



Friedel Wolff

Software localisation by Translate.org.za

Abstract

Prior to 2001, very little software had been localised for South Africa. The numerous languages in South Africa have been cited as problematic. Translate.org.za started to localise software for all eleven official South African languages and has been the catalyst and driving force in the South African localisation community. Initial work included a South African keyboard layout usable for all South African languages. Translations in all eleven official languages were recently completed for OpenOffice.org, Mozilla Firefox and Mozilla Thunderbird. These translation projects also resulted in the development of several localisation tools that simplify the localisation of these big projects and a web-based translation and translation management tool. An active volunteer community has already developed for one language and proprietary vendors have gradually been introducing localised products.

Keywords

localisation, South Africa, localisation tools, Afrikaans, minority languages

South Africa has a long and complex history. It has seen some of the oldest hominids known to modern man, tribes migrating from the north, colonialism by the Netherlands and the United Kingdom, imported slaves, several wars, and the political turmoil of the 20th century.

Currently South Africa has eleven official languages. It is second only to India in terms of the number of official languages. However, the South African language diversity is quite small compared to other African countries. Linguistically, the ten non-English languages are categorised as follows: one West Germanic language (Afrikaans) and nine languages from the Bantu family. Of these nine, four belong to the Nguni language group, three to the Sotho language group, and two are separate languages (Tsonga and Venda). Several of these eleven languages are also spoken in neighbouring countries.

In some parts of the country English serves as a *lingua franca* – mainly in the cities and more so to the south-eastern part of the country. Afrikaans is more dominant in the western part of the country and in rural areas. The use of the other languages is, to varying extents, localised in certain areas.

Translate.org.za was started in 2001 by Dwayne Bailey to localise Free Software for South Africa. A few applications were available in Afrikaans but no software was available in the other local languages at that stage. Having eleven official languages were mostly of symbolic value when it came to computers and technology.

1. Input and Display

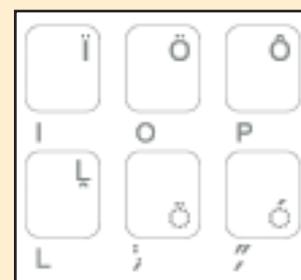
There are relatively few technical problems with input and display for the eleven official languages. All use the Latin character set, with four languages using diacritics. The diacritics for three of these exist in some European languages. However, five characters exist that are unique to Venda. All of the extended characters had already been codified in Unicode.

No keyboard for South African languages has ever been developed.

To this day, many resort to 'Alt-codes' in Windows to input the extended characters, or use application specific character insertion techniques, or simply do not use the correct characters any more. Neglecting the diacritics is truly problematic, as this greatly reduces the morphological wealth of the affected languages. In some cases people still manually insert diacritics after printing, thereby making it impossible to have perfectly correct electronic copies. The situation has probably been worsened by the fact that the non-English languages do not enjoy high status in business and the fact that the diacritics do not occur in all of the languages, and in some cases do not occur with great frequency either.

To rectify this, Translate.org.za developed a keyboard with which all languages of South Africa can be typed [1]. It could not be much different from the standard US layout, since that is what is ubiquitous in South Africa, and many people need to use English regularly. It also had to be taken into account that many people would not use the extended characters often, or might not be adept typists.

Figure 1: A sample of the keys for the right hand on the South African keyboard layouts. The dotted circles indicate dead keys for using the relevant diacritics.



Along with the development of the South African keyboard, the popular DejaVu fonts were extended with the Venda characters that were still missing.

2. Initial Steps

As first projects, several smaller programmes were translated into the major languages of South Africa. Google South Africa was made available in four languages (unfortunately it is still not possible to limit searches to a specific South African language). The popular desktop environments, KDE and GNOME, were

translated and valuable lessons were learnt: skilled translators were not always easy to come by and managing translation efforts for several languages proved to be a mammoth task. Many translators preferred to work with spreadsheets rather than with applications supporting the standard Portable Object (PO) files that are used to localise Free Software. Experience has shown that translators need more training than might be expected.

Work started on the development of some tools as part of the Translate Toolkit to do format conversion and to test the quality of translations automatically. Some of the automated checks included tests for consistent punctuation, spacing, variable use and XML tags. These tools proved to be invaluable for the review process, as many messages could be filtered out for review by a non-native speaker of the language.

3. Locales

Locales were added to the GNU C library and to OpenOffice.org to support all official languages. All locales are now available in the Common Locale Data Repository. Microsoft has supported Afrikaans and South African English for some time and they added three more South African locales in Windows XP SP2. Locales for most of the other languages have been assigned, but will only be part of future releases of Windows. Locales for two of the languages are still lacking [2] [3].

