

# Open-source software and localisation in indigenous South African languages with Pootle

Lorenzo Dalvit, Alfredo Terzoli and Friedel Wolff

## I. INTRODUCTION

**Abstract**—The majority of the South African population is not fully proficient in English. Besides access to the necessary infrastructure, many Africans lack epistemological access to the ICT world simply because of linguistic reasons. Limited availability of resources and lack of interest for the promotion of African languages in the ICT domain among their speakers are often blamed for this. The open-source community has been particularly responsive to the needs of this portion of the population, and a variety of software is already available in African languages. Translate.org.za, an NGO committed to the development of open-source software in all 11 South African languages, has been particularly active with respect to this. On the one hand, they have developed a Web-based application to support collaborative translation on-line. On the other, they have partnered with various institutions to organise translation efforts, mainly relying on volunteers. In this paper we describe an experience showing how these two models can be used to address issues of lack of resources and support for the use of African languages in the ICT domain.

**Index Terms**—Open-source, localisation, translation, African languages.

APPROXIMATELY 80% of the South African population is *not* fully proficient in English [1]. This means that, besides having limited access to ICT infrastructure, most people in townships and rural areas cannot use computers simply because they function in English. Language issues are a possible factor determining the failure of many ICT-for-development projects [2]. Besides the practical problems of operating a computer in a language one hardly knows, lack of English proficiency adds a psychological component to the issue: technology which is already “foreign” in many other ways, also uses what for most Africans is still a foreign language.

Proposing the use of software in the African languages might sound like a daring alternative. Some scholars [1] argue that, under apartheid, such languages were underfunded and marginalised. As a result, even their own speakers would appear to support their use only in traditional and low-status domains, possibly because of a sense of linguistic “inferiority complex”. This could result in a vicious circle, perpetuating the exclusion of African languages and their speakers from the ICT domain. In spite of this, African languages are experiencing a small-scale Renaissance in the ICT World, as more and more software becomes available in these languages [3].

The process of adapting software to local contexts is called *localisation* [4]. This includes a number of different processes ranging from adapting interfaces and colour schemes to different cultural norms, to creating language-specific fonts. Translation of user interfaces into local languages is an integral part of localisation, and it is the most important one for the purposes of this paper. In particular, we consider localisation into the official indigenous South African languages. Although the discussion pertains mainly to languages of African origin, Afrikaans is also included.

Although proprietary software is slowly coming to the fore, the primary drive for localisation in the African languages seems to be the open-source community [5]. Because of its ideological underpinnings, the latter seems to be more responsive to the needs of speakers of less powerful languages. In this paper we will highlight this issue. We will then describe an application developed specifically to support localisation. Finally, we will contextualise its usage within a real-life experience, which seems to address two of the issues mentioned above, i.e. lack of resources and lack of

---

<sup>1</sup>L. Dalvit is Research and ICT coordinator of the SANTED multilingualism programme within the School of Languages (African Studies Section) and Lecturer in ICT Education in the Education Department at Rhodes University, PO Box 94, 6140 Grahamstown, South Africa. (Tel: +27 (0)46 603 7267; e-mail: [l.dalvit@ru.ac.za](mailto:l.dalvit@ru.ac.za)).

A. Terzoli is Head of the Telkom Centre of Excellence in Distributed Multimedia and Professor in the Computer Science Department at Rhodes University, PO Box 94, 6140 Grahamstown, South Africa. (Tel: +27 (0)46 603 7267; e-mail: [a.terzoli@ru.ac.za](mailto:a.terzoli@ru.ac.za)).

F. Wolff is Senior Developer of Translate.org.za, 63A Wenning street, Groenkloof, 0181 Pretoria, South Africa. (Tel: +27 (0)12 460 1095; e-mail: [friedel@translate.org.za](mailto:friedel@translate.org.za)).s

support for the promotion of African languages among their own speakers.

## II. SOFTWARE AVAILABLE IN INDIGENOUS SOUTH AFRICAN LANGUAGES

### A. Proprietary software

While a growing amount of content in African languages is becoming available on the Web [3], comparatively few of the most popular Web-based applications have been localised to cater for speakers of such languages. Piecemeal efforts made by individual companies [5], though interesting in their own right, do not have the potential to impact on the bulk of the African population.

