

Revised Proposal to Encode Lao Characters for Pali

Vinodh Rajan vrs3@st-andrews.ac.uk, Ben Mitchell ben@fontpad.co.uk
Martin Jansche mjansche@google.com, Sascha Brawer sascha@google.com

This proposal requests the encoding of a set of characters that are required to support the Pali orthography in the Lao script.

1. Introduction

Lao is a South-East Asian script that is used to write the Lao language and some minority languages in Laos. Historically, Modern Lao like other South-East Asian scripts can be ultimately derived back to the Pallava ‘Grantha’ script. However, unlike its neighboring scripts that imported the entire “Indic” consonant repertoire, Lao lacks several characters that are required to accurately express the phonology of Pali/Sanskrit.

This lack of additional Indic characters in Modern Lao can be attributed to two major factors. In Laos, the writing of Pali (and other religious texts) was traditionally relegated to the Tham script. Although Tham is capable of representing Pali words accurately, it is not in everyday use, instead being reserved for special use by monks for religious purposes, and by academics for study. Laypeople, who do not generally know Tham, therefore do their everyday reading and writing in the relatively less complex Lao. However, Lao cannot faithfully represent Pali words, and by extension cannot transcribe religious texts faithfully.

The other factor is that the Lao language uses a phonemic spelling as opposed to the etymological spelling of many South-East Asian languages that retain Indic phonemes in loan words. Eventually, both the factors had the combined effect of reducing the core character set of Lao by retaining only those characters required for the native Lao language. This scenario is very similar to the historical development of the Tamil character set. The Grantha script was specifically employed for writing Sanskrit/Prakrit in the Tamil-speaking areas of India and Sri Lanka, while the Tamil language itself used a phonemic orthography that respelled loan words to suit the native phonology. As a result, the Tamil script never developed/adapted the full set of characters to represent Sanskrit sounds (apart from a few ad-hoc borrowings) unlike other Indian languages.

2. Etymological Orthography vs Phonemic Orthography

Lao employs a huge number of Indic loan words in its vocabulary. Due to the phonemic nature of Lao orthography, words with different Indic roots but the same

Lao pronunciation came to be written in the same way in the Lao script. This lack of etymological information results in homophonous words being easily confounded (unless disambiguated through context).

To enable the Lao script to completely and faithfully express the Pali language and also write Indic loan words with proper etymological spelling, Buddhist scholars tried to implement a script reform. In the 1930s Buddhist scholar Maha Sila Viravong at the initiation of the Buddhist Institute in Vientiane and approved by the Buddhist Academic Council, added an additional set of characters to support Pali (and also Sanskrit) by filling in the missing gaps. The Buddhist Institute using the expanded script published several books such as Dhammapada and Pali Grammar. Their aim was to make Pali more accessible to the common public. But the addition met with little widespread support and finally by 1975, these additional characters were mostly out of use. There is a revived interest in the characters and a few modern publications have been printed using them in an attempt to propose an etymological orthography for Lao.

buddham saraṇaṃ gacchāmi

बुद्धं सरणं गच्छामि

ພຸທຸ້ ສຣນຸ້ ຄຈຸຈາມີ

ພຸທຸ້ ສຣນຸ້ ຄຈຸຈາມີ

Fig 1: Sample Pali formula. Extended characters have been highlighted

Currently, due to the lack of these extended characters in Lao Unicode, publications that use the character set cannot be represented properly in plain text except through font hacks. Encoding the additional characters will allow existing materials to be faithfully transcribed into Unicode and also enable interested people to use an etymological orthography for Lao. These characters would additionally improve the round-trip transliteration between Lao and related South-East Asian scripts.

3. Pali Alphabet in extended Lao

The complete Pali alphabet in extended Lao script along with their Thai equivalents (as the closest script to Lao) is presented below. The characters in red are the additions by the Buddhist Institute. Maha Sila Viravong designed the additional Pali characters through his research into various epigraphic sources. Specifically, some precursors to the additional characters can be traced back to the Tai Noi script, which itself is the precursor to the modern Lao script. It must be noted that *Table 1* and *2* do not include Lao-specific characters that cannot be directly mapped to their Indic prototypes.

