

Lao Collation

Phonpasit Phissamay, Nadir Durrani

STEA¹, CRULP²

phonpasit@stea.gov.la¹, nadirdurrani@yahoo.com²

Abstract

Lao has a complex case of collation. It is highly dependant on syllable structure. Same character has different sorting weight based on its position in a syllable. Moreover the input string goes through process of contextual contraction and reordering before it is finally sorted This paper discusses two collation process for two popularly used Lao dictionaries. One is based on Combinational consonant approach while other is based on alphabetical approach.

1. Introduction

Lao collation is a complex phenomenon thoroughly based on an intricate syllable structure therefore before we can actually proceed we need to segment the word into syllables. This has to be done through a syllabification module which uses a procedure, a set of conditions to find out where a syllable boundary exists. To further complicate the process a same Lao character can act hybrid showing different behavior as a main consonant, vowel and a consonantal depending upon its position in the syllable. As it takes different position it exhibits different role in collation depending upon it being treated as a consonant, consonantal or a vowel. This is not it, to further aggravate the situation a Lao character needs to be reordered once it is syllabified because the collation is based on the syllable structure and not the key press order. Generally each syllable has exactly one nuclear consonant, can have one ore more vowels, tone mark is optional and can have consonantal optionally. But again this depends upon the set of conditions defined to find syllable segments.

The other possible solution that exists is to get the purpose fulfilled using services of lexicon so it is a trade-off between complexity and efficiency. This

document aims on discussing the complex way of doing it.

2. Lexicographic Standards

Unfortunately, there aren't any national lexicographic standards defined for Lao language therefore different dictionaries display different sorting sequences. Based school books, different lexicons and dictionaries generally agree that Lao language has four levels of collation.

- Nuclear Consonant and Lao Digits
- Vowels
- Consonantal
- Tone Mark
- Punctuation marks and some other Lao characters are ignorable at all levels.

Most popular dictionaries that prevail generally agree on consonants, difference lies between the vowel ordering, vowel combinations and the consonants having combining consonants. We discuss below two popular approaches based on Unicode collation algorithm.

3. Based on Combinational Consonants and Vowels

3.1. Nuclear and Combinational Consonants and Lao Digit

Lao sorting is primarily done on base of nuclear consonant ordering as given below:

ກ ຂ ຄ ງ ຈ ສ ຊ ຢ ດ ຕ ຖ ທ ນ ບ ປ ຜ ຝ ພ ັ ມ ຢ
ຮ ູ ລ ວ ຫ ທງ ຫຍ ຫນ ຫມ ຫມ ຫລ ຫຼ ຫວ ອ ຮ

As can be observed that first 25 of 36 are single nuclear consonants, same is the case with 35th and

¹ Science Technology and Environment Agency of Lao PDR

² Center for Research in Urdu Language Processing

the vowels; vowels have been assigned weights as per their position given in the table below:

X0	X1	X	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	ຫ	ກ01 ຂ02 ຄ03 ງ04	ຊ1	ຊ1	ຊ1	ຊ1	ວ1	ຊ1	ກ1	ຈ1	ຮ1
X2		ຈ05 ຄ06 ຂ07 ຄ08	ສ2	ຊ2	ຊ2	ຊ2	ອ2	ຊ2	ງ2	ສ2	ຮ2
X3		ດ09 ຕ10 ຖ11 ທ12	ດ3		ຊ3	ຊ3	ຊ3	ຊ3	ຍ3	ຊ3	ຊ3
X4		ນ13 ບ14 ປ15 ຜ16	ລ4		ຊ4	ຊ4		ດ4	ພ4		
X5		ຜ17 ພ18 ພ19 ມ20			ຊ5	ຊ5		ນ5	ພ5		
		ຜ21 ສ22 ມ23 ວ24			ຊ6	ຊ6		ມ6	ລ6		
		ຫ25 ອ26 ສ27 ຫ28			ຊ7			ປ7			
		ຫ29						ອ8			

Categories X0, X3, X4, X6 and X7 all contain vowels. So we assign them code as they appear in the table i.e. ຕ will have minimum weight and ອ has maximum weight amongst the vowels. Followed by vowels we assign weight to consonantal X8/X9 then the tone marks X5 and finally the X10 category. For the words having multiple syllables we need to store this information for each syllable separately such that each word can reference to its syllables.

