the main keyboard. In order to ease the learning of the position of the characters, the approach that has been followed is to (possibly) use the same key for all the variants of the same symbol, assigning the variants to the different keyboards. The assignment of symbols to keys and of variants to keyboards has been driven by statistical measurements of the characters' distribution in the Tibetan language [20].

6.3.2 Hardware keyboards

Hardware keyboard devices that allow direct input of Tibetan character codes do not seem to exist, yet there are very advanced solutions, mostly developed for gamers and IT professionals, that may be used as hardware input devices for symbol rich languages such as Tibetan.

One such example is the Optimus Maximum Keyboard [21], a stunning input device where each key consists of a small display. The action performed as well as the icon displayed on each key can be customized, and also the whole keyboard layout can be changed on the fly by pushing a key combo. The five Tibetan keyboards may be assigned to different logical layers of the device. The user would be able to hit the keyboard selection modifiers and visualize on the keys the symbols of the corresponding layer.

Of course, such devices are very expensive and energy hungry if compared with traditional keyboards, and hence not eligible to be used in low budget projects. Luckily enough, waiting for such technologies to become considerably cheaper or for less sophisticated native hardware input devices to be developed, there are software alternatives that allow a Latin keyboard device to be used to input non-Latin characters.

6.3.3 Software keyboards and input methods

Several software keyboards allow the user to input Latin character transliterations and output the resulting Tibetan script to some application programs [6, 5]. Just a couple of years ago, these tools were the only feasible approach to Tibetan scripting.

All modern operating systems provide some built-in facility to allow users from around the world to input text in virtually any script language, assuming that proper locale, Unicode standards and fonts are available.

For example, for this purpose Linux systems provide several inter-operable technologies, namely the X Input Method (*XIM*), the Unified Input Method (*UIM*) or the Smart Common Input Method (*SCIM*). With such technologies it is possible to define arbitrary mappings between keyboard layouts and Unicode characters, for example using a Latin keyboard to input transliterated text which is then rendered in the desired script.

Unlike the ad-hoc solutions above discussed, these frameworks are general and flexible, allowing different scripts to coexist in the same document. Most importantly, they allow the same input methods to be shared across different applications providing the user with a shared, consistent interface for complex script input.

7 Spellchecker

> *Always run your important documents through a spell checker. It will plane lee mark four you're revue, miss steaks ewe mite knot sea.*
>
> Source: http://lowfatlinux.com/linux-spell-checker.html

Neither GNU Aspell [35] nor Ispell [36], the two most widely used spellcheckers in the Linux world, support Tibetan script, but the latter lists Tibetan among the languages that can be technically supported by the spell-checking engine.

There are some spellcheckers for Tibetan that are part of specific Word Processing suites, such as those described in the next section, but it looks like there is no standalone product. Anyhow, it looks like these checkers actually work against transliterated text rather than Unicode script.

8 Software localization

In order to assess the localization status of software with respect to the Tibetan language, in this section we will focus on those applications that almost any user may be willing to use for her/his daily work and that are generally available on a standard Linux distribution, namely:

- desktop environments (KDE, Gnome);

- office productivity software (OpenOffice.org, KOffice);
- communication software (Skype, Pidgin);
- browsers and mail clients (Firefox, Thunderbird, Opera).

We will also introduce a few software products especially targeted on Tibetan users.

8.1 Desktop environments

The Gnome i18n team [24] classifies languages as:

- supported: more than 80% of the interface strings are translated;
- partially supported: between 50% and 80% of the strings are translated;
- not-supported: less than 50% of the strings are translated.

Tibetan does not appear in any of these categories, meaning that either there is no translation team currently at work on the language, or that such team has not contributed anything back to Gnome's repositories. Anyhow, Dzongkha is listed among the supported languages (as also shown by the screenshot of Dzongkha Linux in Figure 2).

As for the KDE project [25], a Tibetan localization team exists, but according to the stats the current state of string translation for the latest stable version of the desktop environment (4.0) amounts to 0%. Dzongkha (or even Bhutani) is not listed at all.

8.2 Office productivity

The OpenOffice.org l10n status page [26] lists entries for both Tibetan and Dzongkha. While it is not clear what is the translation status for the former (a team exists, but it does not seem to be operative), it looks like the latter was fully supported at the time of the release of OpenOffice.org 2.2. The software is currently at version 2.4, meaning that the interface is largely available in Dzongkha but with several areas of inconsistency, especially concerning new functionalities.

As KOffice l10n efforts are coupled with those for KDE, we can conclude that KOffice is not available neither in Tibetan nor in Dzongkha.