	<i>ka</i>	<i>kha</i>	<i>ga</i>	<i>gha</i>	<i>ṅa</i>
<i>Lao</i>	ກ	ຂ	ຄ	ຢ	ງ
<i>Thai</i>	ก	ข	ค	ฆ	ง
	<i>ca</i>	<i>cha</i>	<i>ja</i>	<i>jha</i>	<i>ṅa</i>
<i>Lao</i>	ຈ	ຊ	ຊ	ຊ	ຍ
<i>Thai</i>	จ	ฉ	ช	ฌ	ญ
	<i>ṭa</i>	<i>ṭha</i>	<i>ḍa</i>	<i>ḍha</i>	<i>ṇa</i>
<i>Lao</i>	ຝ	ຖ	ຜ	ຜ	ື
<i>Thai</i>	ฎ	ฏ	ฒ	ณ	ณ
	<i>ta</i>	<i>tha</i>	<i>da</i>	<i>dha</i>	<i>na</i>
<i>Lao</i>	ຕ	ຖ	ທ	ຣ	ນ
<i>Thai</i>	ต	ถ	ด	ธ	น
	<i>pa</i>	<i>pha</i>	<i>ba</i>	<i>bha</i>	<i>ma</i>
<i>Lao</i>	ປ	ຜ	ພ	ຜ	ມ
<i>Thai</i>	ป	ผ	พ	ภ	ม
	<i>ya</i>	<i>ra</i>	<i>la</i>	<i>va</i>	
<i>Lao</i>	ຢ	ຣ	ລ	ວ	
<i>Thai</i>	ย	ร	ล	ว	
	<i>śa*</i>	<i>ṣa*</i>	<i>sa</i>	<i>ha</i>	<i>ḷa</i>
<i>Lao</i>	ຮ	ຮ	ສ	ຫ	ຣ
<i>Thai</i>	ศ	ษ	ส	ห	ฬ

* Sanskrit specific consonants

Table 1: Pali Consonants

	<i>k</i>	<i>ka</i>	<i>kā</i>	<i>ki</i>	<i>kī</i>	<i>ku</i>	<i>kū</i>	<i>ke</i>	<i>ko</i>	<i>kaṃ</i> *
<i>Lao</i>	ກ	ກາ	ກາ̄	ກີ	ກີ̄	ກຸ	ກູ	ເກ	ໂກ	ກໍ
<i>Thai</i>	ก	กา	กา̄	กิ	กิ̄	กู	กู	เก	โก	กั

* ṃ is the nasalization sign not a vowel

Table 2: Pali Vowel Diacritics

Like most Indic-derived South East Asian languages, the phonetic value of the characters have deviated very much from their Indic prototype. Hence, when reading Pali in extended Lao, the native “Lao” consonants must be read with their “Indic” reading . Also, in the Lao orthography all vowels are always explicitly written, but this is not the case in the Pali orthography. The consonants are now assumed to have the inherent vowel “a”. This inherent vowel is killed using the Virama ◌̣ (analogous Thai Phinthu U+0E3A) (These “Indic” reading/writing rules should be ideally applied when reading Pali in most South East Asian scripts including Thai). Even though, Pali is sometimes written using the native vowel convention of marking the inherent ‘a’ explicitly. (See Fig. 11)

To summarize, 14 consonantal letters and 1 combining character should be additionally encoded in the Lao block to support the Pali orthography.

4. Characters Proposed

Code Point	Glyph	Character Name
U+0E86	ຮ	LAO LETTER PALI GHHA
U+0E89	ຮ	LAO LETTER PALI CHA
U+0E8C	ຮ	LAO LETTER PALI JHA
U+0E8E	ຮ	LAO LETTER PALI NYA
U+0E8F	ຮ	LAO LETTER PALI TTA
U+0E90	ຮ	LAO LETTER PALI TTHA
U+0E91	ຮ	LAO LETTER PALI DDA
U+0E92	ຮ	LAO LETTER PALI DDHA
U+0E93	ຮ	LAO LETTER PALI NNA
U+0E98	ຮ	LAO LETTER PALI DHA
U+0EA0	ຮ	LAO LETTER PALI BHA
U+0EA8	ຮ	LAO LETTER SANSKRIT SHA
U+0EA9	ຮ	LAO LETTER SANSRIT SSA
U+0EAC	ຮ	LAO LETTER PALI LLA
U+0EBA	ຮ	LAO SIGN PALI VIRAMA

Table 3: Proposed character list

4.1 LAO LETTER PALI NYA

The encoding of LAO LETTER PALI NYA deserves a special discussion. Under the current Unicode encoding of Thai and Lao, the character that is analogous to THAI CHARACTER YO YING ຢ (U+0E0D) is LAO LETTER NYO ອ (U+0E8D). But they are not exactly cognate characters. The current Unicode mapping was probably a result of conflating the historical and modern phonetic realizations of the two characters as explained below.

In terms of script evolution, the cognate of ອ U+0E8D is THAI CHARACTER YO YAK ຢ (U+0E22) and the Lao script cognate of ຢ U+0E0D was lost in Modern Lao, until the reintroduction of the reconstructed character LAO LETTER PALI NYA ພ by Maha Sila Viravaong. Historically, ອ represented the Indic phoneme /j/ that later shifted to become /ɲ/ in Modern Lao. Whereas, ຢ still retains the original Indic value in Modern Thai. On the other hand, ຢ which represented the Indic phoneme /ɲ/ came to represent /j/ in Modern Thai but still represents (at least in spelling) /ɲ/ in Pali texts.

In the traditional ordering of Lao consonants, ພ and ອ are placed in the palatal series next to ຈ as “ຈ ພ ຊ ອ”, analogous to the Thai ordering of “จ ฉ ช ซ ฌ ญ” (grey indicates Thai characters that don’t have Lao equivalents). During the encoding process, the false equivalence of ฉ to ພ was recognized and ພ U+0EAA was rightly placed in an equivalent position to ฌ U+0E2A. But the false equivalence of ອ was somehow missed out.