The South African version of the search engine Google [6] is currently available in Afrikaans, Sotho, Zulu and Xhosa. Informal communication with native speakers of Xhosa, however, suggests that more work is needed to ensure relevance and quality of the terminology used. Words like *uphendlo* (“search”) are archaic and their use is not very common among African youth, arguably the primary target of this intervention. In our opinions, efforts such as these seem to pay “lip service” to what is considered the dominant orientation at the time, rather than increasing access for members of marginalised communities.

Microsoft is working in close collaboration with various government bodies, such as the Pan South African Language Board (PanSALB) [7]. In recent years, some of its products have become available in African languages. Windows XP has been localised in Afrikaans, Zulu and Tswana and upcoming Vista will include various other South African official languages [8]. Since Microsoft dominates the end-user market, these products could potentially reach millions of Africans.

Microsoft's approach to localisation, however, is inevitably shaped by its developer-centric orientation [9]. Collaboration with government structures introduces a number of bureaucratic hurdles which inevitably delay the process. Moreover, a “top-down” approach might not work for the promotion of new terminology in the African languages in a field, such as ICT, where this is likely to meet some resistance from the users. Consistently with the recommendations of the South African language policy [10], localisation into African languages might require an innovative, “bottom-up” approach.

### B. Open-source software

Open-source software caters for a comparatively small section of the end-user population. It is characterised by a distributed and user-centred orientation and is informed by ideals of freedom and collaboration [11]. Possibly because of this, it has been comparatively more responsive to the needs of speakers of less powerful languages around the

world [12].

In South Africa, consistently with the Government's orientation, various organisations such as Go-Open Source, for instance, have done considerable work to promote the use of open-source software [13]. This is free to use, modify and distribute and lends itself to adaptation to different contexts. The popular and fast-growing Ubuntu distribution of Linux is a project by Canonical Ltd, initiated by South-African born entrepreneur Mark Shuttleworth. *Ubuntu* is the Zulu and Xhosa word for “humanity” and it encompasses much of the philosophy underlying the open-source movement [14]. Canonical coordinated the localisation of the Gnome desktop environment for Ubuntu in various languages, among which Xhosa.

Translate.org.za is an NGO committed to the localisation of open-source software in all 11 official South African languages. Since 2001, it has been very active in making popular open-source software available in a number of African languages. The OpenOffice.org2 office suite and the Mozilla Firefox Web browser and ThunderBird mail client are available in all 11 languages. Translate.org.za also localised into one or more of the indigenous South African languages a wide range of other applications. These include desktop environments (e.g. Gnome and KDE), ICT-related content (e.g. Creative Common licenses), videogames (e.g. KhangMan) and Web applications (e.g. Pootle, Horde/Imp).

Among the latter, Pootle is a project initiated and managed by Translate.org.za itself to support collaborative localisation on-line. Horde/Imp is a webmail system and has been localised as part of a collaboration between Translate.org.za, Rhodes University and the University of Fort Hare, as described in the last section.

## III. THE APPLICATION

### A. Pootle and the Translate Toolkit

Pootle stands for PO-based, On-line Translation and localisation Engine. It is a rigorously open-source localisation management system developed to enable translators to work collaboratively on-line. Together with the Translate Toolkit, Pootle can be used as a server-based Web application over the Internet or within a local Intranet by a team of localisers, or on a standalone machine.

Although it was developed with open-source software in mind, it can be used to localise all types of products. It uses standard formats such as PO (Gettext Portable Object) and XLIFF (XML localisation Interchange File Format) and adheres to international standards recognised by LISA (localisation Industry Standards Association) [15]. Being file-based, Pootle can work with existing files and does not require a database. Files are indexed using Lucene [16] to speed up searching, which is done with Unicode normalisation. localised files can be submitted directly to

version control systems such as CVS, Subversion etc.

Integration with the Translate Toolkit allows for upload and download of various file types to enable translation off-line. It also makes it possible to run a number of checks to correct errors in localisation. These make sure that the treatment of elements such as variables and escapes is consistent with the conventions of each project. The Translate Toolkit provides statistics such as string and word counts and progress in percentage for each project. Work within a project can be shared and assigned to different translators, possibly with different functions ranging from viewer to administrator.

The Web-based editor (Figure 1) filters untranslated and fuzzy strings supports and can provide localisers with untranslated work to speed up the process. It allows comments from translators and gives suggestions based on existing projects, glossaries or a translation memory. Besides allowing for right-to-left display and customised styling for languages which require it, the Pootle editor can include a set of language-specific clickable characters. This is important for languages such as Afrikaans, Pedi and Venda, which use characters which are not commonly found on standard keyboards.