8.3 Communication software

Skype [37] is a proprietary, closed-source project whose development is not backed by a community of volunteers, hence it is much more complicated to understand what its l10n status is. Anyhow, searching on the official Skype site for the keywords "tibetan" or "dzongkha" did not produce any result. It seems than that it has absolutely no support for either languages.

Pidgin is an open-source, multi-protocol instant messaging client available for almost any platform. Its localization page [38] does not list Tibetan at all, but reports a translation status of almost 75% for Dzongkha.

8.4 Browsers and email clients

Opera is a proprietary web browser whose use is not very widespread, but it is very popular among web developers due to its very good support of web standards. Unluckily, it seems like it offers no support for either Tibetan or Dzongkha [39].

Firefox is a very well known open source web browser developed by the Mozilla foundation (a for-profit organization) with a large contribution from volunteering supporters all around the world. From the project's translation page, it is clear that neither Tibetan nor Dzongkha are supported languages [40]. The same goes for Thunderbird, the email client sponsored by the Mozilla foundation, who has no support for any of the two languages [41].

8.5 Ad-hoc software products

A few software products have been especially developed for the Tibetan user base. Meant to overcome the awkwardness of inputting Tibetan script in word processors or spreadsheet programs, these applications are mostly outdated now that advanced input methods and Tibetan locale, Unicode definitions and keyboard layouts are largely standardized. Nevertheless, as they allow to write texts in the traditional Pecha layout [1], these tools may still be useful for those users interested in this kind of writing. Two such examples are the Tibetan5 [28] and PechaMaker [29] word processors.

9 Our experience with Tibetan and Linux

A computer equipped with a modern operating system is mostly ready to be used by Tibetan native speakers. A modern Linux/Gnome desktop, for example, has all the libraries and utility software that may be needed to support Tibetan script in a consistent manner across all the applications.

To demonstrate this point, we have decided to try using Tibetan script on our English-based installation of Ubuntu Linux 8.04.

First, we tried to include Tibetan strings in this very LaTeX article. Downloading and installing the *ctib* package was a breeze, and also writing some strings was very easy, it is sufficient to write down Wylie transliterated text and the corresponding Tibetan characters are output.

Activating advanced input methods to be able to type Tibetan in desktop applications, such as OpenOffice.org or Firefox, was a bit more complicated as it required some configuration steps which were not obvious and not clearly documented. Different applications have built-in support for one of the three major desktop players in this area (namely: XIM, UIM and SCIM) and therefore it is not clear which one to go for. In the end, we had to install all the three of them and then activate some "bridging" functionality that allowed the three technologies to inter-operate. The procedure was a bit weird, but most likely starting with a fresh Linux install, and selecting the desired language support during the installation process, would result in a much smoother experience.

Figure 5: The Gnome Universal Input Method (UIM) control applet.

As we never happened before to toy around with a complex-input-enabled desktop, when we had it finally working we were quite impressed with the results. Figure 5 shows, on the left, the Gnome controls that allow the user to select "direct" or "composite" input, the latter allowing to intercept sequences of keystrokes that are then mapped onto Unicode character codes. The control shown on the right allows to select what mapping to use: in this case, we could use Wylie transliteration to input Tibetan Unicode characters.

This approach is very flexible, the user can assign a key-combo to switch among different input methods, and Unicode support in applications allows different scripts to be mixed with very little effort.

Figure 6 shows on the background the PDF output of this LaTeX document and in the foreground a text document in OpenOffice.org in which we have written བོད་ སྐད་ again. The screenshot also shows that we were using the Tibetan Machine Unicode font.

Then we wanted to try creating a spreadsheet using mixed scripts. The result is shown in Figure 7, where we have written a small table containing English words, their Wylie transliteration (as reported by the English to Tibetan Online Dictionary) and the resulting Tibetan script.

[1] Pecha is a layout used for religious books in which text is written horizontally in the body of the page and vertically on the side margins. Pecha also implies special boxes to be drawn around the structural elements of the page (main body, header, footer, margins). See [29] for a graphical example.