Given that Thai tends to take an etymological approach to the language, it is possible that at some point during the encoding, the modern phonetic realization of ອ as /ɲ/ and the historical/etymological realization of ຢ as /j/ were conflated. This probably resulted in the characters being marked as analogous to each other in the encoding standard.

Furthermore, the character LAO LETTER YO ອ U+0E22 that is marked as analogous to Thai ຢ is not directly related to the latter. From an etymological perspective, Lao ອ is actually related to the Thai compound ออ. Both eventually developed as alternate representations to denote the phoneme /j/ in a high tone as the default tone register for ຢ/ ອ was low.

Ideally, LAO LETTER PALI NYA ພ should get mapped to U+0E8D but this is not possible as the position is already occupied by ອ. Therefore, we propose to map PALI NYA to the next available position at U+0E8E analogous to THAI CHARACTER DO CHADA ฎ U+0E0E. It is highly unlikely that a Lao equivalent of ฎ U+0E0E would

turn up later for encoding. However, if it does, it must to be relocated somewhere else. At some point, Lao has to be considered on its own terms. Thai and Lao scripts have done different things with different characters for centuries and the two are not fully matchable.

In terms of transliteration between Thai & Lao blocks, implementers must be aware of the historical and modern equivalences among ຢ, ຣູ, ອ & ພ and choose the appropriate character mappings between them. From a purely etymological perspective, we suggest that ຣູ U+0E0D be mapped to ພ U+0E8E and ຢ U+0E22 to ອ U+0E8D.

5. Lao Unicode Block with Proposed Additions

	0E8	0E9	0EA	0EB	0EC	0ED	0EE	0EF
0		ຸ	ກ	ະ	ໄ	ອ		
1	ກ	ຫ	ນ	້	ໄໄ	ອ		
2	ຂ	ຜ	ຢ	າ	ໄ	ຮ		
3		ກ	ຮ	ໍ່າ	ໃ	ຜ		
4	ຄ	ດ		ົ	ໄ	ຜ		
5		ຕ	ລ	ົ		ຜ		
6	ຜ	ຖ		ົ	ງ	ຜ		
7	ງ	ຫ	ວ	ົ		ກ		
8	ຈ	ຜ	ຜ	ຸ	່	ຜ		
9	ຸ	ນ	ຜ	ຸ	້	ລ		
A	ຊ	ບ	ຜ	ຸ	້			
B		ປ	ຫ	ົ	່			
C	ຊ	ຜ	ຜ	ົ	້	ຫ		
D	ຢ	ຜ	ອ	ຸ	ໍ່	ຫ		
E	ຜ	ຜ	ຮ			ກ		
F	ຜ	ຜ	ຮ			ຜ		

6. Character Properties

```
0E86;LAO LETTER PALI GHA;Lo;0;L;;;;N;;;;;
0E89;LAO LETTER PALI CHA;Lo;0;L;;;;N;;;;;
0E8C;LAO LETTER PALI JHA;Lo;0;L;;;;N;;;;;
0E8E;LAO LETTER PALI NYA;Lo;0;L;;;;N;;;;;
0E8F;LAO LETTER PALI TTA;Lo;0;L;;;;N;;;;;
0E90;LAO LETTER PALI TTHA;Lo;0;L;;;;N;;;;;
0E91;LAO LETTER PALI DDA;Lo;0;L;;;;N;;;;;
0E92;LAO LETTER PALI DDHA;Lo;0;L;;;;N;;;;;
0E93;LAO LETTER PALI NNA;Lo;0;L;;;;N;;;;;
0E98;LAO LETTER PALI DHA;Lo;0;L;;;;N;;;;;
0EA0;LAO LETTER PALI BHA;Lo;0;L;;;;N;;;;;
0EA8;LAO LETTER SANSKRIT SHA;Lo;0;L;;;;N;;;;;
0EA9;LAO LETTER SANSKRIT SSA;Lo;0;L;;;;N;;;;;
0EAC;LAO LETTER PALI LLA;Lo;0;L;;;;N;;;;;
0EBA;LAO SIGN PALI VIRAMA;Mn;9;NSM;;;;N;;;;;
```

7. Indic Syllabic Category

The following should be appended to # *Indic_Syllabic_Category=Pure_Killer* :

```
0EBA ; Pure_Killer # Mn LAO SIGN PALI VIRAMA
```

The existing Lao entries under # *Indic_Syllabic_Category=Consonant* must be removed and be replaced by the below:

```
0E81..0E82 ; Consonant # Lo [2] LAO LETTER KO..LAO LETTER KHO SUNG
0E84 ; Consonant # Lo LAO LETTER KHO TAM
0E86..0E8A ; Consonant # Lo [5] LAO LETTER PALI GHA..LAO LETTER SO TAM
0E8C..0EA3 ; Consonant # Lo [24] LAO LETTER PALI JHA..LAO LETTER LO LING
0EA5 ; Consonant # Lo LAO LETTER LO LOOT
0EA7..0EAE ; Consonant # Lo [8] LAO LETTER WO..LAO LETTER HO TAM
0EDC..0EDF ; Consonant # Lo [4] LAO HO NO..LAO LETTER KHMU NYO
```