6.3. Step-II (Re-ordering)

Now that we have syllable boundaries and all the required information we need to reorder the input based on the weights we assigned them during segmentation process. To be more precise the syllable is reordered as 'X' 'X1' 'X2' (Consonants), 'X0' 'X3' 'X4' 'X6' 'X7' (Vowels), 'X8' 'X9' (Consonantal), 'X5' (Tone Marks) and finally 'X10' (Special characters).

Consequently after this step the input is ordered as:

- ນ ອ ນ
- ນ ອ ົ
- ນ ະ
- ກ າ ງ (New Syllable) ຮ ື ນ ົ
- ກ າ ງ (New Syllable) ລ ຫ າ ວ
- ນ ະ ື ນ ົ
- ນ ື ົ
- ນ ະ

As you can see that input has been reordered, but only within the syllable that it belongs. If a word contains more than one syllable reordering is done for each of them separately without involving characters from neighboring syllables.

6.4. Step-III (Form Collation Element Array):

As soon as we are done with reordering each syllable we are ready to form collation element array. This is the step where two approaches take different routes and finally rejoin each other in next step. We have to traverse through each word and within each word every syllable to assign each Lao character or sequence of Lao characters a collation Element that we have already discussed above. Procedure is fairly complex and is the heart of Lao collation. Let see how the two approaches discussed above form collation element array differently then each other. Differences have been highlighted.

Combinational Consonants and Vowel	Based on Alphabetical Order
1. ນ ອ ນ 1. ນ [085c.0200.0020.0002] 2. ນ-ອ+X8:9 [0000.0269.0020.0002] 3. ນ [0000.0000.004S.0002]	1. ນ ອ ນ 1. ນ [085c.0200.0020.0002] 2. ອ [0000.025F.0020.0002] 3. ນ [0000.0000.004S.0002]
2. ນ ອ ົ 1. ນ [085c.0200.0020.0002] 2. ອ [0000.020F.0020.0002] 3. ົ [0000.0000.0000.000C]	2. ນ ອ ົ 1. ນ [085c.0200.0020.0002] 2. ອ [0000.020A.0020.0002] 3. ົ [0000.0000.0000.000C]
3. ນ ະ 1. ນ [085c.0200.0020.0002] 2. ະ or ະ [0000.0232.0020.0002]	3. ນ ະ 1. ນ [085c.0200.0020.0002] 2. ະ [0000.022D.0020.0002] 3. ະ [0000.0205.0020.0002]
4. ກ າ ງ (New Syllable) ຮ ື ນ ົ 4a. ກ າ ງ (First Syllable) 1. ກ [0820.0200.0020.0002] 2. າ [0000.020F.0020.0002] 3. ງ [0000.0000.002A.0002]	4. ກ າ ງ (New Syllable) ຮ ື ນ ົ 4a. ກ າ ງ (First Syllable) 1. ກ [0820.0200.0020.0002] 2. າ [0000.020A.0020.0002] 3. ງ [0000.0000.002A.0002]
4b. ຮ ື ນ ົ (Second Syllable) 1. ຮ [08C0.0200.0020.0002] 2. ື [0000.0255.0020.0002] 3. ນ [0000.0000.004S.0002] 4. ົ [0000.0000.0000.0007]	4b. ຮ ື ນ ົ (Second Syllable) 1. ຮ [08C0.0200.0020.0002] 2. ື [0000.0250.0020.0002] 3. ນ [0000.0000.004S.0002] 4. ົ [0000.0000.0000.0007]
5. ກ າ ງ (New Syllable) ຫ ລ າ ວ	5. ກ າ ງ (New Syllable) ຫ ລ າ ວ