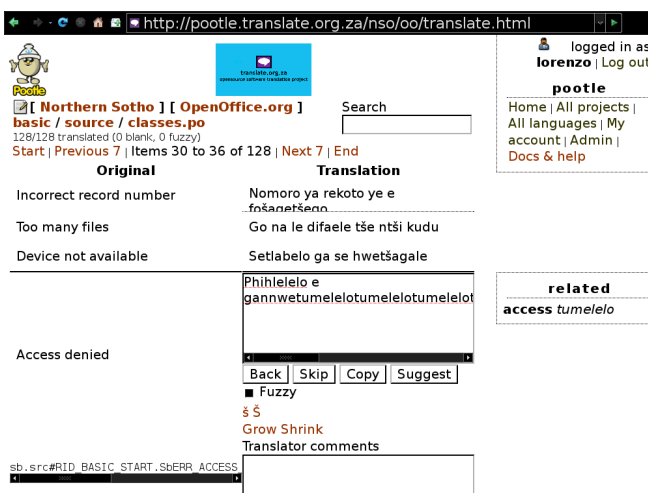


Figure 1: Pootle Editor

§

### B. Background

The project was initiated by Translate.org.za in 2001 as part of WordForge in response to the need for localisers around the world, particularly for less powerful languages. Pootle is currently used by a number of localisation teams [17].

One could say that the applications developed by Translate.org.za are designed to support crowdsourcing in the localisation context. Crowdsourcing is not specifically an ICT term, and can be used to refer to the outsourcing of a task (e.g. designing something, solving an algorithm or process data) to a large and undefined number of people [18]. Its reliance on collaborative, distributed and (in many

ceases) voluntary participation determined its up-take by the open-source community.

Like many other open-source projects, Pootle and the Translate Toolkit have been developed by a number of programmers in different parts of the World. This community communicates through an IRC channel and by means of two mailing lists administered by Translate.org.za. These forums are channels to exchange information and discuss different opinions. A recent example was the issue of supporting XLIFF files as well as the PO files Pootle was originally designed for.

### C. Translate@thon

A community of volunteers uses Pootle to work collaboratively on various localisation projects for indigenous South African languages. These individuals are experts in various fields, ranging from Computer Science to Linguistics, and live in different parts of the country. Reliance on volunteers is not uncommon in the open-source World, but Translate.org.za has uses an original model to encourage participation and maximise efficiency.

While on-line collaboration is an on-going process, major localisation “leaps” have been archived through a series of translate@thons. A translate@thon can be defined as a translation sprint, during which a number of translators gather to accomplish one specific task. Consistently with the crowdsourcing model, the underlying assumption is that breaking a task down in a number of smaller tasks speeds up the process. Besides that, translate@thons usually take place in a cheerful atmosphere, where participants are galvanised and encouraged to take action in supporting their own language beyond the event itself. In this sense, they can be seen as part of an awareness campaign, and normally attract considerable attention from the media [19].

A translate@thon can have any number of participants, and the size of the event determined to some extent the outcomes one can expect. Large translate@thons (40 people and more) often attract a number of unskilled volunteers whose primary motivation is the social nature of the event. These require considerable efforts in explaining and demonstrating how localisation works, and often lead to poor quality of translation. However, a large number of participants is likely to attract attention from the media and can contribute to raising awareness. Small translate@thons (approximately 5 people) usually involve a selected team of dedicated and skilled translators, coupled with a relatively high proportion of technically-skilled people. These events are usually more effective in terms of quality rather than amount of work done.

A number of translate@thons have been organised, usually in collaboration with the languages departments of academic institutions. The localisation of Mozilla FireFox into Xhosa was hosted by the University of Cape Town and

the one into Zulu by the Durban Institute of Technology. The model has also been used for the localisation of various pieces of software (e.g. TuxPaint) into a number of languages (e.g. Yoruba, Hausa, Amharic etc.) at a translate@thon in Ghana, in collaboration with AITI [20]. Academic institutions can contribute with expertise, infrastructure and networks of contacts. These events can be the occasion to find dedicated people who are willing to get involved, possibly integrating localisation into their academic career [21].

After a translate@thon, extensive review is needed to ensure accuracy and consistency of translation. Depending on availability of funds, Translate.org.za can support this by hiring professionals or through its network of expert volunteers. This community communicates via language-specific mailing lists and can contribute over the Internet using the official Translate.org.za Pootle server [22].

