Lao Collation

Phonpasit Phissamay, Nadir Durrani
STEA[1] , CRULP[2]
phonpasit@stea.gov.la[1], nadirdurrani@yahoo.com[2]

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 though 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 ຍ ດ ຕ ຖ ທ ນ ບ ປ ຜ ຝ ພ ຟ ມ ຢ
ຣ ◌ ລ ວ ຫ ຫງ ຫຍ ຫນ ຫມ ຫຣ ຫລ ຫວ ຫ ອ ຮ

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

---

36th. The in-betweens ຫ are to cater behavior of combinational consonant which can combine with several other consonants {ງ ຍ ຫນ ມ ລ ວ} to exhibit different behavior. Consonants {ຫນ, ໝ, ຫຼ} are same as {ຫນ, ຫມ, ຫ ລ} respectively as people sometime prefer to write ໝ instead of writing ຫມ therefore we assign them same collation element to a pair in table.

Another note worthy point is that Lao character ຼ is a lower vowel but we have placed it in nuclear consonant section as it does not play any role as a vowel it just combines with ຫ in Lao words, can combine with other nuclear consonants for foreign language words so we can treat it as a consonant. Moreover we assign it same collation element as ລ.

Lao digits are also placed primary level of collation.

໐ ໑ ໒ ໓ ໔ ໕ ໖ ໗ ໘ ໙

3.2. Vowels and Combinational Vowels

If a syllable has same nuclear consonant then we have to compare the vowels and their combinations with consonantal or themselves at secondary level. These are given below:

ະ ັ +X8/X9 າ ິ ີ ຶ ື ຸ ູ ເXະ ເັ+X8/X9 ເX ແXະ ແັ+X8/X9 ແX ໂXະ ົ ໂX ເXາະ ໍ Xອ+X8/X9 ເິ ເີ ເັຽ ເXຽ ຽ+X8/X9 ົວະ ັວ+X8/X9 ົວ ເຶອ ເືອ Xວ+X8/X9 ໄX ໃX ເົາ ຳ ໍ+າ

Lao Script altogether has 21 vowels. Excluding the lower vow'ຼ' leaves us with 20. These twenty when combined with nuclear consonants, consonantal and with each other have different roles to play when sorting takes place. In above list symbols 'X','X8/X9' have been borrowed from the line breaking document. 'X' represents the nuclear consonant and 'X8/X9' stands for the consonantal. Therefore in addition to finding out the syllable boundaries we have to typify each Lao character in syllable. Another reason to characterize is that we have to reorder each syllable. We will come to details of this in later part of document. With these structuring and combinational possibilities we have to carefully look in the neighboring of vowels to characterize it as one of the above.

3.3. Consonantal (X8/X9)

As already mentioned some of the Lao characters can act hybrid depending upon their position in the syllable structure. When working as consonantal they play their part at tertiary level. The list is given below:

3.4. Tone Marks

Last but not the least the operation goes with four tone marks that play at ternary level. These are given below:

່ ້ ໊ ໋

3.5. Ignorable Level

The punctuation marks, Lao digits and several other characters given below are ignorable at all levels.

໌ ໆ ຯ

Characters ໆ and ຯ are used for word and sentence repetition, instead of writing word or sentence more then once these characters are employed. Third character ໌ is controversial; it can be looked at as a vowel but occurs at position X10 in syllable therefore we can place it anywhere.

4. Based on Alphabetical Order

4.1. Nuclear Consonants and Lao Digit

If a syllable has same nuclear consonant then we have to compare the vowels and their combinations with consonantal or themselves at secondary level. These are given below:

ກ ຂ ຄ ງ ຈ ສ ຊ ຍ ດ ຕ ຖ ທ ນ ບ ປ ຜ ຝ ພ ຟ ມ ຢ ຣ ລ ວ ຫ ຫຼ ອ ຮ ຫນ ໝ

Consequently if we sort ຫມ and ໝ these two approaches will produce different results. First approach will not sort them at all as they have same code in the defined collation element table. However with second technique the decision would be made

by comparing ກ and ໝ . Likewise is the case with all the other combinational consonants.

Lao digits are also placed primary level of collation.

໐ ໑ ໒ ໓ ໔ ໕ ໖ ໗ ໘ ໙

4.2. Vowels

Same as the case with former approach this approach operates with vowels at secondary level too. However the difference is that here we use single vowels without looking into what it is combined with in its neighbors in the syllable. Therefore the number in this category has been dropped down to 20.

ະ າ ິ ີ ຶ ື ຸ ູ ເ ແ ໂ ໍ ໄ
ໃ ໋ ໌ ໆ ວ ອ

Consonantal, Tone Marks and Ignorable Level

Rest remains the same, the consonantal operating at tertiary, tone marks at ternary and rest at ignorable levels.