4. Big Successes

In 2004 Translate.org.za released the first complete, localised Office productivity suite localised for South Africa by releasing OpenOffice.org 1.1 in four official languages [4]. As part of a large sponsorship Translate.org.za was able to extend the effort to update translations for OpenOffice.org 2.0 and include translations for all official languages [5]. The popular web browser and e-mail client from the Mozilla Foundation, Firefox and Thunderbird, were also translated into all official languages.

An important part of a fully localised office suite is a spell checker. Infrastructure was developed for the development of several spell checking systems and existing word lists were used to provide initial spell checkers. The complex morphology of the languages, especially for the languages in the Nguni group, offers severe challenges, especially for traditional UNIX spell checkers. The languages in the Nguni group are agglutinating languages, meaning that up to a whole sentence can be represented as a word. It is hoped that a future project would make it possible to extend Hunspell, the new checker used by OpenOffice.org, to provide support for the rich morphology of all official languages. Another difficulty with spell checkers is that they require good word lists. Often the best lists are to be obtained from lexicographers in the various languages. However, since the spell checkers are released as Free Software, it is very difficult to convince lexicography units – even though they are government supported – that this would not hamper their other commercial efforts in printed dictionaries.

5. Community Building

Part of the effort of Translate.org.za was not only to provide localised software, but also to ignite the flame of community localisation projects and to build a culture of multilingualism and of using localised software. Because English is also an official language, and English proficiency is relatively high amongst

economically and technologically privileged, resistance to change has slowed the uptake of localised software. For some languages the community mailing lists are mostly dormant. The lack of translated teaching material has also been cited as an inhibitor for the adoption of localised software in training programmes.

To spark interest in software localisation, some *Translate@thons* (localisation sprints) were held with focus on specific languages. These events try to attract people to translate some software in a single day. The web-based translation tool, Pootle, created by Translate.org.za, has proven invaluable for these events. Such events can attract a mixed crowd in terms of translation skill, technical skill, and true interest. While a small group usually creates quality localisations, large numbers make it very hard to achieve quality translations and probably serve better for creating awareness and interest. Most people still consider computers and electronics to be something inherently English.

However, interest in localised software is slowly on the increase. The major accomplishments are reported on in local internet news sites, and some interviews were held on national and community radio stations [6] [7] [8].

6. The Afrikaans Localisation Community

Afrikaans arguably sports the most successful localisation community among the endemic African languages. It has active mailing lists, coordinated terminology efforts and many localisation projects undertaken by community volunteers.

It was possible to compile a reasonably good spell checker from previously compiled word lists and this was improved by community members. More recently this work was also extended with hyphenation rules and data for the AutoCorrect feature of OpenOffice.org.

The success of the Afrikaans community has indicated that effort is often required by an individual to take initiative, or to coordinate willing helpers. Without leadership many projects are unlikely to be completed or to achieve good quality. Without an existing volunteer community, newcomers find it hard to become involved. An existing community provides means for newcomers to join, without them needing to provide the initial leadership.

7. The Future

Although significant milestones have been reached, much remains to be done. The work on good spell checking for all languages was mentioned as an outstanding project; the complex morphology also affects development of an effective AutoCorrect functionality. Ideally grammar checking and thesauri should become feasible in the future. Collation specific for some of the languages should be considered, although the decision is a complex one, as powerful morphological analysis might be needed and could render such collating impractical.

Perhaps the most noticeable effect of the localisation work is to see how others have joined in. A few cellular phones are available in a few local languages, recently even with predictive text input for Afrikaans. A Zulu language interface pack for Windows XP was released in April 2006 and others would have followed soon thereafter [9]. Some of the Microsoft website is now also partially available in some South African languages.

Translate.org.za continues its work in the development of tools to simplify Open Source Software localisation; currently as part of the WordForge project. It is believed that this should afford even the smallest of marginalised languages a chance to efficiently manage their localisation projects at little or no monetary cost.

8. Conclusion

Despite common opinion to the contrary, we have proven that it is possible to localise software into all eleven official languages of South Africa. Only localising content for one of each of the language families is a common practice (a total of six languages), but localising for all eleven languages truly puts them all on equal footing. We have also aimed to translate complete user interfaces, rather than only translating the most commonly used messages.