#### IV. IMPLEMENTATION

##### *A. The Translate.org.za Pootle server*

Translate.org.za works in collaboration with a number of other organisations and institutions and different components (e.g. its Web and mail server, its Wiki etc.) are hosted in different locations. Since 2006, the official Translate.org.za Pootle server [22] and the file server from which localised software can be downloaded [23] are hosted by the Telkom Centre of Excellence in Distributed Multimedia at Rhodes University. Although the primary interest of the Centre is telecommunication, its leadership has long realised the importance of tackling language issues to promote access.

Pootle.translate.org.za is hosted on an Ubuntu Linux server, upgraded from version 6.04 (Dapper Drake) to the current 7.04 (Feisty Fawn). In older versions of Pootle, this created some problems, especially because of its integration with Kid and Elementtree [24], which Pootle uses for its HTML templates, and Lucene [25], which is used for searches. In the versions currently installed (1.0.2) these issues have been resolved by the integration of the Pootle package in the official Ubuntu repositories. The new Pootle release (1.1.0) will therefore be available on upgrade to Ubuntu 8.04 (Hardy Heron). Although this entails a delay of a few months compared to the actual Pootle release cycle, it allows maintaining a coherent and robust system

The server has more than 200 registered users and hosts some 20 projects ranging from desktop environments such as Gnome and KDE to videogames such as KhangMan. To different extents, all 11 South African languages (i.e. Afrikaans, South African English, Ndebele, Xhosa, Zulu, Sotho, Pedi, Tswana, Swati, Venda and Tsonga) are represented. Portuguese has been added to support a cooperation with Mozambique.

In some cases, the central server is used for translate@thons. More often, though, local installations within an institution's Intranet are used. This allows translators to use a broadband connection, thus improving their experience in using the server. A more distributed and de-centralised model is also consistent with the principles of crowdsourcing described above. After a [translate@thon](#), projects can be moved to the central server so that dedicated and expert translators can complete them if necessary and review them.

##### *B. The first translate@thon at Rhodes*

In May 2007 Rhodes University, a historically White tertiary institution in the Eastern Cape Province of South Africa, hosted its first translate@thon. This was an initiative of the SANTED (South Africa-Norway Tertiary Education Development) programme at Rhodes University, in collaboration with the Telkom Centre of Excellence and Translate.org.za. SANTED is the fruit of a collaboration between the Norwegian Government and the South African Department of Education. At Rhodes University, it funds a programme, hosted by the African Studies Section of the School of Languages, to promote multilingualism and the use of Xhosa within the institution. Through its extensive collaboration with the Centre of Excellence, localisation has become an integral component of the programme.

The objective of the translate@thon was to localise Horde (a PHP-based Web application framework) and Imp (a Web-based e-mail client). These are the two main components of an open-source e-mail system used by many institutions, among which Rhodes. Since Rhodes is based in a predominantly Xhosa-speaking area, making the e-mail system of the University available in this language could have a profound impact on the university community. In particular, we hoped that it would instil a sense of acceptance and belonging among Xhosa-speaking students. We also hoped a translate@thon would help raise awareness around language issues at the University.

The translate@thon took place in the Peter Mtuzi Multimedia Laboratory, the computer facility of the School of Languages. All 30 machines in the laboratory are dual boot (Windows and Linux) and feature state-of-the-art in terms of open-source software localised in South African languages. To our knowledge, this is the first computer facility at a South African tertiary institution to allow students to operate a computer almost entirely in a language such as Xhosa, for instance [26].

During the weekend of 11 and 12 May, a group of approximately 50 volunteer students engaged in the localisation into Xhosa of Horde/Imp (Figure 2). Students came from various departments ranging from Computer Science to Communication and Media Studies at Both Rhodes and Fort Hare. Cooperation between a historically

“White” and a historically “Black” institution allowed us to combine technical expertise with sensitivity for the language in context. Students from various disciplines were encouraged to work sit together at the same client machine and were assisted by members of staff.



Figure 2: Translate@thon at Rhodes

Although the official server of Translate.org.za is hosted at Rhodes and can be reached through the Rhodes Intranet, one of the machines in the laboratory was used as an *ad hoc* Pootle server instead. This allowed us to test Pootle outside bandwidth constraints and network “bottlenecks”. The server was a fairly recent machine originally intended for desktop use, with 2 GB RAM and a 1.6 GHz processor. Under these circumstances, serving 30 connections simultaneously did not seem to affect the performance of the server at all.