<p>5a. ກາງ (First Syllable)</p> <p>1. ກ [0820.0200.0020.0002]</p> <p>2. ກ [0000.020F.0020.0002]</p> <p>3. ຕ [0000.0000.002A.0002]</p>	<p>5a. ກາງ (First Syllable)</p> <p>1. ກ [0820.0200.0020.0002]</p> <p>2. ກ [0000.020A.0020.0002]</p> <p>3. ຕ [0000.0000.002A.0002]</p>
<p>5b. ຫລາວ (Second Syllable)</p> <p>1. ຫລ [08B1.0200.0020.0002]</p> <p>2. ກ [0000.020F.0020.0002]</p> <p>3. ດ [0000.0000.0066.0002]</p>	<p>5b. ຫລາວ (Second Syllable)</p> <p>1. ຫ [088E.0200.0020.0002]</p> <p>2. ຫ [0898.0200.0020.0002]</p> <p>3. ກ [0000.020A.0020.0002]</p> <p>4. ດ [0000.0000.0066.0002]</p>
<p>6. ນ ເ ສິ ນ ຕ</p> <p>1. ນ [085c.0200.0020.0002]</p> <p>2. ເ-X8:9 [0000.0237.0020.0002]</p> <p>3. ນ [0000.0000.0048.0002]</p> <p>4. ສ [0000.0000.0000.0002]</p>	<p>6. ນ ເ ສິ ນ ຕ</p> <p>1. ນ [085c.0200.0020.0002]</p> <p>2. ເ [0000.022D.0020.0002]</p> <p>3. ສ [0000.024B.0020.0002]</p> <p>4. ນ [0000.0000.0048.0002]</p> <p>5. ສ [0000.0000.0000.0002]</p>
<p>7. ນ ສິ ຕ</p> <p>1. ນ [085c.0200.0020.0002]</p> <p>2. ສ [0000.0219.0020.0002]</p> <p>3. ສ [0000.0000.0000.0002]</p>	<p>7. ນ ສິ ຕ</p> <p>1. ນ [085c.0200.0020.0002]</p> <p>2. ສ [0000.0214.0020.0002]</p> <p>3. ສ [0000.0000.0000.0002]</p>
<p>8. ນ ເ</p> <p>1. ນ [085c.0200.0020.0002]</p> <p>2. ເX or X̄ [0000.023C.0020.0002]</p>	<p>8. ນ ເ</p> <p>1. ນ [085c.0200.0020.0002]</p> <p>2. ເ [0000.022D.0020.0002]</p>

While allocating each character or sequence of characters their respective collation elements we have to form collation element array. For example collation element array for ກາງ using combinational technique will be:

[085c.0200.0020.0002],
[0000.0269.0020.0002], [0000.0000.0048.0002]

And ‘ເງ’ will be [085c.0200.0020.0002],
[0000.023C.0020.0002].

6.5. Step IV (Form Sort Key)

This is the last step before we could actually use any of the sorting techniques to sort the key. But in order to collate we have to form a sort key first based on the collation elements that we have assembled together inform of collation element array for each word and for each syllable within each word. The sort key is formed by successively appending weights from the collation element array. We start from the primary level pick up the primary level weight for each character if it is not zero and append it to the sort key. At the end of syllable we put zero as a level separator then start again this time picking weight of each character at secondary level. The procedure ends when we are done with ternary level each time appending zero at the end as level separator. Each syllable should have separate sort key. For instance sort key for:

‘ນອນ’ is [085c 0000 0200 0269 0000 0020 0020 0048 0000 0002 0002 0002]

And ‘ເງ’: [085c 0000 0200 023C 0000 0020 0020 0000 0002 0002]

6.6. Step V (Compare the Sort Key)

Finally we are ready to compare the sort keys and to sort the Lao strings. Compare the sort keys for each of the input strings, using a binary comparison. This means that:

- Level 3 differences are ignored if there are any Level 1 or 2 differences
- Level 2 differences are ignored if there are any Level 1 differences
- Level 1 differences are never ignored.

