

20

The Localisation into isiXhosa of the iLanga Telephone System

*Lorenzo Dalvit, Mosiuoa Tsietsi, Alfredo Terzoli,
Pamela Maseko Msindisi Sam and Thandeka Mapi*

In this paper we describe the localisation of iLanga, a VoIP PBX system, into isiXhosa. In South Africa, for social and political reasons, making systems available in the African languages is becoming more and more important. On the one hand, it gives access to the use of technology to people who are not fluent in English. On the other hand, it has a strong symbolic value. Although a variety of software is already available in a number of African languages, localisation in such language is still in its early stage. The process presented interesting challenges both from the technical and the linguistic point of view.

1. Introduction

In South Africa, like in other parts of Africa, English is the dominant language of Science and Technology. In most cases, proficiency in English is therefore required in order to access Information and Communication Technology (ICT). Various sources [Department of Education and Department of Communication 2001] acknowledge the potential role of English as a “gatekeeper” to the use of ICT by members of disadvantaged communities and emphasise the need to promote technological discourse in the African languages.

We start our paper with an overview of the presence of African languages in ICT and on localisations in the African languages. We then focus on the localisation process itself and the specific challenges to localisation in the African languages. After that, we describe our experience in localising iLanga, a VoIP PBX system developed in the Department of Computer Science at Rhodes University. We end with some feedback and some reflections on the possible scenarios for the implementation of the localised system.

2. Localisation in the African Languages

2.1. African Languages And ICT

The *localisation* (i.e. translation of software interfaces) in the South African indigenous languages is a comparatively recent but fast growing phenomenon. Microsoft Windows XP has been released in Setswana, isiZulu and Afrikaans [Ashword 2006]. Up-coming Microsoft Vista is expected to include more African

languages. The popular search engine Google is also available in a number of South African languages.

Wikipedia [Wikipedia 2007] features a number of articles in various South African indigenous languages. A growing number of resources in the African languages are available on the Web [De Schryver 2003]. These are mainly on-line dictionaries, often maintained by speakers of other languages.

Translate.org.za [Translate 20007] is a South African NGO committed to the localisation of open-source software into all South African official languages. Software localised so far includes OpenOffice, Mozilla Firefox, Gnome and KDE as well as a number of other applications such as. Tuxpaint, for instance [see <http://downloads.translate.org.za>]. This makes it possible to run a computer almost entirely in a language such as isiXhosa, for instance.

Translate.org.za developed an alternative methodology for the localisation of open-source software called `translate@thon`. This is basically a translation marathon, during which volunteer translators work collaboratively on-line using PO-based On-line Translation and Localisation Engine (Pootle). This is an application developed by Translate.org.za themselves, which allows translators to work on the same project collaboratively and provides suggestions from a database to ensure consistency. The main Pootle server is maintained by the Telkom Centre of Excellence of Rhodes University and is available at <http:pootle.translate.org.za>.

2.2. The localisation process

The two main components of localisation are translation and modification [Keniston 1999]. Translation is the linguistic component of localisation. Ideally, it consists of five phases: translation into the target language, back-translation into the original language, comparison between the two versions, adjustments to the version in the target language and integration into the interface [4]. The localised items are usually stored and indexed into a separate file.

Modification refers to more “structural” changes, such as scrolling patterns (e.g. for Arabic), character set (e.g. for Russian or Chinese) and box sizes. It can involve cultural aspects (e.g. colours or icons) and linguistic aspects (e.g. dictionary search patterns).

2.3. Challenges

Localisation can be a lengthy and expensive process. The localisation of an operating system, which can involve up to half a million words, can cost up to one US Dollar per word and take one to two years [Keniston 1999].

Another problem is finding appropriate terms which are at the same time effective and compliant to a common standard. This is particularly true for African languages, which often have very different dialects and varieties. Moreover, since localisation in such languages is a comparatively new phenomenon, various interventions are usually piecemeal and lack coordination.

Related to this, there is the issue of lack of appropriate terminology [Heugh 2002]. This affects the use of African languages in Science and Technology in general. A considerable effort is put by the South African Government into developing new terminology and promoting its use [National Language Services 2007]. In spite of this, many speakers of an African language prefer to use English borrowings [Heugh 2002], such as *ipassword* (password) or *imeyile* (e-mail) in isiXhosa.

3. Background to the Project

3.1. A Multi-disciplinary Collaboration

In recent years, a common interest for localisation issues has brought together Translate.org.za, the Telkom Centres of Excellence (CoE) of Rhodes and Fort Hare University and the South Africa-Norwegian Tertiary Education Development (SANTED) multilingualism programme at Rhodes.