The original POT (PO templates) files necessary for Pootle to start a new project were retrieved from the official Horde project site [27]. In order to maximise efficiency, strings relating to the user interface were prioritised over those of the administrative interface. Initially, the workload was distributed evenly among the participants. Following the suggestions of the coordinating staff members, it was subsequently re-allocated to maximise the use of very good translators.

Approximately 10 800 out of 12 800 strings (i.e. 84%) were translated. After the translate@thon, the project was moved to the official Translate.org.za Pootle server, where entries were thoroughly reviewed by members of the SANTED team. Logs show that some of the students who took part in the translate@thon registered with the main server to continue contributing. Existing translations were merged into various other components of the Horde framework (i.e. Ingo, Mimp, Nag and Turba). Using the Translate Toolkit, the PO files used for the translation were converted into the MO (Machine Object) files used by the application. Thanks to a PHP script supplied by the Horde development team, Xhosa could be added to the list of

languages available for the Rhodes e-mail system and is now available to staff and students (Figure 3).

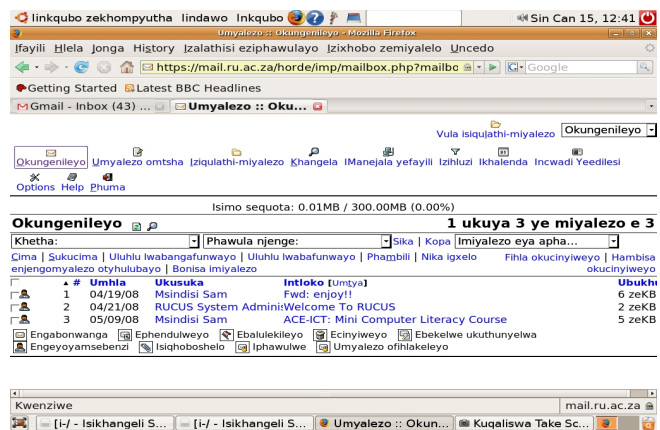


Figure 3: Webmail (plus Gnome and FireFox) in Xhosa

## V.CONCLUSION

In this paper we describe Pootle, a Web application primarily designed with the localisation of open-source software into indigenous South African languages in mind. This is expected to improve access to computers and the Internet by speakers of an African language with low levels of English proficiency. Pootle can contribute to tackling some of the problems commonly associated with the use of such languages in the ICT domain, such as lack of resources and low levels of support among their own native speakers.

The experience of the first translate@thon at Rhodes shows that freely-available open-source software can be localised in any language in a relatively inexpensive way. Most importantly, the enthusiastic participation and subsequent follow-up indicate a clear interest on the part of Rhodes and Fort Hare students to support the use of their language in the ICT domain. These type of events have the potential to catalyse young Africans around ICT-related issues and make them more relevant to them.

Being open-source, the software localised by Translate.org.za can be freely downloaded and used by educational institutions, Government bodies and NGOs. It therefore has the potential to reach millions of Africans in marginalised communities, and promote access by making ICT a little bit less “foreign” to them.

## ACKNOWLEDGEMENT

We would like to acknowledge the sponsors of the Telkom Centres of Excellence of Rhodes University and the University of Fort Hare, of Translate.org.za and of the SANTED programme. We would also like to thank the *Opera Universitaria di Trento*, the Carnegie-Mellon foundation. Special thanks to the community of developers and translators who have worked with Pootle, to the Rhodes

IT Division for their support, to all of the Rhodes SANTED team and the Rhodes and Fort Hare students for their enthusiastic participation in the translate@thon. The translate@thon photo is courtesy of David McGregor of the Daily Dispatch.

