

Notes on Lao Character Names

L2/05-357

Andrew West
2005-10-30

1. Background

As has recently been discussed on the Unicode public mailing list, there are some serious errors in the names of several characters in the Lao block [0E80..0EFF]. Based on input received from several experts in the Lao script, this document attempts to summarise what the problems are, and recommends possible actions to clarify the identification of Lao character names.

2. Consonant Names

The Unicode names for Lao consonants are based on the syllabic pronunciation of the character (i.e. consonant plus inherent vowel). All consonants belong to one of three tone classes: high, mid and low. Where two letters are only distinguished phonetically by their tone class, the modifiers **SUNG** "high" and **TAM** "low" are used to indicate the tone class of the letter (e.g. 0E82 **LAO LETTER KHO SUNG** and 0E84 **LAO LETTER KHO TAM**). However, the Unicode names for two of the consonants have the wrong tone class applied to them:

U+0E9D ະ **LAO LETTER FO TAM** is a high tone class letter, and should have been named **LAO LETTER FO SUNG**

U+0E9F ັ **LAO LETTER FO SUNG** is a low tone class letter, and should have been named **LAO LETTER FO TAM**

Whilst the Unicode names for 25 of the 27 consonants use this naming scheme, the names of two of the consonants use mnemonic names (presumably because they share the same vowel and tone class, and so could not otherwise be differentiated). Mnemonic names are how the consonants are normally identified in the Lao language, although there is no official list of standard mnemonic names for consonants, and different sources may use different mnemonic names for some letters. The most common mnemonic names are listed in Table 1.

The two letters whose Unicode names are based on mnemonic names are:

U+0EA3 ຣ **LAO LETTER LO LING**

U+0EA5 ອ **LAO LETTER LO LOOT**

The mnemonic names for these two letters are the wrong way round. U+0EA5 is the normal letter [l] and is universally identified by the mnemonic name **lo ling** "lo as in ling [monkey]". On the other hand, U+0EA3 is a letter that is used to represent [r] in foreign words; however this letter has been officially deprecated by the Lao government since 1975, and is no longer in common use. The name element LO LOOT applied to U+0EA5 would seem to represent the mnemonic **ro rot**, "rot" meaning automobile, that should be applied to U+0EA3.

Table 1 : Lao Consonants

Code Point	Character	Unicode Name	Tone Class	Mnemonic Name
0E81	ກ	LAO LETTER KO	mid	ko kay ໄກ່ (chicken)
0E82	ຂ	LAO LETTER KHO SUNG	high	kho khay ໄຂ່ (egg)
0E84	ຄ	LAO LETTER KHO TAM	low	kho khuay ຄວາຍ (buffalo)
0E87	ງ	LAO LETTER NGO	low	ngo ngu ງູ (snake) ngo ngua ງົວ (cow)
0E88	ຈ	LAO LETTER CO	mid	co cok ຈອກ (cup, glass) co cua ຈົວ (buddhist novice)
0E8A	ຊ	LAO LETTER SO TAM	low	so sang ຊ້າງ (elephant)
0E8D	ຍ	LAO LETTER NYO	low	nyo nyung ຍຸງ (mosquito)
0E94	ດ	LAO LETTER DO	mid	do dek ໄດ້ກ (child)
0E95	ຕ	LAO LETTER TO	mid	to ta ຕາ (eye)
0E96	ຖ	LAO LETTER THO SUNG	high	tho thong ຖົງ (bag) tho thung [?] (to ask)
0E97	ທ	LAO LETTER THO TAM	low	tho thung ທຸງ (flag)
0E99	ນ	LAO LETTER NO	low	no nok ນົກ (bird)