Therefore when comparing

‘ນອນ’ [085c 0000 0200 0269 0000 0020 0020 0048 0000 0002 0002 0002]

‘ເງ’ [085c 0000 0200 023C 0000 0020 0020 0000 0002 0002] we find out that difference is at secondary i.e. at vowel level.

We move on to sort key of next syllable when we current syllables are same, following example illustrates:

1. ກາງເສີນ
ກາງ (First Syllable)

Collation Element Array:

[0820.0200.0020.0002], [0000.020F.0020.0002],
[0000.0000.002A.0002]

Sort Key:

[0820 0000 0200 020F 0000 0020 0020 002A 0000
00002 0002 0002]

ເສີນ (Second Syllable)

Collation Element Array:

[08C0.0200.0020.0002], [0000.0255.0020.0002],
[0000.0000.0048.0002], [0000.0000.0000.0007]

Sort Key:
 [08C0 0000 0200 0255 0000 0020 0020 0048 0000
 0002 0002 0002 0007]

2. ກາງຫລາວ

ກາງ (First Syllable)

Collation Element Array:
 [0820.0200.0020.0002], [0000.020F.0020.0002],
 [0000.0000.002A.0002]

Sort Key:
 [0820 0000 0200 020F 0000 0020 0020 002A 0000
 00002 0002 0002]

ຫລາວ (Second Syllable)

Collation Element Array:
 [08B1.0200.0020.0002], [0000.020F.0020.0002],
 [0000.0000.0066.0002]

Sort Key:
 [08B1 0000 0200 020F 0000 0020 0020 0066 0002
 0002 0002]

As can be observed in above example the decision can not be made by comparing the first syllable as they are same in both the word so we have to leave it to second syllable. While comparing the sort keys for second syllables we find out that difference is at primary level:

[08C0 0000 0200 0255 0000 0020 0020 0048 0000
 0002 0002 0002 0007]
 [08B1 0000 0200 020F 0000 0020 0020 0066 0002
 0002 0002]

Therefore ‘ຫລາວ’ is sorted before ‘ອິນ’

Output (Differences highlighted)

Combinational Consonants and Vowel	Based on Alphabetical Order
<ul style="list-style-type: none"> • ກາງຫລາວ • ກາງອິນ • ນ້າ • ນີ້ • ເນະ • ເນິນ • ເນ • ນອນ 	<ul style="list-style-type: none"> • ກາງຫລາວ • ກາງອິນ • ນ້າ • ນີ້ • ເນ • ເນະ • ເນິນ • ນອນ

7. Conclusion

This document discusses one possible solution for Lao collation based on syllabification. It discusses how it can cater both the popular lexicographic standards that prevail. It discusses how we can use Unicode Collation Algorithm to sort Lao strings. The algorithm is totally based on the line breaking module which means if segmentation could not be done properly the results would effected. Because if the module is unable to segment it could not find out what role is the character playing within syllable. Generally correct segmentation has been achieved up to 98% but there are few exceptions when dealing with foreign language words.