5. Lao Collation Elements

In order to realize the above given sequence we need to define the collation element table and employ Unicode Collation Algorithm to obtain desired results. These collation elements can be used to generate the sort keys for the input strings to realize the required lexicographic sorting for Lao. The Lao Collation element consist of four weights, e.g.0861.0020.0002.0061 for Latin Small Letter a. The first (left-most) weight is primary weight (or Level 1 weight), the second is secondary weight (or Level 2 weight), the third is tertiary weight (or Level 3 weight), and the fourth is quaternary weight (or Level 4 weight). These levels are used to distinguish between characters, diacritics, capitalization ('a' vs. 'A') and other lesser significant differences respectively. 0000 sequence is used to make the character ignorable at the level this weight is assigned. Unicode also describes default values for collation elements for all Unicode letters in Allkeys.txt. However, this default sequence cannot realize collation for a specific language. In order to define collation properly for any particular language, these collations codes must be re-ordered in such a way that the coding sequence conforms to the standardized linguistic sorting order.

As has been mentioned repeatedly to achieve Lao collation we need 4 levels plus one level of Unicode. Appendix A & B define the Collation Element Table for both the approaches.

6. Lao Collation Algorithm

Lao Collation Algorithm has been based on Unicode Collation Algorithm. But we need to add a few steps and mould it to our purpose. Most prime addition is line breaking module which divides the word into syllables, then each syllable act has individual entity i.e. each syllable will have its own array of collation elements and its own sort key. Given below are the steps that we go by while sorting Lao strings. We will proceed by an example and show step by step how things work out.

6.1. Sample Input

1. ນອນ
2. ນ້າ
3. ເນະ
4. ກາງຮັບ
5. ກາງຫລາວ
6. ເນັ້ນ
7. ນີ້
8. ເນ

6.2. Step-I (Segmentation)

First of all we need to find out the how many syllables are there in each word and what role is each character playing within syllable. Therefore we send each word one by one to Line Breaking Module which based on some test conditions find outs the required information as given below. More details of this can be found out in Line Breaking document.

For input stri 'ກ້າງ' we have information that it has only one syllable whose boundary is at 4th character. Other than this we come to know that in syllable ກ is the main consonant (X), (X2) also is treated as main consonant, ້ as tone mark (X5), າ as vowel (X7) and ງ as consonantal (X8/X9). Other than that we assign weights to each character in the syllable that are later to be used during reordering. Consonants have minimum weights, and then come

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
ເX$_1$	ຫ	ກ$_{01}$ ຂ$_{02}$ ຄ$_{03}$ ງ$_{04}$	X$_1$	X$_1$	X$_1$	X'$_1$	ວ$_1$	Xຍ$_1$	ກ$_1$	າ$_1$	ໆ$_1$
ແX$_2$		ຈ$_{05}$ ສ$_{06}$ ຊ$_{07}$ ຍ$_{08}$	X$_2$	X$_2$	X$_2$	X$_2$	ອ$_2$	Xຫ$_2$	ຢ$_2$	ຳ$_2$	໅$_2$
ໂX$_3$		ດ$_{09}$ ຕ$_{10}$ ຖ$_{11}$ ທ$_{12}$	ວ$_3$		X$_3$	X$_3$	]$_3$	Xຳ$_3$	ຍ$_3$	ຂ$_3$	X$_3$
ໄX$_4$		ນ$_{13}$ ບ$_{14}$ ປ$_{15}$ ຜ$_{16}$	ລ$_4$		X$_4$	X$_4$			ຄ$_4$	ພ$_4$	
ໃX$_5$		ຝ$_{17}$ ພ$_{18}$ ຟ$_{19}$ ມ$_{20}$			X$_5$				ບ$_5$	ຟ$_5$	
		ຢ$_{21}$ ຣ$_{22}$ ລ$_{23}$ ວ$_{24}$			X$_6$				ມ$_6$	ລ$_6$	
		ຫ$_{25}$ ອ$_{26}$ ຮ$_{27}$ ຫ$_{28}$			X$_7$				ບ$_7$		
		ຫ$_{29}$							ວ$_8$		

Categories X0, X3, X4, X6 and X7 all contain vowels. So we assign them code as they appear in the table i.e. ເ will have minimum weight and ໍາ 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 າ ໍ
- ນ ເ ະ
- ກ າ ງ (New Syllable) ຮ ິ ນ ໍ
- ກ າ ງ (New Syllable) ລ ຫ າ ວ
- ນ ເ ິ ນ ໍ
- ນ ໍ ໍ
- ນ ເ