0E9A	ບ	LAO LETTER BO	mid	bo be ແບ້ (goat) bo bet ເບັດ (fishhook)
0E9B	ປ	LAO LETTER PO	mid	po pa ປາ (fish)
0E9C	ຜ	LAO LETTER PHO SUNG	high	pho pheng ເຜີ້ງ (bee)
0E9D	ຝ	LAO LETTER FO TAM	high	fo fon ຝົນ (rain) fo fa ຝາ (wall)
0E9E	ພ	LAO LETTER PHO TAM	low	pho phu ພູ (mountain)
0E9F	ຟ	LAO LETTER FO SUNG	low	fo fay ໄຟ (fire)
0EA1	ມ	LAO LETTER MO	low	mo mew ແມວ (cat) mo ma ມ້າ (horse)
0EA2	ຢ	LAO LETTER YO	mid	yo ya ຢາ (medicine or tobacco)
0EA3	ຣ	LAO LETTER LO LING	low	ro rot ຣົດ (car)
0EA5	ລ	LAO LETTER LO LOOT	low	lo ling ລິງ (monkey)
0EA7	ວ	LAO LETTER WO	low	wo wi ວີ (fan)
0EAA	ສ	LAO LETTER SO SUNG	high	so sya ເສືອ (tiger)
0EAB	ຫ	LAO LETTER HO SUNG	high	ho hay ໄຫ (jar) ho han ຫ່ານ (goose)

0EAD	ອ	LAO LETTER O	mid	ອ ໂອ (bowl) ອ ອວ [ʔ] ເອົາ (to take)
0EAE	ຮ	LAO LETTER HO TAM	low	ຮ ຮາ ເຮືອ (boat) ຮ ຮານ ເຮືອນ (house)

3. Vowel Names

Table 2 : Lao Vowels

Code Point	Character	Unicode Name	Traditional Name
0EB0	ຳ	LAO VOWEL SIGN A	sala a
0EB1	ຳ̣	LAO VOWEL SIGN MAI KAN	mai kan
0EB2	ຳ̆	LAO VOWEL SIGN AA	sala aa
0EB3	ຳ̆̄	LAO VOWEL SIGN AM	sala am
0EB4	ຳ̆̇	LAO VOWEL SIGN I	sala i
0EB5	ຳ̆̈	LAO VOWEL SIGN II	sala ii
0EB6	ຳ̆̉	LAO VOWEL SIGN Y	sala y
0EB7	ຳ̆̊	LAO VOWEL SIGN YY	sala yy
0EB8	ຳ̆̋	LAO VOWEL SIGN U	sala u
0EB9	ຳ̆̌	LAO VOWEL SIGN UU	sala uu

0EBB	◌ [◌]	LAO VOWEL SIGN MAI KON	mai kong
0EBC	◌ [◌]	LAO SEMIVOWEL SIGN LO	
0EBD	◌ [◌]	LAO SEMIVOWEL SIGN NYO	nyo fyang
0EC0	◌ [◌]	LAO VOWEL SIGN E	sala e
0EC1	◌ [◌]	LAO VOWEL SIGN EI	sala ei
0EC2	◌ [◌]	LAO VOWEL SIGN O	sala o
0EC3	◌ [◌]	LAO VOWEL SIGN AY	mai muan
0EC4	◌ [◌]	LAO VOWEL SIGN AI	mai may
0ECD	◌ [◌]	LAO NIGGAHITA	sala oo

There are no serious problems with the Unicode names of Lao vowels, although they do not entirely correspond to the traditional names for the vowels, which are mostly formed from the word **sala** "vowel" followed by the vowel in question (cf. Thai where the Unicode names for the vowels are formed from the word SARA followed by the vowel). In addition, several vowel signs have special names, including U+0EBB **LAO VOWEL SIGN MAI KON** which should actually have been named **LAO VOWEL SIGN MAI KONG**.

The one character which is somewhat problematic is U+0ECD **LAO NIGGAHITA**. In the Unicode code charts this character is included under the sub-heading "Signs", and annotated as "final nasal". Although it may originally have been used to represent a final nasal when writing Pali (equivalent to the anusvara of Sanskrit), its function in modern Lao is quite different. By itself this sign is used to represent a long [ɔ̃] in open syllables, for example ບໍ່ **bo** "not" and ສີ່ **so** "pencil", and as such it acts as a vowel sign. That U+0ECD is used as a vowel sign is also shown in Table 10-2 of the Unicode Standard. However, when used in combination with U+0EB2 **LAO VOWEL SIGN AA**, it does act as a final nasal, for example ນໍ້າ **nam** "to lead" and ລໍ້າ **lam** "to sing".