8. Reference

[1] “Unicode Conference Abstract”
<http://www.unicode.org/iuc/iuc27/biosabstracts/a344.html>

Glyph	Unicode	Collation Elements	Unicode Name
← Consonants →			
ກ	0E81	0820 0200 0020 0002	LAO LETTER KO
ຂ	0E82	0822 0200 0020 0002	LAO LETTER KHO SUNG
ຄ	0E84	0824 0200 0020 0002	LAO LETTER KHO TAM
ງ	0E87	0826 0200 0020 0002	LAO LETTER NGO
ຈ	0E88	0828 0200 0020 0002	LAO LETTER CO
ສ	0EAA	082A 0200 0020 0002	LAO LETTER SO SUNG
ຊ	0E8A	082C 0200 0020 0002	LAO LETTER SO TAM
ຍ	0E8D	082E 0200 0020 0002	LAO LETTER NYO
ດ	0E94	0830 0200 0020 0002	LAO LETTER DO
ຕ	0E95	0832 0200 0020 0002	LAO LETTER TO
ຖ	0E96	0834 0200 0020 0002	LAO LETTER THO SUNG
ທ	0E97	0836 0200 0020 0002	LAO LETTER THO TAM
ນ	0E99	0838 0200 0020 0002	LAO LETTER NO
ບ	0E9A	083A 0200 0020 0002	LAO LETTER BO
ປ	0E9B	083C 0200 0020 0002	LAO LETTER PO
ຜ	0E9C	083E 0200 0020 0002	LAO LETTER PHO SUNG
ຝ	0E9D	0840 0200 0020 0002	LAO LETTER FO TAM
ພ	0E9E	0842 0200 0020 0002	LAO LETTER PHO TAM
ຟ	0E9F	0844 0200 0020 0002	LAO LETTER FO SUNG
ມ	0EA1	0846 0200 0020 0002	LAO LETTER MO
ຢ	0EA2	0848 0200 0020 0002	LAO LETTER YO
ລ	0EA3	084A 0200 0020 0002	LAO LETTER LO LING
ວ	0EBC	084C 0200 0020 0002	LAO SEMI VOWEL SIGN LO
ຸ			

	0EA5	084E 0200 0020 0002	LAO LETTER LO LOOT
ວ	0EA7	0850 0200 0020 0002	LAO LETTER WO
ຫ	0EAB	0852 0200 0020 0002	LAO LETTER HO SUNG
ຫງ	0EAB+0E87	0854 0200 0020 0002	LAO LETTER HO SUNG+ LAO LETTER NGO
ຫຍ	0EAB+0E8D	0856 0200 0020 0002	LAO LETTER HO SUNG + LAO LETTER NYO
ຫນ	0EAB+0E99	0858 0200 0020 0002	LAO LETTER HO SUNG + LAO LETTER NO
ຫມ	0EDC	0858 0200 0020 0002	LAO LETTER HO NO
ຫມ	0EAB+0EA1	0860 0200 0020 0002	LAO LETTER HO SUNG + LAO LETTER MO
ຫມ	0EDD	0860 0200 0020 0002	LAO LETTER HO MO
ຫລ	0EAB+0EA5	0864 0200 0020 0002	LAO LETTER HO SUNG + LAO LETTER LO LOOT
ຫຼ	0EAB+0EBC	0864 0200 0020 0002	LAO LETTER HO SUNG + LAO SEMIVOWEL SIGN LO
ຫວ	0EAB+0EA7	0868 0200 0020 0002	LAO LETTER HO SUNG + LAO LETTER WO
ອ	0EAD	086A 0200 0020 0002	LAO LETTER O
ຮ	0EAE	086C 0200 0020 0002	LAO LETTER HO TAM
← Vowels →			
ຮ	0EB0	0000 0202 0020 0002	LAO VOWEL SIGN A
ຮ+X8/X9	0EB1+X8/X9	0000 0204 0020 0002	LAO VOWEL SIGN MAI KAN + CONSONANTAL
າ	0EB2	0000 0206 0020 0002	LAO VOWEL SIGN AA
ິ	0EB4	0000 0208 0020 0002	LAO VOWEL SIGN I
ີ	0EB5	0000 020A 0020 0002	LAO VOWEL SIGN II
ົ	0EB6	0000 020C 0020 0002	LAO VOWEL SIGN Y
ື	0EB7	0000 0210 0020 0002	LAO VOWEL SIGN YY
ູ	0EB8	0000 0212 0020 0002	LAO VOWEL SIGN U
ູ	0EB9	0000 0214 0020 0002	LAO VOWEL SIGN UU
ຸ			