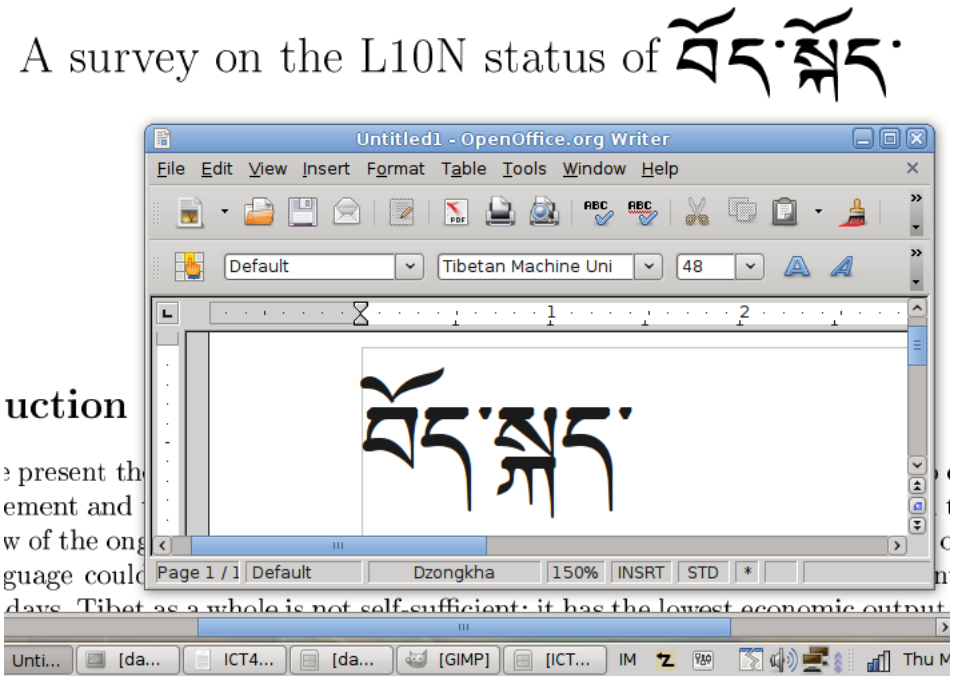

Figure 6: Support for Tibetan script in Latex (on the background) and OpenOffice.org 2.4 under Ubuntu 8.04.

Finally, we wanted to test the support provided by Firefox for Unicode pages written in Tibetan script, so we visited a Tibetan page, with both title and contents in Tibetan script, and we used the advanced input method to search for a sequence of Tibetan characters in the text. The screenshot in Figure 8 shows the displayed page, with the search term in the search boxes and the matching string highlighted in the page.

All this is a demonstration of the current status of support for Tibetan input and visualization on a standard computer. The downside of Tibetan software i18n is in the l10n part, as it seems very difficult to find software products whose interfaces have been translated into Tibetan.

10 Conclusions

In this paper we have surveyed the degree of support that a modern computing environment can offer to a Tibetan user. At the end of our experience we can draw the following conclusions.

First, Tibetan script complexity is greater than scripts based on Latin alphabets, but not as technologically challenging as traditional Chinese or other Eastern languages, which may require totally different text layouts and thousands of distinct glyphs to be rendered. Its less than 200 distinct characters make it feasible to design good quality fonts with a relative effort and to conceive reasonable keyboard layouts. Indeed, a decent supply of Tibetan fonts is available for free use, the language glyphs made it to be encoded in the Unicode standard and a keyboard layout has been standardized.

As for the possibility of producing content in Tibetan script, modern computer systems are generally built around full-fledged internationalization frameworks that allow non-Latin glyphs to be input quite conveniently, and which are flexible enough to cope with the ligature and visualization requirements of Tibetan script.

On the other hand, we have observed that it's very difficult to find applications whose interfaces have been translated into Tibetan, and the same goes for a Tibetan spellchecker. Curiously enough, it is more frequent to find software translations for Dzongkha (indeed, almost completely localized operating systems such as Dzongkha Linux), although it is spoken by a much smaller number of people. The reason behind this may be that Dzongkha is the official language of Bhutan, and