8. Indic Positional Category

The existing entry *0EB8..0EB9* under # *Indic_Positional_Category=Bottom* must be removed and replaced by the below:

```
0EB8..0EBA ; Bottom # Mn [3] LAO VOWEL SIGN U..LAO SIGN PALI VIRAMA
```


11. Attestations

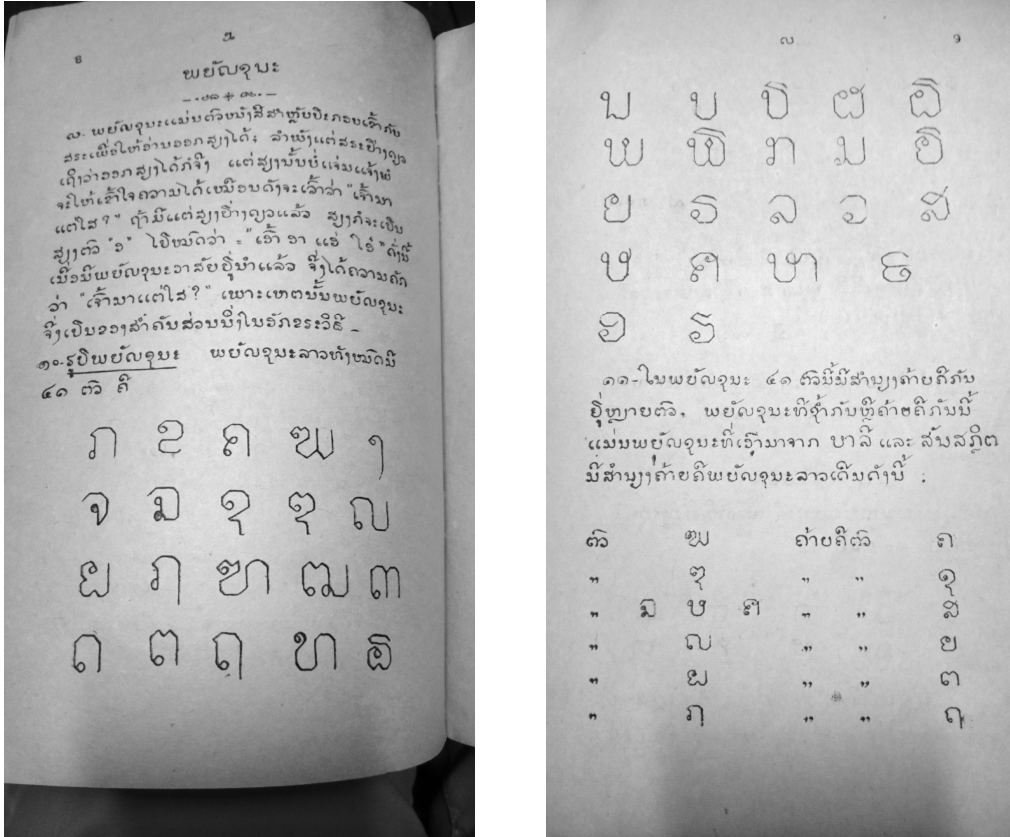


Fig 2: Consonant repertoire of Pali (Sila Viravong 1935)

ອັກສອນລາວ

ກ	ຂ	ຄ	ພ	ງ	ຈ	ຊ	ຊ	ສ	ຜ	ຖ	ຜ		
k	kh	kh	kh	ŋg	c	s	x	x	ñ, ny	t	th	th, d	th
[k]	[kʰ]	[kʰ]	[kʰ]	[ŋ]	[t͡ɕ]	[s]	[s]	[s]	[ɲ]	[t]	[tʰ]	[tʰ, d]	[tʰ]
M	H				M	H				M	H		
ດ	ດ	ຕ	ຖ	ທ	ຜ	ນ	ບ	ປ	ຜ	ຝ	ພ	ຟ	ຟ
n	d	t	th	th	th	n	b	p	ph	f	ph	f	f
[n]	[d]	[t]	[tʰ]	[tʰ]	[tʰ]	[n]	[b]	[p]	[pʰ]	[f]	[pʰ]	[f]	[f]
M	M	H				M	M	H	H				
ກ	ມ	ຍ	ຢ	ຣ	ລ	ວ	ຮ	ປ	ສ	ຫ	ຮ	ອ	ຮ
ph	m	ñ, ny	y	r	l	v, w	s	s	s	h	l	-	h
[pʰ]	[m]	[ɲ]	[j]	[r]	[l]	[v, w]	[s]	[s]	[s]	[h]	[l]	[ʔ]	[h]
			H				H	H	H	H		M	

H = high consonant, M = middle consonant, unmarked = low consonant
 red = obsolete
 note: only initial consonants are indicated

Alif Silpachai

Fig 3: Consonant repertoire from <http://alif-shinobi.blogspot.co.uk/2014/01/indic-lao-script-script-with-pali-and.html>

