Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation internationale de normalisation
Международная организация по стандартизации

Doc Type:	**Working Group Document**
Title:	**Proposal to add minority characters to Lao script**
Source:	**Martin Hosken**
Action:	**For consideration by JTC1/SC2/WG2**
Date:	**2010-10-05**

Introduction: This proposal is to add two extra consonants to the Lao block outside the main consonant list. The characters are used for the Khmu[1] language. The Khmu are a major ethnolinguistic group of northern Southeast Asia. The largest proportion of the estimated 700,000 Khmu resides in Laos, and smaller communities are found in Vietnam, Thailand, and China. Khmu is a Mon-Khmer language and encompasses a complex web of ethnic and dialect groupings. Khmu was first documented in the late 19th century by French scholars in Luang Prabang, Laos. Systematic linguistic analysis, documentation, and efforts toward language development began among Khmu in Laos in the mid-1950's with the work of William Smalley, and have been continued by numerous other scholars. A body of Khmu literature has been produced using both Lao- and Roman- script orthographies that are based on an 'Eastern' Khmu dialect. More recently the Lao script based orthography is being taught among the Khmu and it is felt that now is the right time for ISO 10646 to be able to represent this orthography.

The two consonants represent /g/ and final /ny/, neither of which occur in Lao. The characters are produced by the overlaying a modified LAO LETTER MAI KAN (U+0EB1) over a LAO LETTER KO (U+0E81) and LAO LETTER NYO (U+0E8D) respectively in a productive manner to create two new characters. The mai kan itself and the original characters have no separate meaning that combines to produce the new characters, the production is purely graphical and not linguistic. A similar approach to character production is seen in Thai with THAI CHARACTER SO RUSI (U+0E29).

As such, since these are new characters that have no counterparts in Thai, there is no reason to insert them into the main consonant list for Lao. Instead, they should be added as extra characters at the end of the block. We propose the characters be added as:

```
0EDE;LAO LETTER KHMU GO;Lo;0;L;;;;;N;;;;;
0EDF;LAO LETTER KHMU NYO;Lo;0;L;;;;;N;;;;;
```

Khmu also makes use of LAO SEMIVOWEL SIGN NYO (U+0EBD) as a consonant. But this requires no extra encoding.

Collation: The characters receive a sort order based on their relative positions in the consonant chart for Khmu. In the case of LAO LETTER KHMU NYO, it is given a primary relative order in case the character gets used in other languages. In each case the added character is sorted directly before the character with the unadorned glyph, on which it is based:

```
&[before 1]0E81 < 0EDE
&[before 1]0E8D < 0EDF
```

---

1 ISO 639-3 code: KJG

Lao

0ED
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E ꫞
F ꫟

Consonants

0EDE   ꫞   LAO LETTER KHMU GO

0EDF   ꫟   LAO LETTER KHMU NYO

ພະຍັນຊະນະຂອງພາສາກ຺ຶມຫມຸ
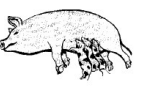
ຮ g ຮາງ gaang	ກ k ກະ ka'	ຄ kh ຄວດ khuat	ງ ng ງວງ nguang	ຫງ hng ເຫງາະ hngo'	ງ່ 'ng ແງ່ດ 'ngèèt
ຈ j ຈ຺ວຫ joh	ຈ c ຈອກ cook	ຊ ch ຊາງ chaang	ສ s ເສາະ so'	ຍ (_ຠ) ກ̃ñ ເຍືອຣ ñwar	ຫຍ hñ ຫຍອດ hñoot
ດ d ດາງ daang	ຕ t ຕ຺ິຫ tih	ທ th ທັບ thap	ນ n ນ຺ົບ nôp	ຫນ hn ຫນ຺ິມ hnwm	ນ່ 'n ນ່ວງ 'niang
ບ b ແບມ bèèm	ປ p ປັດ pat	ພ ph ເພ຺ືອນ phwan	ຟ f ເຟ຺ິ fe	ມ m ເມາະ mo'	ຫມ hm ຫມູຈ hmuuc
ມ່ 'm ເມັ຺ດ 'mét	ຍ y ຍະ ya'	ຫຍ hy (ຍຫ) (yh) ຫຍຯງ hyaang	ຍ່ 'y ຍັ່ງ 'yang	ຣ r ຣາງ raang	ຫຣ hr ຫຣາງ hraang
ລ l ແລ lè	ຫລ hl ຫລະ hla'	ວ v ວາກ vaak	ຫວ hv ຫວັດ hvat	ວ່ 'v ວ່າຍ 'vaay	ຫ h ຫວລ hual

ອ ອອງ oong	**ພະຍັນຊະນະໃຫຍ່ຂອງກ຺ຶມຫມຸໂຣມັນ** G K Kh Ng Hng 'Ng J C Ch S Ñ Hñ D T Th N Hn 'N B P Ph F M Hm 'M Y Hy 'Y R Hr L Hl V Hv 'V H

Figure 1: Khmu consonant alphabet chart. Proposed characters are circled in red.