	0EC0+X+0EB0	0000 0216 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT + LAO VOWEL SIGN A
ຸ+X8/X9	0EC0+0EB1+X8/X9	0000 0218 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN MAI KAN + CONSONANTAL
໊X	0EC0+X	0000 021A 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT
ຸX໋	0EC1+X+0EB0	0000 021C 0020 0002	LAO VOWEL SIGN EI + MAIN CONSONANT + LAO VOWEL SIGN A
ຸ໋+X8/X9	0EC1+0EB1+X8/X9	0000 0220 0020 0002	LAO VOWEL SIGN EI + LAO VOWEL SIGN MAI KAN + CONSONANTAL
໋X	0EC1+X	0000 0222 0020 0002	LAO VOWEL SIGN EI + MAIN CONSONANT
໊X໋	0EC2+X+0EB0	0000 0224 0020 0002	LAO VOWEL SIGN O + MAIN CONSONANT + LAO VOWEL SIGN A
໋	0EBB	0000 0226 0020 0002	LAO VOWEL SIGN MAI KON
໊X	0EC2+X	0000 0228 0020 0002	LAO VOWEL SIGN O + MAIN CONSONANT
໊X໋	0EC0+X+0EB2+0EB0	0000 022A 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT + LAO VOWEL SIGN AA + LAO VOWEL SIGN A
໋	0ECD	0000 022C 0020 0002	LAO NIGGAHITA
X໋+X8/X9	X+0EAD+X8/X9	0000 022E 0020 0002	MAIN CONSONANT + LAO LETTER O + CONSONANTAL
໋	0EC0+0EB4	0000 0230 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN I
໋	0EC0+0EB5	0000 0232 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN II
໋໋	0EC0+0EB1+0EBD	0000 0234 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN MAI KAN + LAO SEMI VOWEL SIGN NYO
໋໋	0EC0+X+0EBD	0000 0236 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT + LAO SEMI VOWEL SIGN NYO
໋໋+X8/X9	0EBD+X8/X9	0000 0238 0020 0002	LAO SEMI VOWEL SIGN NYO + CONSONANTAL
໋໋	0EBB+0EA7+0EB0	0000 023A 0020 0002	LAO VOWEL SIGN MAI KON + LAO LETTER WO + LAO VOWEL SIGN A
໋໋+X8/X9	0EB1+0EA7+X8/X9	0000 023C 0020 0002	LAO VOWEL SIGN MAI KON + LAO LETTER WO + CONSONANTAL
໋໋	0EBB+0EA7	0000 023E 0020 0002	LAO VOWEL SIGN MAI KON + LAO LETTER WO
໋໋໋	0EC0+0EB6+0EAD	0000 0240 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN Y + LAO LETTER O
໋໋໋			LAO VOWEL SIGN E + LAO

	0ECA	0000 0000 0000 0008	LAO TONE TI
◌◌	0ECB	0000 0000 0000 0008	LAO TONE MAI CATAWA
← Numerals→			
◌	0ED0	0700 0200 0020 0002	LAO DIGIT ZERO
◌	0ED1	0702 0200 0020 0002	LAO DIGIT ONE
◌	0ED2	0704 0200 0020 0002	LAO DIGIT TWO
◌	0ED3	0706 0200 0020 0002	LAO DIGIT THREE
◌	0ED4	0708 0200 0020 0002	LAO DIGIT FOUR
◌	0ED5	070A 0200 0020 0002	LAO DIGIT FIVE
◌	0ED6	070C 0200 0020 0002	LAO DIGIT SIX
◌	0ED7	070E 0200 0020 0002	LAO DIGIT SEVEN
◌	0ED8	0710 0200 0020 0002	LAO DIGIT EIGHT
◌	0ED9	0712 0200 0020 0002	LAO DIGIT NINE
← Various Symbols→			
◌◌	0ECC	0000 0000 0000 0000	LAO CANCELLATION MARK
◌	0EC6	0000 0000 0000 0000	LAO KO LA
◌	0EAF	0000 0000 0000 0000	LAO ELLIPSIS