We have shown that Free and Open Source Software can act as a driver for localisation. Almost nobody had anything on the cards a few years ago, but since Translate.org.za has delivered, proprietary vendors have at least started doing lip service and stopped denying that there is a demand for localised products.

On the downside, we have found that dominance of English in the economic sphere makes it hard for localised software to be adopted, as those who would use it, do not want it to hamper their career prospects. Furthermore, the dominance of proprietary software suppresses localisation. The use of proprietary software in government, education, the work place, etc., means that people don't necessarily even have the choice to use software in their mother tongue, even now that it exists.

We expect Free and Open Source Software to continue to dominate in localisation. As proprietary vendors follow this lead and create even more awareness, we expect more people to be drawn to community localisation efforts, where much greater depth and breadth is possible. Harnessing the power of communities empowers both the languages and the communities, and builds a culture of multilingualism.

Friedel Wolff obtained his Masters degree at the University of Johannesburg and currently works for Translate.org.za as part of the WordForge project. He is also involved in the Afrikaans localisation community with spell checkers and grammar checkers as specific fields of interest. He may be reached at friedel@translate.org.za.

References

- [1] [Translate.org.za \(2005\) South African keyboard](http://translate.org.za/content/view/1526/51/) [online], available: <http://translate.org.za/content/view/1526/51/> [accessed 11 July 2006]
- [2] [Microsoft Corporation \(2006\) List of Locale ID \(LCID\) Values as Assigned by Microsoft](http://www.microsoft.com/globaldev/reference/lcid-all.msp) [online], available: <http://www.microsoft.com/globaldev/reference/lcid-all.msp> [accessed 11 July 2006]
- [3] [Microsoft Corporation \(2006\) Windows XP/Server 2003 - List of Locale IDs, Input Locale, and Language Collection](http://www.microsoft.com/globaldev/reference/winxp/xp-lcid.msp) [online], available: <http://www.microsoft.com/globaldev/reference/winxp/xp-lcid.msp> [accessed 11 July 2006]
- [4] [Translate.org.za \(2004\) OpenOffice 1.1.2 released in Afrikaans, Zulu and Northern Sotho](http://translate.org.za/content/view/1545/51/) [online], available: <http://translate.org.za/content/view/1545/51/> [accessed 11 July 2006].
- [5] [Translate.org.za \(2005\) OpenOffice.org Launch Celebration](http://translate.org.za/content/view/1499/51/) [online], available: <http://translate.org.za/content/view/1499/51/> [accessed 11 July 2006]
- [6] [Translate.org.za \(2004\) Translate on the airwaves](http://translate.org.za/content/view/1506/51/) [online], available: <http://translate.org.za/content/view/1506/51/> [accessed 11 July 2006]
- [7] [Translate.org.za \(2005\) SAfm with John Perlman](http://translate.org.za/content/view/1496/51/) [online], available: <http://translate.org.za/content/view/1496/51/> [accessed 11 July 2006]
- [8] [Translate.org.za \(2004\) OpenOffice.org on Radio Pretoria](http://translate.org.za/content/view/1563/51/) [online], available: <http://translate.org.za/content/view/1563/51/> [accessed 11 July 2006]
- [9] [International Marketing Council of South Africa \(2006\) Microsoft Windows in Zulu](http://www.southafrica.info/public_services/citizens/your_rights/microsoft200406.htm) [online], available: http://www.southafrica.info/public_services/citizens/your_rights/microsoft200406.htm [accessed 11 July 2006]

<p>Production solutions for multilingual work items in separate pieces or in various combinations plus important extras saving significant amount of time and effort</p>	<h2 style="text-align: center;">Logrus Technology:</h2> <ul style="list-style-type: none"> • Multilingual localization, engineering and testing • Rare languages and complex scripts • Large production site in Russia • Complex projects, large and small • Authoring, test case writing, tools development • Multilingual DTP in 36+ languages, including ME and FE • Most field area specializations: ERP, mechanical, engineering, telecom, marcom, finance, medical, consumer and more
 <p>www.logrus.ru management@logrus.ru +1 (215) 947-4773</p>	<p>We provide multilingual services into a large number of languages, including complex scripts and rare tongues. We specialize in large and highly technical projects requiring unique technical experience, a high level of self-sufficiency, and outstanding problem-solving capabilities. We go beyond simple standard separate work pieces and complete solutions of several of them combined, saving your time with very nice extras which include out-of-the-box workarounds, writing tools to cut hundreds of hours of effort, managing in-country approval, doing design, layout, DTP, sound recording and so much more.</p> <p style="text-align: right;"><i>Standing above a background of expectations</i></p>