Bibliography

Suksavang Simana, Somseng Sayavong and Elisabeth Preisig. 1994. ວັດຈະນານຸກົມ ຂະມຸ ລາວ ຝະລັ່ງ ອັງກິດ Kmhmu'-Lao-French-English dictionary. Vientiane, Lao P.D.R.: Ministry of Information and Culture, Institute of Research on Culture.

Suksavang Simana and Elizabeth Preisig (English). 1998. ແນວ ຢັດ ມັຫ ເຄະ ກອນ ກົມ ຫມຸ Kmhmu' livelihood: farming the forest.Vientiane, Lao PDR: Institute for Cultural Research, Ministry of Information and Culture.

ປຶມ ຫຽມ ຫຶຣເລຊະ ກົມຫມຸ (Book for Learning Kmhmu').

ຕຶຣຄ້ອຫ ຄອມ ອານ ສຶມມ່າຍ ເຫລັມ 1 (Easy-Reading Fun Stories – Storybook 1)

ຕຶຣຄ້ອຫ ຄອມ ອານ ສຶມມ່າຍ ເຫລັມ 2 (Easy-Reading Fun Stories – Storybook 2)

Acknowledgements

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[2]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html **for guidelines and details before filling this form.**
Please ensure you are using the latest Form from http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:**	*Lao Extensions*
2. Requester's name:	*Martin Hosken*
3. Requester type (Member body/Liaison/Individual contribution):	*Individual contribution*
4. Submission date:	*08/09/10*
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	*X*
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:
   a. This proposal is for a new script (set of characters):
   Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: *X*
   Name of the existing block: *Lao*

2. Number of characters in proposal: *2*

3. Proposed category (select one from below - see section 2.2 of P&P document):
   A-Contemporary   *X*   B.1-Specialized (small collection)   B.2-Specialized (large collection)
   C-Major extinct   D-Attested extinct   E-Minor extinct
   F-Archaic Hieroglyphic or Ideographic   G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?   *yes*
   a. If YES, are the names in accordance with the "character naming guidelines"
      in Annex L of P&P document?   *yes*
   b. Are the character shapes attached in a legible form suitable for review?   *yes*

5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
   publishing the standard?   *SIL*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
   used:

6. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?   *yes*
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources)
   of proposed characters attached?   *yes*

7. Special encoding issues:
   Does the proposal address other aspects of character data processing (if applicable) such as input,
   presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?   *no*

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UCD.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?　　*no*
　　If YES explain

2. Has contact been made to members of the user community (for example: National Body,
　　user groups of the script or characters, other experts, etc.)?　　*yes*
　　　　If YES, with whom?　　*Suksavang Simana*
　　　　If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:
　　size, demographics, information technology use, or publishing use) is included?　　*yes*
　　Reference:　　*this document*

4. The context of use for the proposed characters (type of use; common or rare)　　*common*
　　Reference:

5. Are the proposed characters in current use by the user community?　　yes
　　If YES, where?　Reference:

6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
　　in the BMP?　　*yes*
　　　　If YES, is a rationale provided?　　*yes*
　　　　　If YES, reference:　　*addition to existing BMP block*

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?　　*no*

8. Can any of the proposed characters be considered a presentation form of an existing
　　character or character sequence?　　*no*
　　　　If YES, is a rationale for its inclusion provided?
　　　　　If YES, reference:

9. Can any of the proposed characters be encoded using a composed character sequence of either
　　existing characters or other proposed characters?　　*no*
　　　　If YES, is a rationale for its inclusion provided?
　　　　　If YES, reference:

10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
　　to an existing character?　　*yes*
　　　　If YES, is a rationale for its inclusion provided?　　*yes*
　　　　　If YES, reference:　　*this document*

11. Does the proposal include use of combining characters and/or use of composite sequences?　　*no*
　　If YES, is a rationale for such use provided?
　　　　If YES, reference:　　*no*
　　Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?　　*no*
　　　　If YES, reference:

12. Does the proposal contain characters with any special properties such as
　　control function or similar semantics?　　*no*
　　　　If YES, describe in detail (include attachment if necessary)

13. Does the proposal contain any Ideographic compatibility character(s)?　　*no*
　　If YES, is the equivalent corresponding unified ideographic character(s) identified?
　　　　If YES, reference: