

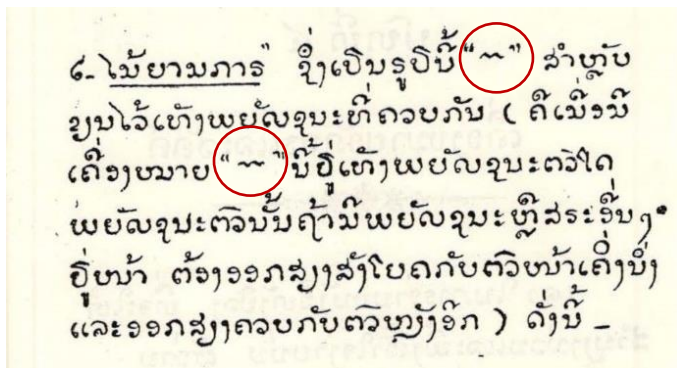
Proposal to Encode Lao Yamakkan

Vinodh Rajan vinodh@virtualvinodh.com

Jayasāro Bhikkhu ven.chaisiri@gmail.com

L2/17-106R proposed the encoding of several characters required for the writing of Pali in the Lao script. This proposal requests the encoding of one more character, *Lao Yamakkan*, to further support Lao Pali in the UCS. This will not only facilitate proper digitization of existing documents but also allow contemporary use of the character.

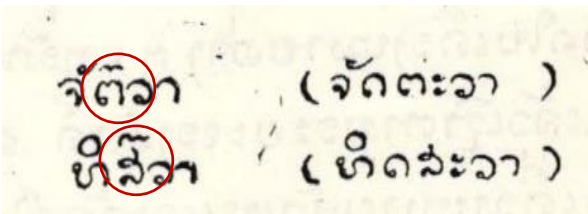
Certain Pali publications in the Lao script use a sign cognate with *Thai Character Yamakkan* U+0E4E to denote a consonant that is shared between the preceding syllable and the succeeding syllable at the inter-syllabic boundary. From [1]:



໒. ໄມ້ຍາມກາຣ “ ຊຶ່ງແປນຣູປນີ້ “~” ສຳຫຼັບ ຊຽນໄວ້ເທິງພັດສຸນະທີ່ຄວບກັນ (ຄືເມື່ອມີ ເຄື່ອງໝາຍ “~” ນີ້ຢູ່ເທິງພັດສຸນະຕວິໄດ ພັດສຸນະຕວິນັ້ນຖ້າມີພັດສຸນະຫຼືສະຣີຍໆ ຢູ່ໜ້າ ຕ້ອງອອກສຽງສັງໂຍຄກັບຕົວໜ້າເຄື່ອງນີ້ ແລະອອກສຽງຄວບກັບຕົວຫຼັງອີກ) ດັ່ງນີ້ -

This may be translated as:

“2. *Maiyāma(k)kāra*“ This figure “~” is to be written on top of consonantal clusters (When the sign “~” is on top of a consonant and if there are other consonants or vowels in front of it, it is to be pronounced 1st consonant half and a 2nd consonant half) as follows”



ຈັຕວາ (ຈັດຕະວາ)

ທິສວາ (ທິດສະວາ)

21-Apr-21

In essence, this sign appears to mark a consonant that is simultaneously a part of two consonant clusters (by splitting itself into two halves). The first half is associated with the preceding syllable and the second half with the succeeding syllable. This is different from gemination where a consonant doubles itself.

ga^ˆt_vā = ga_t_vā

di^ˆs_vā = di_s_vā

It probably applies to consonantal clusters that are composed of a *varga* (plosive/nasal) and *non-varga* (glide/liquid/fricative) consonants. However, the current attestation is only limited to clusters with -v, -y or -h as the final consonant.

The Lao term *Yamakkan* is most likely to be derived from the Pali/Sanskrit term *yamaka-kāra*, which can be roughly translated as *the twin letter*.

A parallel usage is also seen in some Thai orthographies for Pali that differentiate consonant clusters based on their components. Here, Thai *Thanthakhat* U+0E4C is used for normal clusters and *Yamakkan* U+0E4E for *varga/von-varga* clusters.

สํกัยปุตโต *sakyaputto*

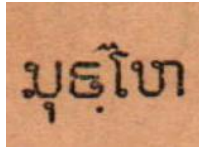
(From <https://www.scribd.com/document/318062469/2016-LahuGaru-Thai-Phonetic-Alphabet-Pa-l-i>)

Visual Appearance

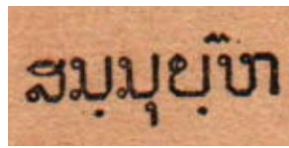
The character is attached at the top-right position of a consonant and usually extends over the succeeding character as seen below. An alternate positioning over a single base is also attested in [1].

(From [2])

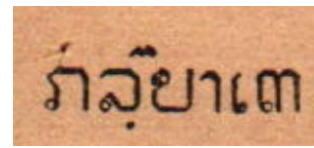
21-Apr-21



मुहु *muḥ*



समुय *sammuy*



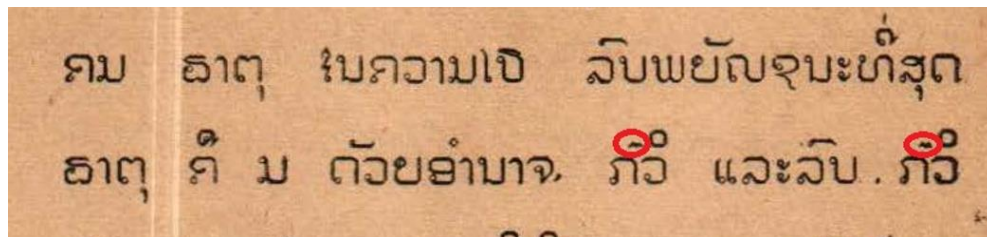
कल्य *kalyā*

In Pali text, the character appears almost always with the sign Pali *Virama* U+0EBA as evident by the above samples. As the sign applies to a dead consonant (that becomes eventually split), it should be recommended that the sign follow the sign Pali *Virama* in terms of input order as shown below.

ຍືຫ = ຍ + ຸ + ື + ຫ

ລືຍ = ລ + ຸ + ື + ຍ

However, when appearing amidst Lao text it is common for it to appear without the Pali *Virama*.



There is also a glyphic variation of the character taking a form similar to that of Thai *Yamakkan* over a single base.

ຍ	ື	ຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ
ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື	ຍຸ	ຍື

(From https://laofonts.tripod.com/keymap_Khunburom.html)

Interaction with Tone Marks

As a Pali-specific character, it is very unlikely that Lao *Yamakkan* will ever appear alongside tone marks (There are currently no attestations that we know of). However, if there is a need to place a tone mark over a syllable, it can adhere to the usual Lao script grammar, by following the *Yamakkan* in terms of input order and be placed over it similar to other top-positioned vowel signs.

21-Apr-21

Character to be encoded

We propose to encode the following character in the Lao block of the UCS with the associated character properties.

OECE  Lao Yamakkan

```
OECE;LAO YAMAKKAN;Mn;0;NSM;;;;;N;;;;;
```

Indic Syllable Properties

It must be noted that the Thai *Yamakkan* is assigned a Indic Syllable Property of “Pure Killer”. But this does not match Lao *Yamakkan*’s attested usage. As discussed earlier, even though the description might appear to suggest gemination, it rather denotes the sharing of a consonant between two syllables. Therefore, it is proposed to classify it as a syllable modifier rather than a gemination mark.

The following addition should be made to the IndicSyllabicCategory.txt file under:

```
# Indic_Syllabic_Category=Syllable_Modifier
OECE                   ; Syllable_Modifier # Mn                   LAO YAMAKKAN
```

Indic Positional Category

The following addition should be made to the IndicPositionCategory.txt file under:

```
# Indic_Positional_Category=Top
OECE                   ; Top # Mn                   LAO YAMAKKAN
```

Collation

The character must be collated after Lai Sign Pali *Virama* U+OEBA.

21-Apr-21

Additional Attestations

From [3]

116

ຍະນະ	ນ. ຄູ, ແຜ່ດ (ປ.)	ຍາມະໂຍກ	ນ. ເວລາທີ່ປະກອບດ້ວຍອິກດີ.
ຍະນະກະ-ຍະນິກ	ນ. ຄູ, ເປັນຊື່ສັນພິນັງໃນພາສາສາກາວດັດອ- ພັນທີ່ເປັນຄຳກັມ ແລະຄຳນິລາວໃຊ້ວ່າ ຍອດນຸກ.	ຍາມັກການ	ນ. ເຄື່ອງໝາຍກຳກັບພັນຊນະ ສຳລັບໃຊ້ ອ່ານຄວບກັນ (ສ.) (ຮູບດັ້ງນີ້: ^)
ຍະນະກະປະຕິຫານ	ນ. ອິດທີ່ສະແດງເປັນຄູ່, ເປັນອິດທີ່ພະຍາຍາມເຈົ້າ ທັງສະແດງສູ່ພວກເດັງຂະຕິຍູ່ວິໄນສາກສອງ ແຂວງຮາຊ ສີ (ປ.)	ຍານາ	ນ. ຊື່ເມືອງສວັນຊັນທ໌ ມີ ເທວະດາ ໃນສວັນຊັນທ໌ - ເອີ້ນວ່າ ຍາມະເທວະບູດ (ປ.)

ຍາມັກການ ນ. ເຄື່ອງໝາຍກຳກັບພັນຊນະ ສຳລັບໃຊ້ອ່ານຄວບກັນ (ສ.) (ຮູບດັ້ງນີ້: ^)

Yāmakkan n. mark directed consonants for reading the cluster consonants (s.) (figure as follows: ^)

From [2]:

- ດິວ -

ອຊຊຕຕຕິ (ລີ ອ ອາຄົມ)

ບຸຣິສາ ເອກວຈນໍ	ພຫຸວຈນໍ
ປ. ອກຕຸທິ	ອກຕຸທິ, ອກຕຸທິສຸ
ນ. X ອກຕຸທິ	ອກຕຸທິຕຸ
ອຸຕ. ອກຕຸທິ	ອກຕຸທິນິທາ

ກວິສສນຕິ (ລີ ອິ ອາຄົມ)

ບຸຣິສາ ເອກວຈນໍ	ພຫຸວຈນໍ
ປ. ອກຕຸທິສສຕິ	ອກຕຸທິສສນຕິ
ນ. ອກຕຸທິສສສິ	ອກຕຸທິສສຕຸ
ອຸຕ. ອກຕຸທິສສນນິ	ອກຕຸທິສສນນ
ອກຕຸທິສສສິ	

ບຸນທັງ ແກ້ວບົວສາຍ
ສ. ສາຍບຸນມີ

ບາລິເວຍຍາກອຕ

ວຈິວິກາຄ ກາຄທິສອງ

ອາຂຸຍາຕ ແລະ ກິຕກ

ໂດຍ

ນຫາລິລາ ວິຈວິຈິສ

ພ. ສ. ເຊຊື່

21-Apr-21

References

1. VeyyākaraṇaLao (1) Akkharavidhī (1935). 1st Edition. Lao Library.

ພ.ສ. ໒໔໗໘, ໄວຍບາກຣຸດລາວ (໑) ອັກຂະວິຊີ, ເປນຂອງຫໍພູສະໝຸດເບສລາວ
ພິມພໍ່ເກືອບີ່ນຶ່ງ

2. Mahā Silā Vīravong (1939). Pāli Veyyākaraṇa: Ākhyāta et Kitaka, Grammaire Pālie. Buddhist Institute of Vientiane.

ພ.ສ. ໒໔໘໒, ບາລີເວຍບາກຣຸດ, ວຈີວິກາດ ກາດບໍ່ສອງ, ອາຂບາດແລະກິຕກ ໄດຍ
ມະຫາສິລາ ວິຣະວົງສ໌

3. Mahā Silā Vīravong (1953). Dictionary Phāsā Lao.

ພ.ສ. ໒໕໐໓, ວັຈນານຸກົມພາສາລາວ ໄດຍ ມະຫາສິລາ ວິຣະວົງ

21-Apr-21

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest *Roadmaps*.

A. Administrative

1. Title:	<i>Proposal to Encode Lao Yamakkan</i>	
2. Requester's name:	<i>Vinodh Rajan, Jayasāro Bhikkhu</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual</i>	
4. Submission date:	<i>21-Apr-2021</i>	
5. Requester's reference (if applicable):		
6. Choose one of the following:		
This is a complete proposal:	<input checked="" type="checkbox"/> <i>Yes</i>	
(or) More information will be provided later:	<input type="checkbox"/>	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/> <i>Yes</i>	
Name of the existing block:	<i>Lao</i>	
2. Number of characters in proposal:	<i>1</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/> <i>Yes</i>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input type="checkbox"/>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/> <i>Yes</i>	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/> <i>Yes</i>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Jayasāro Bhikkhu, Sanit Phokhaphan, Vinodh Rajan</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>ven.chaisiri@gmail.com, sanit2001@hotmail.com, viinodh@virtualvinodh.com</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/> <i>Yes</i>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/> <i>Yes</i>	
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)?	<input checked="" type="checkbox"/> <i>Yes</i>	
<i>Collation</i>		

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 Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

21-Apr-21

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain	No
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.)? If YES, with whom? If YES, available relevant documents:	Yes <i>Jayasāro Bhikkhu is a native user</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	Yes <i>See Proposal</i>
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare <i>See Proposal</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Yes. By Jayasāro Bhikkhu</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	Yes Yes <i>Lao is encoded in the BMP</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	Yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:	No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	No
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference:	No
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	Yes Yes <i>See Proposal</i> Yes <i>See Proposal</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	No