#### REFERENCES

- 1] V. N. Webb and J. Kembo-Sure, *African Voices: An Introduction to the Languages and Linguistics of Africa*. Oxford: Oxford University Press, 2000.
- 2] G. J. de Vreede. Field experiences with collaboration technology: A comparative study in Tanzania and South Africa. *Information Technology for development*, vol. 10, no. 3, pp. 201—219, 2003.
- 3] G. M. de Schryver. A perspective for the African Languages. *Nordic Journal of African Studies*, vol. 11, no. 2, pp. 266–282, 2002.
- 4] K. Keniston. (1997). Software localisation: Notes on Technology and Culture. [Online]. Available: <http://web.mit.edu/kken/public/papers.htm>
- 5] Pan African localisation WikiL [Online]. Available: <http://www.panafrill10n.org/>
- 6] Google (South African version). [Online]. Available <http://www.google.co.za/>
- 7] Microsoft translates Windows XP into isiZulu. SA Goodnews, 20 April 2006. [Online]. Available: [http://www.sagoodnews.co.za/science\\_technology/microsoft\\_translates\\_windows\\_xp\\_into\\_isizulu.html](http://www.sagoodnews.co.za/science_technology/microsoft_translates_windows_xp_into_isizulu.html)
- 8] List of localised MS Operating Systems. [Online]. Available: <http://www.microsoft.com/globaldev/reference/oslocversion.msp>
- 9] Microsoft. Wikipedia. [Online]. Available: <http://en.wikipedia.org/wiki/Microsoft>
- 10] Draft Report of the Language Plan Task Group (LANGTAG). Language Plan Task Group (LANGTAG). Pretoria: Department of Arts, Culture, Science and Technology, 1996.
- 11] Open Source. Wikipedia. [Online]. Available: [http://en.wikipedia.org/wiki/Open\\_source](http://en.wikipedia.org/wiki/Open_source)
- 12] S. Freeman. Contribution and Motivation in Hybrid Open Source Language Technology Development. [Online]. Available: [http://folk.uio.no/lrisan/Linux/workshop/Freeman\\_Motivation\\_manuscript.pdf](http://folk.uio.no/lrisan/Linux/workshop/Freeman_Motivation_manuscript.pdf)
- 13] South Africans told “Go Open”. Southafrica.info. 17 February 2005. [Online]. Available: <http://www.southafrica.info/about/science/leadeross.htm>
- 14] Ubuntu. [Online]. Available: <http://www.ubuntulinux.com/>
- 15] localisation International Standards Association. [Online]. Available: <http://www.lisa.org/>
- 16] Lucene . [Online]. Available: <http://lucene.apache.org/>
- 17] List of Pootle Servers. [Online]. Available: [http://translate.sourceforge.net/wiki/pootle/live\\_servers](http://translate.sourceforge.net/wiki/pootle/live_servers)
- 18] D. C. Brabham. Crowdsourcing as a model for problem solving: and introduction and cases. *Convergence*, vol. 14, no. 1, p. 75, 2008.
- 19] Students translate webmail into local language. Independent On-Line (IOL) Technology. 17 May 2007. [Online]. Available: [http://www.ioltechnology.co.za/article\\_page.php?iSectionId=2888&iArticleId=3836432](http://www.ioltechnology.co.za/article_page.php?iSectionId=2888&iArticleId=3836432)
- 20] List of translate@thons. [Online]. Available: [http://translate.sourceforge.net/wiki/guide/translateathon\\_halloffame](http://translate.sourceforge.net/wiki/guide/translateathon_halloffame)
- 21] M. Sam, L. Dalvit and R. Kaschula. “Development and implementation of ICT terminology in isiXhosa”. Kenton conference. Phumula, South Africa, 24-28 October 2008.
- 22] Translate.org.za Pootle Server. [Online]. Available: <http://pootle.translate.org.za/>
- 23] Translate.org.za File Server. [Online]. Available: <http://downloads.translate.org.za/>
- 24] Unable to run Pootle. Bug #83639 in Pootle (Ubuntu). [Online]. Available: <https://bugs.launchpad.net/ubuntu/+source/pootle/+bug/83639>
- 25] Problem using Lucene on Ubuntu. [Online]. Available: <http://www.nabble.com/Problem-using-Lucene-on-Ubuntu-td15543843.html>
- 26] Fully localised lab at Rhodes. Translate.org.za blog. 16 March 2007. [Online]. Available: <http://translate.org.za/blogs/dev/category/translation/>
- 27] Horde. [Online]. Available: <http://horde.org/>

Dr Lorenzo Dalvit (*Laurea* in Sociology (UNITN, Italy); MA in Linguistics (Rhodes, South Africa)).

Lorenzo Dalvit is a Lecturer in ICT Education at Rhodes University and the Research and ICT coordinator of the SANTED programme within the Rhodes School of Languages (African Studies Section). He has worked for several years as a researcher on multilingualism and ICT for the Telkom Centres of Excellence of Rhodes and Fort hare Universities in South Africa.

Dr Dalvit's areas of academic interest are ICT for development and localisation into African languages. As part of his multi-disciplinary interests, he is currently completing a PhD in Education. He has published in academic journals and presented at conferences within various disciplines both in South Africa and internationality. He is active in post-graduate supervision and in establishing collaborations with universities in other parts of Africa.