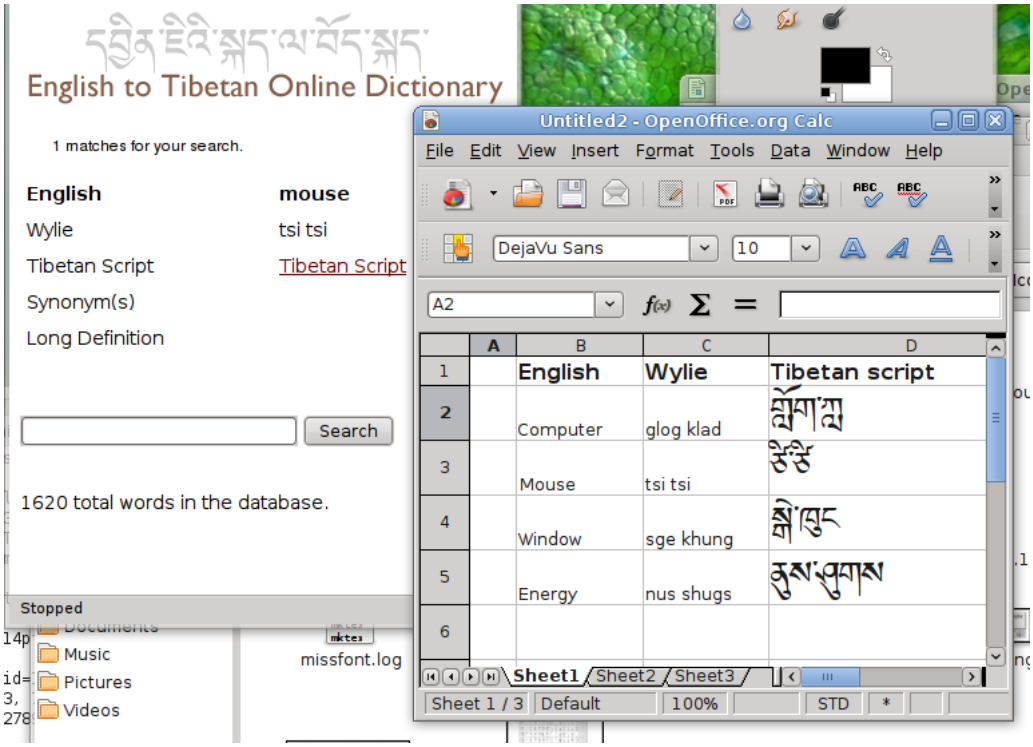

Figure 7: A spreadsheet with mixed Latin and Tibetan scripts.

as such localization efforts may be supported by Bhutanese national authorities. Conversely, it's easy to imagine that the People's Republic of China may not be as much supportive in backing localization efforts for the Tibetan people.

Anyhow, localization itself is the least technologically critical aspect of internationalization. As all the needed infrastructure is there, with a reasonable effort it should be possible to translate core desktop elements and applications into Tibetan, so that also the less educated people may conveniently access the instruments for the production and consumption of digital contents.

The lack of a technical glossary is also a critical aspect to be considered for software localization. Indeed, it may reflect a cultural gap in the field of computer science and the necessity to develop a local shared knowledge in these matters before attempting a large-scale translation process.

References

[1] *Tibetan language.*
http://en.wikipedia.org/wiki/Tibetan_language

[2] *Tibetan script.*
http://en.wikipedia.org/wiki/Tibetan_script

[3] *About the Unicode Character Database.*
http://www.unicode.org/ucd/

[4] *Tibetan Unicode.*
http://www.unicode.org/charts/PDF/U0F00.pdf

[5] *An Online Tibetan Keyboard.*
http://www.incks.com/en/tibetan.html

[6] *Tibetan!: a Tool to Input Tibetan Text in Corel WordPerfect and MS Word.*
http://www.tibet.dk/tcc/tibetan5.htm

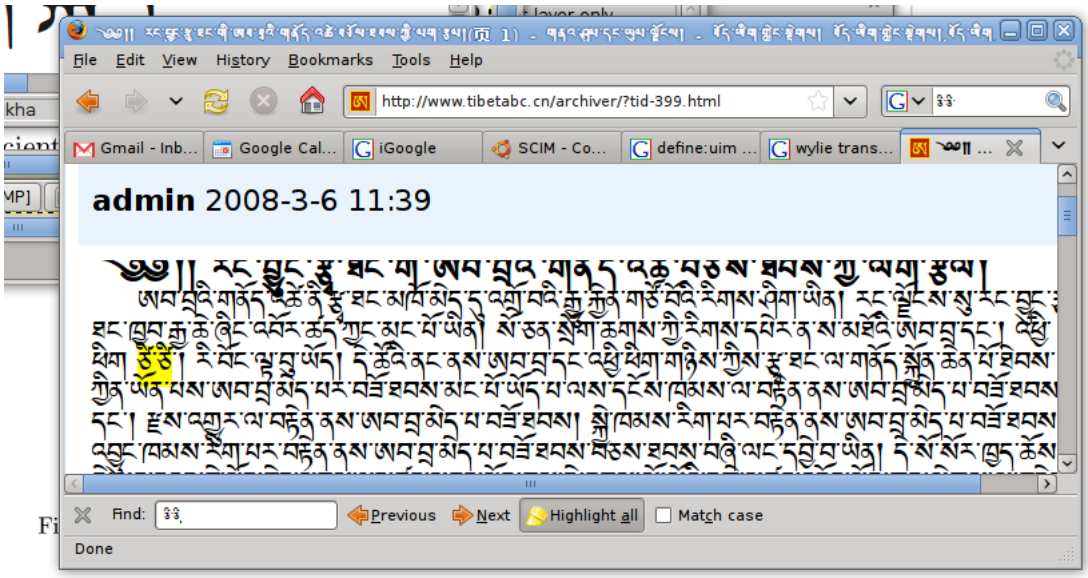

Figure 8: Firefox showing a Tibetan page, with Tibetan title and contents and highlighting a search term input in Tibetan script.