The SANTED programme is specifically tasked with the implementation of the Rhodes Language Policy. Rhodes is an historically English-medium institution, but is situated in a predominantly isiXhosa-speaking area. Its new language policy involves mainly the promotion of isiXhosa in all domains. Although the CoE of both Rhodes and Fort Hare focus mainly on telecommunication, their leaderships have long realised the importance of language issue in promoting access to ICT. For this reason both Centres have sponsored research on localisation in the past.

3.2. An Asterisk-based Telephone System

The telecommunication landscape has been evolving at a fast pace, spurred on by the exciting possibility of delivering a range of new multimedia services to consumers largely through the adoption and development of open source telephony applications, among which Asterisk is chief [Meggelen et al 2005]. Asterisk is a softswitch that is able capable of handling voice and data communication between heterogeneous endpoints in a seamless fashion. It provides a framework for producing a variety of useful telephony services such as IVR (Interactive Voice Response), voicemail storage and retrieval and multi-party conferencing.

Since 2004, postgraduate students sponsored by the CoE at Rhodes University have been developing an integrated telephony solution called iLanga which is based on Asterisk [Penton and Terzoli 2004]. The initial installation was complimented in subsequent years with an aesthetically pleasing, dynamic interface built with Flash, Action Script and PHP [Hitchcock 2006]. Its main objective was to provide an alternative method through which users could interact with the system. Many of the functions that are provided on the user telephones, such as accessing voicemail messages and initiating calls, are also available via the interface.

iLanga is a complete, cost effective, telephony solution based exclusively on open source telephony and non-telephony software, which embeds asterisk at the core. Since 2005, it has been connected to the University's PBX and tested with postgraduate students and staff in the Department of Computer Science. During

2006, a similar system was deployed in Dwesa, a rural community in the Eastern Cape, South Africa, as part of an ICT-for-Development project discussed at a previous edition of the conference [Dalvit et al 2007]. The system is now available to a wider and linguistically (as well as culturally) diverse community of users. This prompted its localisation into isiXhosa, the most widely spoken African language in the area.

4. Localisation of iLanga

4.1. A Bilingual Interface

iLanga currently allows a user to choose which language they want to use in their interaction with the system, either English or isiXhosa [Dalvit et al 2006]. The system interface has two components: a Web-based interface and a telephone-based audio interface.

The Web-based interface comprises a total of sixty items, divided into six screen shots. The initial translation was performed by approximately a dozen isiXhosa-speaking students of Computer Science using the translate@thon method. The initial phase took approximately six hours. The translations were later reviewed by a language expert before being integrated into the system (see Figure 1).

Fig.1: iLanga Main Page



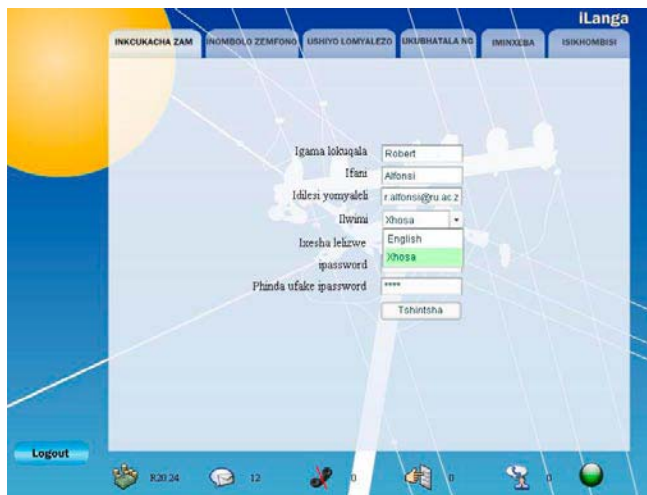
To interact with users via phone, Asterisk uses a set of pre-recorded voice prompts, stored in .gsm format. A complete language pack consists of 436 prompts, equivalent to approximately 2500 words. Obviously, some of these are more frequently used than others and some others refer to features (e.g. video conference) which are not yet implemented in iLanga. For this reasons we focused primarily on frequently used services, such as voicemail for instance.

The majority of the prompts were translated by a team of postgraduate students from various disciplines. These ranged from Computer Science to African Languages and Communication. On a smaller scale, the translate@thon method was again used.

Prompts were recorded and converted into .gsm format using standard open-source tools [Audacity 2007][Sox 2007]. The translated prompts were included in a separate directory identified by the relevant language code (i.e. “xh” for isiXhosa). The system automatically used prompts from the right directory based on the value of the “language” variable set through the user interface.

This configuration option is shown in Figure 2. The same does not apply to the audio portion of the service, which replays prompts completely in English. Fortunately, asterisk does provide configuration options for multi-language support so that users can listen to audio prompts in their language of preference (see Figure 2).

Fig 2: Language Configuration in iLanga



4.2. Challenges

To maintain the cohesion between the iLanga interface and the audio system, it is necessary to ensure that any language preference changes made on the interface are implemented in the audio system. iLanga stores the preference in a MySQL database, thus the Asterisk configuration on iLanga must be extended to interact with the database to serve the appropriate audio content on the user’s telephones.

Localisation of Asterisk systems is available for most European languages, such as French and Spanish. African languages have been marginalised in this respect, and currently there is no isiXhosa language package available for users of Asterisk systems to download and install. The compilation of an isiXhosa package, containing the voice prompts and working configuration settings, is one of the envisaged spin-offs of our project.

The main technical challenge was to get the system to respect the language-specific order of the prompt for isiXhosa instead of the default one for English. For example, in isiXhosa the adjective follows the noun it refers to, while in English the adjective precede the noun. In the initial stages, we relied on the example configurations for other languages (i.e. Italian) which follow a similar word order. A completely independent configuration, specific to isiXhosa, is currently being developed.

5. Summary

Localisation of software in African languages is important for both practical and symbolic reasons. This presents specific challenges, such as lack of resources and terminology for these languages. During the localisation of the iLanga VoIP PBX, a collaborative approach called translateathon was used for the translation component. The modification component presented a few technical challenges, mainly related to the structure of the languages involved and the differences compared to English. While the Web interface of the system is complete, the localisation of the voice prompts is limited mainly to the voicemail system, but suitable for demonstration purposes. By the time of the conference, we are confident we will be able to run a complete demonstration of the system. On a conclusive note it should be emphasised that, to our knowledge, this is the first VoIP PBX system ever to be localised in an African language. We hope it will constitute a valuable experience to inform future similar efforts.

Acknowledgements

We would like to thank the Telkom Centre of Excellence in Distributed Multimedia and the South African-Norwegian Tertiary Education Development Programme of Rhodes University for sponsoring this work.

References

- ASHWORD, W. (2006 April). Windows goes African. *ITWEB*. Available: <http://www.itweb.co.za/sections/software/2006/0604211033.asp?S=MICROSOFT%20WATCH&A=MSW&O=FRGN>
- DALVIT, L., ALFONSI, R., MINI, B., MURRAY, S., & TERZOLI, A. (2006). "The localisation of iLanga, a next generation PBX". Poster presented at the *Southern African Telecommunication Networks and Applications Conference (SATNAC)*, 3 - 6 September 2006, Cape Town.
- DALVIT, L., MUYINGI, H., TERZOLI, A., & THINYANE, M. (2007). "The deployment of an e-commerce platform and related projects in a rural area in South Africa". *International Journal of Computing and ICT Research*, 1(1): 7-19. Available: <http://www.ijcir.org/volume1-number1/index.php> (last accessed 30 November 2007).
- Department of Education and Department of Communication. (2001). *Strategy for Information and Communication Technology in Education*. Pretoria: Government Printer.
- HITCHCOCK, J. (2006). *Decorating Asterisk: Experiments in Service Creation for a Multi-Protocol Telephony Environment Using Open Source Tools*. Master's Thesis, Rhodes University, South Africa
- Gtranslator project details. Available: <http://freshmeat.net/projects/gtranslator/>

HEUGH, K. "The Case Against Bilingual and Multilingual Education in South Africa: Laying Bare the Myths". *Perspectives in Education*, vol. 20, no. 1, pp. 1-19.

KENISTON, K. "Language, power and software" in *Cultural Attitudes towards Technology*. C. Ess, Ed. NY: Suny Press. 1999.

MEGGELEN, J.V, SMITH, J. AND MADSEN, L (2005) Asterisk, the Future of Telephony. National Language Services. Available: http://www.dac.gov.za/about_us/cd_nat_language/home.htm

PENTON, J., AND TERZOLI, A (2004). iLanga: A Next Generation VOIP-based, TDM-enabled PBX. In *SATNAC Conference Proceedings (2004)*

Pootle <http://pootle.translate.org.za>

Translate project. Available: <http://www.translate.org.za>