4. Recommendations

4.1 Source for Lao Encoding

The Unicode code charts note that the Lao block is "Based on TIS 620-2529". This statement is misleading as TIS 620-2529 is a Thai standard for representing the Thai script in an 8-bit code, and does not define names or code points for the Lao script. The Unicode Lao block is based on a mapping of Lao characters to the equivalent Thai characters in TIS 620, but is not actually based on this standard. I would therefore recommend removing this statement from the Unicode Standard.

4.2 Formal Aliases

A solution to the problem of perniciously misnamed Unicode characters that has been suggested is to devise formal aliases for the characters in question. This would only apply to cases where the misnaming is so extreme as to cause probable misidentification of the character, not to cases where there is merely academic dispute as to the best name for the character. For the Lao block, the names of two pairs of characters have been inadvertently swapped, which means that in isolation (i.e. without sight of the corresponding glyphs) the characters would be misidentified by anyone familiar with the Lao script and language. Therefore, if the UTC does adopt a system of formal aliases for perniciously misnamed characters, I would suggest that the four Lao characters with misapplied names be given formal aliases, as shown in Table 3.

Table 3 : Suggested Formal Aliases

Code Point	Character Name	Suggested Formal Alias
0E9D	LAO LETTER FO TAM	LAO LETTER FO FON
0E9F	LAO LETTER FO SUNG	LAO LETTER FO FAY
0EA3	LAO LETTER LO LING	LAO LETTER RO
0EA5	LAO LETTER LO LOOT	LAO LETTER LO

4.3 Annotations and Aliases

In order to improve the identification of Lao characters, I suggest that the mnemonic names of the consonants and special names of vowels be added as aliases to the Unicode code charts, as shown below. I also suggest that the fact that the names for 0E9D, 0E9F, 0EA3 and 0EA5 are mistakes be explicitly annotated. Note that there is no standard system for romanising Lao, and so the transcriptions of the Lao names suggested here may not be ideal.

Suggested Annotations and Aliases

```

@@      0E80      Lao      0EFF
@----- Based on TIS 620-2529
@              Consonants
0E81     LAO LETTER KO
          = ko kay
0E82     LAO LETTER KHO SUNG
          = kho khay
0E84     LAO LETTER KHO TAM
          = kho khuay
0E87     LAO LETTER NGO
          = ngo ngu, ngo ngua
0E88     LAO LETTER CO
          = co cok, co cua
0E8A     LAO LETTER SO TAM
          = so sang
0E8D     LAO LETTER NYO
          = nyo nyung