[7] *Survey of Language Computing in Asia.*
http://www.panl10n.net/english/outputs/Survey/Dzongkha.pdf

[8] *The Tibetan and Himalayan Digital Library.*
http://www.thdl.org/

[9] *OpenType Tibetan Fonts from THDL.*
http://www.thdl.org/tools/toolbox/index.php?#wiki=/wiki/site/26a34146-33a6-48ce-001e-f16ce7908a6a/tibetan%20fonts.html

[10] *Official National Standard for the Dzongkha-Bhutan Locale.*
http://www.dit.gov.bt/guidelines/locale_culture.pdf

[11] *The Common Locale Data Repository.* http://unicode.org/cldr/

[12] *The locale Generator.*
http://www.it46.se/localegen/locale/1179113539_bo_CN.xml

[13] *Dzongkha Linux Website.*
http://dzongkha.sourceforge.net/

[14] *China Tibetology.*
http://www.tibet.cn/english/zt/TibetologyMagazine/..%5CTibetologyMagazine/2003120031216162127.htm

[15] *Online version of the Rangjung Yeshe Tibetan-English Dictionary.*
http://www.nitartha.org/dictionary_search04.html

[16] *The Tibetan to English Dictionary and Translation Tool.*
http://www.gaugeus.com/ramblings/tibetan-to-english-translation-tool

[17] *Jim Valby Tibetan-English Dictionary for MS-DOS.*
http://snark.ptc.spbu.ru/~uwe/tibet/t.html

[18] *English to Tibetan Online Dictionary.*
http://eng-tib.zanwat.org/

[19] *UTF Diacritics for Tibetan Script (THDL).*
http://www.thdl.org/tools/fonts/diafonts.html

[20] *How to Use Microsoft Tibetan Keyboard in Windows Vista.*
http://www.yalasoo.com/English/docs/yalasoo_en_MStbKb.html

[21] *The Optimus Maximum Keyboard.*
http://www.artlebedev.com/everything/optimus/

[22] *Qt support for i18n.*
http://doc.trolltech.com/4.0/i18n.html

[23] *Pango: supporting i18n for Gtk applications.*
http://www.pango.org/

[24] *Gnome l10n status.*
http://www.gnome.org/i18n/

[25] *KDE l10n status.*
http://l10n.kde.org/stats/gui/stable-kde4/team/

[26] *OpenOffice.org l10n status.*
http://wiki.services.openoffice.org/wiki/Languages

[27] David Germano. *Getting Started with Unicode Tibetan.*
http://www.thdl.org/xml/showEssay.php?xml=/tools/fonts/getstartuni.xml&m=all

[28] *Tibetan5.*
http://www.tibet.dk/tcc/tibetan5.htm

[29] *PechaMaker.*
http://www.pechamaker.com/

[30] *Apple Advanced Typography.*
http://developer.apple.com/textfonts/

[31] *MicroSoft OpenType Font Technlogy.*
http://www.microsoft.com/typography/default.mspx

[32] *The otibet LaTeXpackage.*
http://www-hep2.fzu.cz/tex/texmf-dist/doc/latex/otibet/

[33] *The ctib LaTeXpackage.*
http://www.tug.org/texlive/devsrc/Master/texmf-dist/tex/latex/ctib/

[34] *The MetaFont font description language.*
http://en.wikipedia.org/wiki/METAFONT

[35] *ASpell supported languages.*
http://aspell.net/man-html/Supported.html

[36] *ISpell supported languages.*
ftp://ftp.gnu.org/gnu/aspell/dict/0index.html

[37] *Skype* http://www.skype.com

[38] *Pidgin l10n status.*
http://developer.pidgin.im/l10n/

[39] *Opera l10n status.*
http://www.opera.com/download/languagefiles/

[40] *Firefox l10n status.*
http://www.mozilla.com/en-US/firefox/all.html

[41] *Thunderbird l10n status.*
http://www.mozilla.com/en-US/thunderbird/all.html

[42] *The official Tibetan keyboard layout.*
http://en.wikipedia.org/wiki/Image:Tibetan_Keyboard.png