ຕົວອັກສອນ

ຕົວອັກສອນທີ່ໃຊ້ໃນກາສາບາລິນມີຜູ້ ໔໑ ຕົວ ແບ່ງເປັນ
ສະໝຸດ ຕົວ ພັດທະນະ ມີ ໓ ຕົວ ຕໍ່ນີ້:

ສະໝຸດ ຕົວ ຕໍ່:

ອ ອາ ອິ ອີ ອຸ ອູ ເອ ໂອ

ພັດທະນະ ມີ ໓ ຕົວ ຕໍ່:

ກະວັດຕິກໍ ກ ຂ ຄ ພ ງ

ຈະວັດຕິກໍ ຈ ຄ ຊ ຈ ລ

ບະວັດຕິກໍ ບ ກ ຂາ ຕ ຫ

ຕະວັດຕິກໍ ຕ ດ ທ ສ ນ

ປະວັດຕິກໍ ປ ຜ ພ ກ ມ

ເສສວັດຕິກໍ ບ ຣ ລ ວ ສ ຫ ສ ະ

Fig 4: Consonant repertoire

<https://www.facebook.com/photo.php?fbid=10200750935910010&set=a.10200750920749631.1073741853.1060676515&type=3&theater>










































ກ = ໄກ່ 	ຂ = ຂັນເງິນ 	ຄ = ຄົນ 	ພ = ຄໍາສະໝັ່ງ 
ງ = ງົວ 	ຈ = ຈອກ 	ສ = ສີ່ອ 	ຮ = ຮີລັບ 
ຜ = ສິກາ 	ອ = ອົມ 	ຊ = ຊິງ 	ຊ = ຊ້າງ 
ຍ = ຍົນຕ໌ 	ຝ = ຜູ້ລົງ 	ດ = ດັກນ້ອຍ 	ຫ = ບັດຫິດ 
ຕ = ຕາ 	ຝ = ຝ່າຍູກ 	ຝ = ຜູ້ເໝ້າ 	ຖ = ໄຖນາ 
ກຸ = ກຸ່ນຫີນ 	ທ = ໂທລະສັບ 	ຮ = ຮັ່ມນະນູດ 	ນ = ນົກ 
ດ = ດາວ 	ບ = ແບ້ 	ປ = ປາ 	ຜ = ເຜັ່ງ 
ຝ = ປາຝາ 	ພ = ພຣະ 	ກ = ກິກຂຸ 	ຟ = ໂຟ 
ມ = ມ້າ 	ຢ = ຢາງ 	ຮ = ຮົ່ຖ 	ລ = ໂລກ 
ຣ = ລ່ຽລາ 	ວ = ວີ 	ຫ = ຫອຍ 	ອ = ອະນຸສາວະຣີ 
ຮ = ເຮືອໃບ 	ຫລັກຖານອ້າງອີງ "ວັຈນານຸກົມລາວລ້ານຊ້າງ"		ຫລັກຖານອ້າງອີງ "ວັຈນານຸກົມ ບັ ກາບາ"

Fig 6: Consonant repertoire

<https://www.facebook.com/photo.php?fbid=10200690765285782&set=a.10200888419507014.1073741859.1060676515&type=3&theater>

ອັກສອນລາວ

(ສັບປະທັບພັນທິຕິສກາຈັນທະບູຣີ)

ພັບໂຊນະເລົ້າ					ສຣະ							
ກ ກໍ່ ໄກ່		ຂ ຂໍ່ ຂັນ		ຄ ຄໍ່ ແຄນ		ພ ພໍ່ ອະພັງ		ຸຸ	ຸຸ	ຸຸ	ຸຸ	
ງ ງໍ່ ງົວ		ຈ ຈໍ່ ຈຳປາ		ສ ສໍ່ ສີອ		ຜ ຜໍ່ ສີລປະ		ຸຸ	ຸຸ	ຸຸ	ຸຸ	
ຜ ຜໍ່ ສີກບາ		ຟ ຟໍ່ ຈິມ		ຊ ຊໍ່ ຈິງ		ຊ ຊໍ່ ຈຶ່ງ		ຸຸຸ	ຸຸ	ຸຸຸ	ຸຸຸ	
ຍ ຍໍ່ ຍັນດໍ		ລ ລໍ່ ພິງ		ດ ດໍ່ ເດັກ		ຫ ຫໍ່ ບັດທິດ		ຸຸຸ	ຸຸ	ຸຸຸຸ	ຸຸ	
ຕ ຕໍ່ ຕາ		ຝ ຝໍ່ ພ່າຫຼາກ		ຜ ຜໍ່ ຜູ້ເຜົ່າ		ຖ ຖໍ່ ໂຖນາ		ຸຸ	ຸຸ	ຸຸຸຸ	ຸຸຸຸ	
ກຸ ກຸ່ ກຸ່ນ		ທ ທໍ່ ທຽນ		ລ ລໍ່ ສົມມະ		ນ ນໍ່ ນົກ		ຸຸຸ	ຸຸ	ຸຸຸ	ຸຸຸ	
ທ ທໍ່ ທອນ		ບ ບໍ່ ບັບ		ປ ປໍ່ ປາ		ຜ ຜໍ່ ຜຶງ		ຸຸ	ຸຸ	ຸຸຸ	ຸຸຸ	
ຝ ຝໍ່ ປາປາ		ພ ພໍ່ ພອນ		ກ ກໍ່ ກິກຂຸ		ຟ ຟໍ່ ໂຟ		ຸຸ	ຸຸຸຸ	ຸຸຸຸ	ຸຸຸຸ	
ມ ມໍ່ ແມວ		ຢ ຢໍ່ ຢາ		ຣ ຣໍ່ ຣົຖ		ລ ລໍ່ ລິງ		ຕົວສັກົດ ກ (ຂຄພ) ດ (ຈສຜບອຊຊຸຕພຝກຸທຸລ) ບ (ປພກຟ) ງ ນ (ນຜຜລຮ) ມ ຍ ວ				
ຮ ຮໍ່ ກິຮາ		ວ ວໍ່ ວີ		ຫ ຫໍ່ ໂຫ		ອ ອໍ່ ໂອ						
ຮ ຮໍ່ ເຮືອນ		ພັບໂຊນະປະສົມ ຫງ ຫຍ(ຫງ) ຫນ ຫນ ຫູ ຫລ ຫວ										
ເລຂ ໐ ໑ ໒ ໓ ໔ ໕ ໖ ໗ ໘ ໙					ວັດທະຍຸຕ ຸ ຸ ຸ ຸ				ແຫລ່ງອ້າງອີງ: ວັຈນານຸກົມລາວລ້ານຊ້າງ ແຕ່ງແຕ້ມໂດຍ: ປ. ພຣະນຸນ (໒໕໕໕)			

