



PAN
Localization

Survey of Language Computing in Asia 2005

Sarmad Hussain
Nadir Durrani
Sana Gul

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences



www.nu.edu.pk

IDRC  CRDI

Canada

www.idrc.ca

Published by

Center for Research in Urdu Language Processing
National University of Computer and Emerging Sciences
Lahore, Pakistan

Copyrights © International Development Research Center, Canada

Printed by Walayatsons, Pakistan

ISBN: 969-8961-00-3

This work was carried out with the aid of a grant from the International Development Research Centre (IDRC), Ottawa, Canada, administered through the Centre for Research in Urdu Language Processing (CRULP), National University of Computer and Emerging Sciences (NUCES), Pakistan.

Lao

Lao is a Thai-Kadai language spoken by approximately 15 million people in Laos and Thailand [1]. It is one of the many Tai languages of the Southwestern branch. The origins of the Tai language can be traced back to Guangxi-Guizhou-Hunan region in southern China and bordering areas of northern Vietnam about 2000 years ago [2].

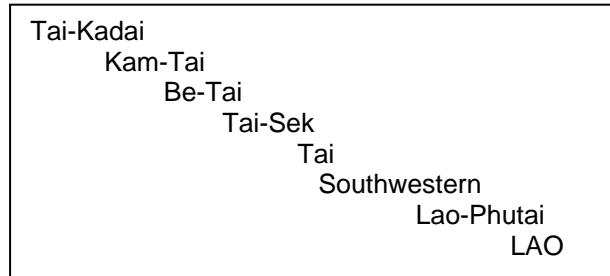


Figure 1: Language Family Tree for Lao [1]

Traditionally, Lao language and its literature have been written in two scripts, Lao and Tham. Lao script has been largely influenced by the ancient Thai script that was developed from the old Khmer script [2].

Character Set and Encoding

Lao computing has progressed slowly due to lack of Lao encoding standards [3]. Laos' Science Technology and Environment Agency (STEA) is currently working towards IT standardization. Due attention is being given to formulation and implementation of national standards and adoption of available international standards [3]. Unicode has also included Lao script from 0E80-0EFF but is not widely used. Current ad hoc standard is based on encoding developed Lao Script for Windows [4]. Earlier Lao encoding includes IBM ISO-8 Code Page 01033 and EBCDIC Code Page 01032 developed by STEA [3].

Fonts and Rendering

Multiple fonts are available for Lao based on ad hoc encodings [4] and based on Unicode standard (e.g. [5]), in TTF and OTF formats. Even though basic research in Lao font development has been conducted, advanced issues e.g. mark to mark positioning and hinting, are still unresolved. The most widely used font for Lao is Saysettha OT. It is developed by Lao Script for Windows [4] and includes proper hinting. STEA, through PAN Localization project, has also developed two Lao fonts [5].

Microsoft Platform

Windows XP does not provide in-built fonts that support Lao script. Locally some of the fonts have been developed that adequately satisfy Lao script requirements. The following figure presents the Lao font rendering on Microsoft platform.

ໂຄງການພັດທະນາພາສາລາວໃນຄອມພິວເຕີ ໂຄງການພັດທະນາພາສາລາວໃນຄອມພິວເຕີ

Figure 2: Two Lao fonts Displayed on Microsoft Platform [5]

Linux Platform

Several Lao Open Type fonts are developed to be used on the Linux platform. Popularly used are Saysettha OT and Jason variety Lao Unicode fonts. These can also be rendered in Open Office and Mozilla.

Keyboard Standard

No keyboard layout has been standardized for Lao language. However there are few ad hoc standards that are popularly used for Lao text input. A commonly used Lao keyboard is based on the Lao typewriter layout. Figure 3 shows the keyboard layout based on the typewriter.



Figure 3: The Old Typewriter Based Keyboard Layout [6]

Lao Script for Windows is a shareware Lao solution for Windows. Its LSWin version 5.0 has a separate keyboard interface specifically designed for Lao input [4].



Figure 4: Keyboard Layout Based for LSWIN [4]

Lao Software [7] has developed Lao Unikey, the Lao keyboard drivers compatible with Unicode. Five popularly used Lao keyboards have been included in Lao Unikey. They are Duang Jan, Lao US, Lao France, Lao France New and Sida Thong. Figure 5 shows the Lao Unikey developed by Lao software. Other keyboards are also available, e.g. [6].



Figure 5: Lao Unikey [7]

Microsoft Windows does not provide keyboard for Lao language. However many Lao localization groups have developed keyboards for Lao. The following figure displays the facility to enable Lao keyboard support on Windows platform.



Figure 6: Setting-up Lao Keyboard on MS Windows

Keyboards based on popularly used Lao ad hoc encodings are being developed for KDE and GNOME through LaoNux project [8].

Collation Sequence

There is no collation standard for Lao language. However two most popularly used schemes are based on lexical order in dictionaries. Projects are underway at National University of Laos and STEA, through PAN Localization project, to standardize the collation order of Lao.

Lao sorting utility has been developed by STEA based on lexicographic orders. It does syllable based sorting, after rule-based syllabification of input string. Sorting utility is also available through LSWIN and Lao Software projects. These orders are based on existing dictionaries.