Table 1: Collation Elements – Combinational Consonants and Vowels

Glyph	Unicode	Collation Elements	Unicode Name
← Consonants→			
ກ	0E81	0820 0200 0020 0002	LAO LETTER KO
ຂ	0E82	0822 0200 0020 0002	LAO LETTER KHO SUNG
ຄ	0E84	0824 0200 0020 0002	LAO LETTER KHO TAM
ງ	0E87	0826 0200 0020 0002	LAO LETTER NGO
ຈ	0E88	0828 0200 0020 0002	LAO LETTER CO
ສ	0EAA	082A 0200 0020 0002	LAO LETTER SO SUNG
ຮ			

	0E8A	082C 0200 0020 0002	LAO LETTER SO TAM
ຢ	0E8D	082E 0200 0020 0002	LAO LETTER NYO
ດ	0E94	0830 0200 0020 0002	LAO LETTER DO
ຕ	0E95	0832 0200 0020 0002	LAO LETTER TO
ຖ	0E96	0834 0200 0020 0002	LAO LETTER THO SUNG
ທ	0E97	0836 0200 0020 0002	LAO LETTER THO TAM
ນ	0E99	0838 0200 0020 0002	LAO LETTER NO
ບ	0E9A	083A 0200 0020 0002	LAO LETTER BO
ປ	0E9B	083C 0200 0020 0002	LAO LETTER PO
ຜ	0E9C	083E 0200 0020 0002	LAO LETTER PHO SUNG
ຝ	0E9D	0840 0200 0020 0002	LAO LETTER FO TAM
ພ	0E9E	0842 0200 0020 0002	LAO LETTER PHO TAM
ຟ	0E9F	0844 0200 0020 0002	LAO LETTER FO SUNG
ມ	0EA1	0846 0200 0020 0002	LAO LETTER MO
ຢ	0EA2	0848 0200 0020 0002	LAO LETTER YO
ຮ	0EA3	084A 0200 0020 0002	LAO LETTER LO LING
ລ	0EA5	084E 0200 0020 0002	LAO LETTER LO LOOT
ວ	0EA7	0850 0200 0020 0002	LAO LETTER WO
ຫ	0EAB	0852 0200 0020 0002	LAO LETTER HO SUNG
ຸ	0EAB+0EBC	0866 0200 0020 0002	LAO LETTER HO SUNG + LAO SEMIVOWEL SIGN LO
ອ	0EAD	086A 0200 0020 0002	LAO LETTER O
ຂ	0EAE	086C 0200 0020 0002	LAO LETTER HO TAM
ຫງ	0EDC	0870 0200 0020 0002	LAO LETTER HO NO
ຫຍ	0EDD	0872 0200 0020 0002	LAO LETTER HO MO
← Vowels →			
ູ	0EB0	0000 0202 0020 0002	LAO VOWEL SIGN A
າ	0EB2	0000 0206 0020 0002	LAO VOWEL SIGN AA