Fig 7: Consonant repertoire using a digital font
<https://www.facebook.com/photo.php?fbid=10200944115619382&set=a.10200888419507014.1073741859.1060676515&type=3&theater>



Fig 8: Consonant repertoire using a digital font
<https://www.facebook.com/photo.php?fbid=10200888420547040&set=a.10200888419507014.1073741859.1060676515&type=3&theater>

ພັດທະນະເຈົ້າ	ຊື່ຄວາມໝາຍ	ທຽບສຽງ
ກ	ໄກ່	K
ຂ	ໄຂ່	Kh
ຄ	ຄວາຍ	C
ພ	ຄ ຣະພັງ	Kh
ງ	ງ	Ng
ຈ	ຈອກ	Ch
ສ	ເສືອ	S
ຮ	ສ ສາລາ	S
ຜ	ສ ຣີບີ	S
ອ	ສ ຈິ່ງ	S
ຊ	ຊ້າງ	Z
ຊ	ຊ (ໄຊ້)	Zh
ຍ	ຍັກຂໍ(ຍັກບໍ່)	Y
ຝ	ຍ ສັພພັນລູ	Yh
ດ	ເດັກ	D
ທ	ດ ນາງນຸນໂຫ	Dh
ຕ	ຕາ	T
ຕ	ຕ ປັກ	Th
ຝ	ຝຶງ	Th
ຖ	ຖ ປົກ	Th
ຖ	ຖ ຜູ່ເຜົາ	Th
ຝ		Th
ທ	ທຫາອ	

Fig 9: Consonant repertoire (Phanlak, 2012)

ສ	ທ ສຸງ	Th
ນ	ນິກ	N
ຜ	ນ ເຕຣ	Nh
ບ	ແປ້	B
ປ	ປາ	P
ຜ	ຜັງ	Ph
ຝ	ຝິນ	Fh
ພ	ພານ, ພູຄັງວໝາກ	Ph
ກ	ພ ກຄວາ, ກູຜາ	Ph
ຟ	ໄຟ	F
ມ	ແມວ	M
ຢ	ຢາ	Yh
ຣ	ຣາຊສີຫໍ່	R
ລ	ລິງ	L
ຮ	ລ ກິຮາ	Lh
ວ	ວິ	V
ຫ	ຫ່ານ	Hh
ອ	ໄອ	Or
ຮ	ເຮືອນ	H

ໝາຍເຫດ:

໑. ພັດທະນະທັມຈາກກາສາປາລີມີ ມີ ໂຕ ແລະ ສະ ຕູ
ໂຕ ລວມເປັນ ໔໑ ໂຕດັ່ງນີ້:
ອ ອາ ອິ ອີ ອ ອ

Fig 10: Consonant repertoire (Phanlak, 2012)