```

0E94 LAO LETTER DO
 = do dek
 0E95 LAO LETTER TO
 = to ta
 0E96 LAO LETTER THO SUNG
 = tho thong
 0E97 LAO LETTER THO TAM
 = tho thung
 0E99 LAO LETTER NO
 = no nok
 0E9A LAO LETTER BO
 = bo be, bo bet
 0E9B LAO LETTER PO
 = po pa
 0E9C LAO LETTER PHO SUNG
 = pho pheng
 0E9D LAO LETTER FO TAM
 * name is a mistake for FO SUNG
 = fo fon, fo fa
 0E9E LAO LETTER PHO TAM
 = pho phu
 0E9F LAO LETTER FO SUNG
 * name is a mistake for FO TAM
 = fo fay
 0EA1 LAO LETTER MO
 = mo mew, mo ma
 0EA2 LAO LETTER YO
 = yo ya
 0EA3 LAO LETTER LO LING
 * name is a mistake, LO LING is the mnemonic for 0EA5
 = ro rot
 0EA5 LAO LETTER LO LOOT
 * name is a mistake, LO LOOT is the mnemonic for 0EA3
 = lo ling
 0EA7 LAO LETTER WO
 = wo wi
 0EAA LAO LETTER SO SUNG
 = so sya
 0EAB LAO LETTER HO SUNG
 = ho hay, ho han
 0EAD LAO LETTER O
 = o o
 0EAE LAO LETTER HO TAM
 @ Sign
 0EAF LAO ELLIPSIS
 @ Vowels
 0EB0 LAO VOWEL SIGN A
 0EB1 LAO VOWEL SIGN MAI KAN
 * vowel shortener
 0EB2 LAO VOWEL SIGN AA
 0EB3 LAO VOWEL SIGN AM
 # OECD 0EB2
 0EB4 LAO VOWEL SIGN I
 0EB5 LAO VOWEL SIGN II
 0EB6 LAO VOWEL SIGN Y
 0EB7 LAO VOWEL SIGN YY
 0EB8 LAO VOWEL SIGN U
 0EB9 LAO VOWEL SIGN UU
 0EBB LAO VOWEL SIGN MAI KON
 = mai kong
 @ Signs
 0EBC LAO SEMIVOWEL SIGN LO
 0EBD LAO SEMIVOWEL SIGN NYO
 = nyo fyang

@	Vowels
0EC0	LAO VOWEL SIGN E
0EC1	LAO VOWEL SIGN EI
0EC2	LAO VOWEL SIGN O
0EC3	LAO VOWEL SIGN AY = mai muan
0EC4	LAO VOWEL SIGN AI = mai may
@	Sign
0EC6	LAO KO LA * repetition
@	Tone marks
0EC8	LAO TONE MAI EK
0EC9	LAO TONE MAI THO
0ECA	LAO TONE MAI TI
0ECB	LAO TONE MAI CATAWA
@	Signs
0ECC	LAO CANCELLATION MARK
0ECD	LAO NIGGAHITA * final nasal or long o vowel
@	Digits
0ED0	LAO DIGIT ZERO
0ED1	LAO DIGIT ONE
0ED2	LAO DIGIT TWO
0ED3	LAO DIGIT THREE
0ED4	LAO DIGIT FOUR
0ED5	LAO DIGIT FIVE
0ED6	LAO DIGIT SIX
0ED7	LAO DIGIT SEVEN
0ED8	LAO DIGIT EIGHT
0ED9	LAO DIGIT NINE
@	Digraphs
0EDC	LAO HO NO # 0EAB 0E99
0EDD	LAO HO MO # 0EAB 0EA1

5. Errors in the Unicode Standard

Table 10-2 of the Unicode Standard has the following errors:

ku' ກີ U+0E81 U+0EB5 – should be U+0E81 U+0EB6

ku' : ກີ U+0E81 U+0EB6 – should be U+0E81 U+0EB7

6. Web Resources

- <http://www.omniglot.com/writing/lao.htm>
- http://www.seasite.niu.edu/lao/LaoLanguage/LaoAlphabet/lao_consonants_poster.htm
- http://www.seasite.niu.edu/lao/LaoLanguage/LaoWritingSystem/alpha_body.htm
- <http://www.seasite.niu.edu/lao/lao1.htm>
- <http://www.seasite.niu.edu/lao/lao2.htm>
- <http://www.seasite.niu.edu/lao/lao3.htm>
- <http://www.thailao.net/laoalph.htm>

- <http://www.thailao.net/laovowel.htm>
 - <http://www.thailao.net/laochart.htm>
 - <http://www.cs.colorado.edu/~main/cs1300/emacs/lisp/language/lao-util.elc>
 - <http://www.undplao.org/UNV%20Site/Laos/Lao%20PDR.pdf>
-

7. Acknowledgments

My thanks to the following people who provided valuable input into this document. Of course any errors are mine, and my opinions do not necessarily reflect those of the undernamed.

- Kent Spielmann
 - Fraser Bennett
 - Marc Durdin
 - John M. Durdin
 - Houmphanh Thongvilu
-