໐	0EB4	0000 0208 0020 0002	LAO VOWEL SIGN I
໑	0EB5	0000 020A 0020 0002	LAO VOWEL SIGN II
໒	0EB6	0000 020C 0020 0002	LAO VOWEL SIGN Y
໓	0EB7	0000 0210 0020 0002	LAO VOWEL SIGN YY
໔	0EB8	0000 0212 0020 0002	LAO VOWEL SIGN U
໕	0EB9	0000 0214 0020 0002	LAO VOWEL SIGN UU
໖	0EC0	0000 0216 0020 0002	LAO VOWEL SIGN
໗	0EC1	0000 0222 0020 0002	LAO VOWEL SIGN EI
໘	0EC2	0000 0224 0020 0002	LAO VOWEL SIGN O
໙	0ECD	0000 022C 0020 0002	LAO NIGGAHITA
໑໐	0EC4	0000 0246 0020 0002	LAO VOWEL SIGN AI
໑໑	0EC3	0000 0248 0020 0002	LAO VOWEL SIGN AY
໑໒	0EB1	0000 024A 0020 0002	LAO VOWEL SIGN MAI KAN
໑໓	0EBB	0000 024C 0020 0002	LAO VOWEL SIGN MAI KON
໑໔	0EBD	0000 0250 0020 0002	LAO SEMI VOWEL SIGN NYO
໑໕	0EA7	0000 0252 0020 0002	LAO LETTER WO
໑໖	0EAD	0000 0254 0020 0002	LAO LETTER O
← Consonantal →			
໑໑໗	0E81	0000 0000 0022 0002	LAO LETTER KO
໑໑໘	0E87	0000 0000 0024 0002	LAO LETTER NGO
໑໑໙	0E88	0000 0000 0026 0002	LAO LETTER CO
໑໑໑໐	0EAA	0000 0000 0028 0002	LAO LETTER SO SUNG
໑໑໑໑	0E8A	0000 0000 002A 0002	LAO LETTER SO TAM
໑໑໑໒	0E8D	0000 0000 002C 0002	LAO LETTER NYO
໑໑໑໓	0E94	0000 0000 002E 0002	LAO LETTER DO
໑໑໑໔	0E99	0000 0000 0030 0002	LAO LETTER NO
໑໑໑໕			

	0E9A	0000 0000 0032 0002	LAO LETTER BO
ພ	0E9E	0000 0000 0034 0002	LAO LETTER PHO TAM
ຟ	0E9F	0000 0000 0036 0002	LAO LETTER FO SUNG
ມ	0EA1	0000 0000 0038 0002	LAO LETTER MO
ລ	0EA5	0000 0000 003A 0002	LAO LETTER LO LOOT
ວ	0EA7	0000 0000 003C 0002	LAO LETTER WO
← Tone Marks→			
◌◌◌◌	0EC8	0000 0000 0000 0004	LAO TONE MAI EK
◌◌◌◌	0EC9	0000 0000 0000 0006	LAO TONE MAI THO
◌◌◌◌	0ECA	0000 0000 0000 0008	LAO TONE TI
◌◌◌◌	0ECB	0000 0000 0000 0008	LAO TONE MAI CATAWA
← Numerals→			
໐	0ED0	0700 0200 0020 0002	LAO DIGIT ZERO
໑	0ED1	0702 0200 0020 0002	LAO DIGIT ONE
໒	0ED2	0704 0200 0020 0002	LAO DIGIT TWO
໓	0ED3	0706 0200 0020 0002	LAO DIGIT THREE
໔	0ED4	0708 0200 0020 0002	LAO DIGIT FOUR
໕	0ED5	070A 0200 0020 0002	LAO DIGIT FIVE
໖	0ED6	070C 0200 0020 0002	LAO DIGIT SIX
໗	0ED7	070E 0200 0020 0002	LAO DIGIT SEVEN
໘	0ED8	0710 0200 0020 0002	LAO DIGIT EIGHT
໙	0ED9	0712 0200 0020 0002	LAO DIGIT NINE
← Various Symbols→			
◌◌◌◌	0ECC	0000 0000 0000 0000	LAO CANCELLATION MARK
໘	0EC6	0000 0000 0000 0000	LAO KO LA
໘	0EAF	0000 0000 0000 0000	LAO ELLIPSIS

Table 2: Collation Elements – Alphabetical Order