ເຈຕະສິກາ ລັມມາ ຈິຕຕະສະມຸຢຸຖານານັງ ຣູປານັງ ອະວິຄະຕະປັຈຈະ
ເຍນະ ປັຈຈະໂຍ ຯ ຈັກຂວາຍະຕະນັງ ຈັກຂຸວິນລາຕະລາຕຸຍາ ຕັງ
ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ,
ໂສຕາຍະຕະນັງ ໂສຕະວິນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕຕະການິນຈະ
ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ, ພານາຍະຕະນັງ ພາ
ນະວິນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ ອະວິຄະ
ຕະປັຈຈະເຍນະ ປັຈຈະໂຍ, ຊິວຫາຍະຕະນັງ ຊິວຫາວິນລາຕະລາຕຸ
ຍາ ຕັງ ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ
ປັຈຈະໂຍ, ກາຍາຍະຕະນັງ ກາຍະວິນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕ
ຕະການິນຈະ ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ, ຣູປາຍະ
ຕະນັງ ຈັກຂຸວິນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕຕະການິນຈະ ລັມມາ
ນັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ, ສັທລາຍະຕະນັງ ໂສຕະວິ
ນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ ອະວິຄະຕະປັຈ
ຈະເຍນະ ປັຈຈະໂຍ, ພັນລາຍະຕະນັງ ພານະວິນລາຕະລາຕຸຍາ ຕັງ
ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ,
ຣະສາຍະຕະນັງ ຊິວຫາວິນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕຕະການິນ
ຈະ ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ, ໂຜຣຸຖັພພາຍະຕະ
ນັງ ກາຍະວິນລາຕະລາຕຸຍາ ຕັງ ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ
ອະວິຄະຕະປັຈຈະເຍນະ ປັຈຈະໂຍ, ຣູປາຍະຕະນັງ ສັທລາຍະຕະນັງ
ພັນລາຍະຕະນັງ ຣະສາຍະຕະນັງ ໂຜຣຸຖັພພາຍະຕະນັງ ມະໂນລາຕຸຍາ
ຕັງ ສັມປະຍຸຕຕະການິນຈະ ລັມມານັງ ອະວິຄະຕະປັຈຈະເຍນະ ປັຈ
ຈະໂຍ ຯ ຍັງ ຣູປັງ ນິສສະຍະ ມະໂນລາຕຸ ຈະ ມະໂນວິນລາຕະລາ
ຕຸຈະ ວັຕຕັນຕິ ຕັງ ຣູປັງ ມະໂນລາຕຸຈະ ມະໂນວິນລາຕະລາຕຸຍາ

Fig 11: Pali text in extended Lao with Lao-like vowel notation (Suriyo, 2003)

Texte Pāli Tham

ສ ທຸ ອ ທິ ພ ມ ບ ຸ ຣິ ສ ສ ສ ັ ຫ ັ ທິ
 ໂ ໄ ເ ຈ ອ ສ ສ ທ ັ ພ ັ ມ ວ ທິ ມ ຸ ດ ັ ທິ
 ຍ ໂ ສ ຈ ກິ ຕ ັ ທິ ຈ ຕ ຕ ວ ສ ສ ັ ຫ ັ ທິ
 ສ ຄ ດ ຕ ັ ຈ ໂ ສ ອ ຈ ັ ທິ ສ ຣິ ຣິ ັ ປ ຫ ັ ພ ັ ທິ
 ສ ຣ ັ ັ ຊ ັ ຫ ັ ວິ ປ ັ ປ ຊ ັ ຫ ັ ພ ັ ຍ ຍ ມ ັ ນ ັ ມ
 ສ ັ ນ ັ ນ ັ ອ ຈ ັ ນ ັ ມ ສ ັ ບ ບ ັ ມ ັ ທິ ກ ກ ັ ມ ັ ມ ັ ຍ ຍ
 ຕ ັ ມ ນ ັ ມ ັ ຣ ຸ ປ ັ ສ ມິ ນ ສ ຊ ຊ ມ ັ ນ ັ ມ
 ອ ກິ ນ ັ ຈ ັ ນ ັ ມ ນ ັ ນ ຸ ປ ັ ຕ ັ ນ ັ ທິ ສ ັ ນ ັ ທິ ັ

Texte Pāli Lao

ສ ທ ຸ ອ ທ ັ ພ ມ ບ ຸ ຣິ ສ ສ ສ ັ ຫ ັ ທິ
 ໂ ໄ ເ ຈ ອ ສ ສ ທ ັ ພ ັ ມ ວ ທິ ມ ຸ ດ ັ ທິ
 ຍ ໂ ສ ຈ ກິ ຕ ັ ທິ ຈ ຕ ຕ ວ ສ ສ ັ ຫ ັ ທິ
 ສ ຄ ດ ຕ ັ ຈ ໂ ສ ອ ຈ ັ ທິ ສ ຣິ ຣິ ັ ປ ຫ ັ ພ ັ ທິ
 ສ ຣ ັ ັ ຊ ັ ຫ ັ ວິ ປ ັ ປ ຊ ັ ຫ ັ ພ ັ ຍ ຍ ມ ັ ນ ັ ມ
 ສ ັ ນ ັ ນ ັ ອ ຈ ັ ນ ັ ມ ສ ັ ບ ບ ັ ມ ັ ທິ ກ ກ ັ ມ ັ ມ ັ ຍ ຍ
 ຕ ັ ມ ນ ັ ມ ັ ຣ ຸ ປ ັ ສ ມິ ນ ສ ຊ ຊ ມ ັ ນ ັ ມ
 ອ ກິ ນ ັ ຈ ັ ນ ັ ມ ນ ັ ນ ຸ ປ ັ ຕ ັ ນ ັ ທິ ສ ັ ນ ັ ທິ ັ