Locale

CLDR 1.3 includes locale definitions for Lao (lo_LA). It includes date, time, currency symbol and names of the days and months. In addition the National University of Laos (NUoL) is working on compiling NLS file for Lao language. Currently Lao locale is not included in any version of Microsoft Windows. Locale file is being developed under LaoNux project to enable Lao locale on Linux platform. Until now the date, time and currency formats have been defined. LaoNux is a KDE based Lao Linux distribution [8].

Interface Terminology

No terminology standard exists for Lao. Lao version does not exist for Microsoft software. A Lao Editor has been developed on Microsoft platform by STEA through PAN Localization project. The Lao Editor provides Lao and English interface. Help files have also been translated in Lao language.

The LaoNux project team is working to develop Lao technical glossary for localization of open source software. Currently there are no considerable outputs on the glossary translation of Linux applications. The LaoNux project is at its initial phases [8]. There is also some work on KDE translation of Lao, as shown in the figure below.

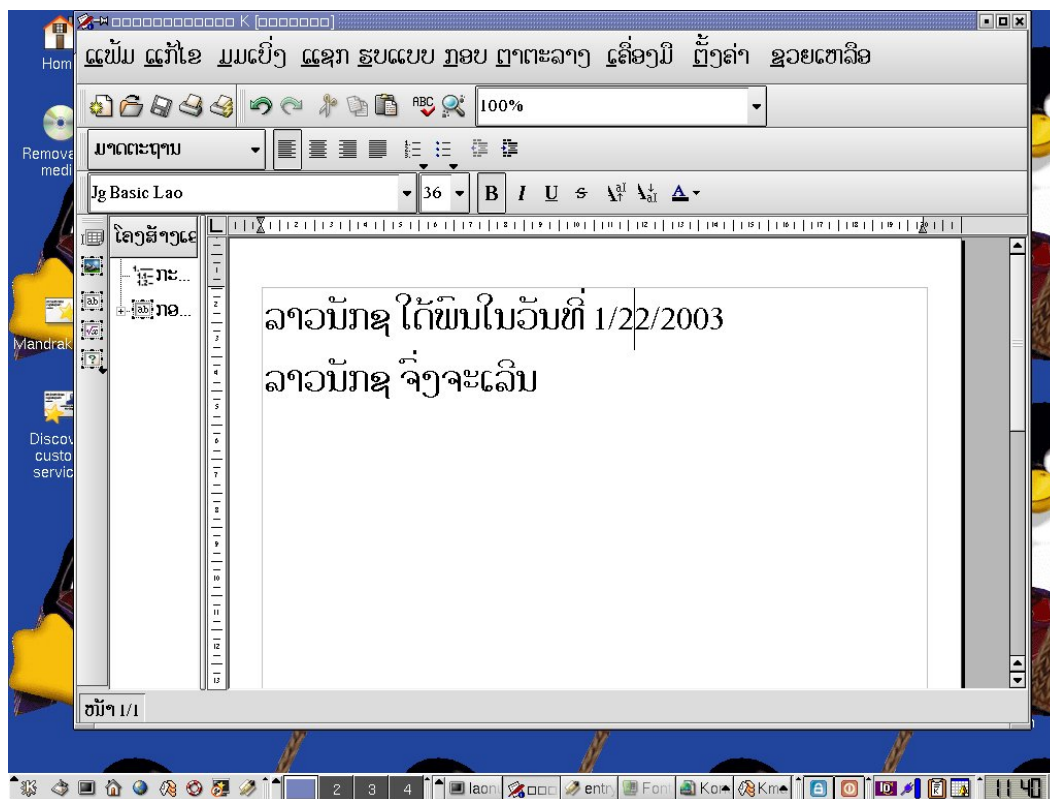


Figure 7: Localized KDE environment in Lao [9]

Status of Advanced Applications

Some research is presently underway for the development of Lao language processing applications. STEA, through PAN Localization project, has already developed encoding converters, Lao syllabification and sorting utility and a Lao lexicon, all incorporated within a Lao text editor. Lao lexicon is also available through Lao Software [7]. Lao Software has developed support for Lao online translation service. The Lao translation applications developed include, Lao- French-Lao word translation and transcription of texts utility from Thai to Lao [7].

Lao script does not have space between words. Thus, line breaking utilities are needed for basic word processing. LSWIN has developed a lexicon based utility. STEA has also developed a utility based on the syllable structure in Lao language [5]. Reordering is also done during line breaking algorithm when some of the vowels and tone marks are used in reverse order. Work is under progress on more advanced applications through PAN Localization project.

References

- [1] http://www.ethnologue.com/show_language.asp?code=lao
- [2] <http://www.omniglot.com/writing/lao.htm>
- [3] <http://www.undplao.org/governance/ProjectDocuments/ICT%20ProDoc.pdf>
- [4] <http://www.laoscript.net/>
- [5] <http://www.laol10n.info.la/Eng.htm>
- [6] http://www.laohub.com/static_html/lang/keyboard/laokey05v11.pdf
- [7] <http://www.laosoftware.com/index.php?Langue=en>
- [8] <http://opensource.muanglao.com/laonux.htm>
- [9] <http://i18n.KDE.org/screenshots/>