Fig 12: Pali text in Tham and extended Lao script using Pali Virama (Lafont, 1962)

From *Saddhā Sutta* in the *Saṃyuttanikāya*, the text reads:

saddhā dutiyā purisassa hoti	ສ ທ ຸ ອ ທ ັ ພ ມ ບ ຸ ຣິ ສ ສ ສ ັ ຫ ັ ທິ
no ce assaddhiyamavati ^t thati	ໂ ໄ ເ ຈ ອ ສ ສ ທ ັ ພ ັ ມ ວ ທິ ມ ຸ ດ ັ ທິ
yaso ca kittī ca tatvassa hoti	ຍ ໂ ສ ຈ ກິ ຕ ັ ທິ ຈ ຕ ຕ ວ ສ ສ ັ ຫ ັ ທິ
saggaṇ ^[1] ca so gacchati sarīraṃ pahāyā ti	ສ ຄ ດ ຕ ັ ຈ ໂ ສ ອ ຈ ັ ທິ ສ ຣິ ຣິ ັ ປ ຫ ັ ພ ັ ທິ
kodhaṃ jahe vippajaheyya mānaṃ	ໂ ກ ັ ັ ຊ ັ ຫ ັ ວິ ປ ັ ປ ຊ ັ ຫ ັ ພ ັ ຍ ຍ ມ ັ ນ ັ ມ
saññojanaṃ sabbamatikkameyya	ສ ັ ນ ັ ນ ັ ອ ຈ ັ ນ ັ ມ ສ ັ ບ ບ ັ ມ ັ ທິ ກ ກ ັ ມ ັ ມ ັ ຍ ຍ
taṃ nāmarūpasminasajjamānaṃ	ຕ ັ ມ ນ ັ ມ ັ ຣ ຸ ປ ັ ສ ມິ ນ ສ ຊ ຊ ມ ັ ນ ັ ມ
akiñcanaṃ ^[2] nānupatanti saṅgā ti	ອ ກິ ນ ັ ຈ ັ ນ ັ ມ ນ ັ ນ ຸ ປ ັ ຕ ັ ນ ັ ທິ ສ ັ ນ ັ ທິ ັ

[1] should read saggañ [2] should read akiñcanaṃ



Fig 13: Buddhist triple refuge formula in Tham and extended Lao scripts

<https://www.facebook.com/laosongfang/photos/a.363997627026327.86552.255417617884329/363997653692991/?type=1&theater>

VIRAVONG, S. (MAHA SILA) (1937), ບາລິເຢຍາກອດ ອັກຂະວິຣິກາຄທິນິງ ສັມມາທິສານແລະສັນສິ, [Pali Grammar, Akkharavidhi, Part one, Samaññābhidāna - Sandhi], Vientiane, Institut Bouddhique, 68 p.

Fig 14: <http://www.unicode.org/L2/L2005/05166-dekalb-gk-vb.pdf>

References

1. Diller, A. (1996). Thai and Lao writing. *The world's writing systems*, 457-466.
2. Enfield, N. J. (2007). *A grammar of Lao* (Vol. 38). Walter de Gruyter.
3. Amaravati: Abode of Amritas. Retrieved January 08, 2017, from <http://www.amritas.com/100410.htm#04102320>
4. Kourilsky, G. J. D. (2006). *L'écriture tham du Laos, Rencontre du sacré et de la technologie: Éléments historiques, linguistiques, sociologiques et pratiques pour l'informatisation d'une écriture bouddhique majeure d'Asie du Sud-Est*. Paris: Cahiers de Péninsule.
5. Sila Viravong, M (1935). *Waiyākō'n Lāō, akkhara withī*. Vientiane: Buddhist Institute.
6. Suriyo, S (2003). *Matika Abhidhamma Nhamaka*.
7. Phanlak, S (2012). *Phasa Lan Xang*.
8. Lafont, P. B. (1962). Les écritures du Pāli au Laos. *Bulletin de l'École française d'Extrême-Orient*, 50(2), 395-405.

**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: *Proposal to Encode Lao Characters for Pali*

2. Requester's name: *Vinodh Rajan, Ben Mitchell, Martin Jansche, Sascha Brawer*

3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*

4. Submission date: *20th July 2017*

5. Requester's reference (if applicable):

6. Choose one of the following:

This is a complete proposal: *Yes*

(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) *No*

Proposed name of script:

b. The proposal is for addition of character(s) to an existing block: *Yes*

Name of the existing block: *Lao*

2. Number of characters in proposal: *15*

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary B.1-Specialized (small collection) B.2-Specialized (large collection) *Yes*

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. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? *Ben Mitchell*

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): *Ben Mitchell*

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)? *Yes*

Sorting, Transliteration (Sections 4.1 & 10)

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)

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>Alif Silpachai</i> <i>Contacted through e-mail</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
4. The context of use for the proposed characters (type of use; common or rare) Reference:	Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	Yes <i>Laos. See Section 2 & Phanlak (2012)</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 <i>See Section 2</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	
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:	Yes <i>See Section 9</i>
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
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