As you can see that input has been reordered, but only within the syllable that it belongs. If a word contains more then 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 ອ ນ
1. ນ [085c.0200.0020.0002]	1. ນ [085c.0200.0020.0002]
2. X~ອ+X8/9 [0000.0269.0020.0002]	2. ອ [0000.025F.0020.0002]
3. ນ [0000.0000.0048.0002]	3. ນ [0000.0000.0048.0002]
2. ນ າ ໍ	2. ນ າ ໍ
1. ນ [085c.0200.0020.0002]	1. ນ [085c.0200.0020.0002]
2. າ [0000.020F.0020.0002]	2. າ [0000.020A.0020.0002]
3. ໍ [0000.0000.0000.000C]	3. ໍ [0000.0000.0000.000C]
3. ນ ເ ະ	3. ນ ເ ະ
1. ນ [085c.0200.0020.0002]	1. ນ [085c.0200.0020.0002]
2. ເXະ or Xແ [0000.0232.0020.0002]	2. ເ [0000.022D.0020.0002]
	3. ະ [0000.0205.0020.0002]
4. ກ າ ງ (New Syllable) ຮ ິ ນ ໍ	4. ກ າ ງ (New Syllable) ຮ ິ ນ ໍ
4a. ກ າ ງ (First Syllable)	4a. ກ າ ງ (First Syllable)
1. ກ [0820.0200.0020.0002]	1. ກ [0820.0200.0020.0002]
2. າ [0000.020F.0020.0002]	2. າ [0000.020A.0020.0002]
3. ງ [0000.0000.002A.0002]	3. ງ [0000.0000.002A.0002]
4b. ຮ ິ ນ ໍ (Second Syllable)	4b. ຮ ິ ນ ໍ (Second Syllable)
1. ຮ [08C0.0200.0020.0002]	1. ຮ [08C0.0200.0020.0002]
2. ິ [0000.0255.0020.0002]	2. ິ [0000.0250.0020.0002]
3. ນ [0000.0000.0048.0002]	3. ນ [0000.0000.0048.0002]
4. ໍ [0000.0000.0000.0007]	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

5a. ກ າ ງ (First Syllable)			5a. ກ າ ງ (First Syllable)		
1.	ກ	[0820.0200.0020.0002]	1.	ກ	[0820.0200.0020.0002]
2.	າ	[0000.020F.0020.0002]	2.	າ	[0000.020A.0020.0002]
3.	ຽ	[0000.0000.002A.0002]	3.	ຽ	[0000.0000.002A.0002]

5b. ທ ລ າ ວ (Second Syllable)			5b. ລ ທ າ ວ (Second Syllable)		
1.	ທລ	[08B1.0200.0020.0002]	1.	ລ	[088E.0200.0020.0002]
2.	າ	[0000.020F.0020.0002]	2.	ທ	[0898.0200.0020.0002]
3.	ວ	[0000.0000.0066.0002]	3.	າ	[0000.020A.0020.0002]
			4.	ວ	[0000.0000.0066.0002]

6. ບ ເ ບ ຶ			6. ບ ເ ບ ຶ		
1.	ບ	[085c.0200.0020.0002]	1.	ບ	[085c.0200.0020.0002]
2.	ເບ–X8/9	[0000.0237.0020.0002]	2.	ເ	[0000.022D.0020.0002]
3.	ບ	[0000.0000.0048.0002]	3.	ຶ	[0000.024B.0020.0002]
4.	ຶ	[0000.0000.0000.000C]	4.	ບ	[0000.0000.0048.0002]
			5.	ຶ	[0000.0000.0000.000C]

7. ບ ຶ ຶ			7. ບ ຶ ຶ		
1.	ບ	[085c.0200.0020.0002]	1.	ບ	[085c.0200.0020.0002]
2.	ຶ	[0000.0219.0020.0002]	2.	ຶ	[0000.0214.0020.0002]
3.	ຶ	[0000.0000.0000.000C]	3.	ຶ	[0000.0000.0000.000C]

8. ບ ເ			8. ບ ເ		
1.	ບ	[085c.0200.0020.0002]	1.	ບ	[085c.0200.0020.0002]
2.	ເX or Xເ	[0000.023C.0020.0002]	2.	ເ	[0000.022D.0020.0002]

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   using combinational technique will be:

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

And 'ເບ' 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:

'ບອນ' is [085c 0000 0200 0269 0000 0020 0020 0048 0000 0002 0002 0002]
And 'ເບ' , [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

'ບອນ' [085c 0000 0200 0269 0000 0020 0020 0048 0000 0002 0002 0002]
'ເບ' [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. ກາງຮິ່ນ
    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]

ຮິ່ນ 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 'ຊື່ນ'

Output (Differences highlighted)

Combinational Consonants and Vowel	*Based on Alphabetical Order*
• ກາງຫລາວ	• ກາງຫລາວ
• ກາງຊື່ນ	• ກາງຊື່ນ
• ນ້ຳ	• ນ້ຳ
• ນີ້	• ນີ້
• ເນະ	• ເນ
• ເນັ້ນ	• ເນະ
• ເນ	• ເນັ້ນ
• ນອນ	• 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
ຍ	0EA2	0848 0200 0020 0002	LAO LETTER YO
ຣ	0EA3	084A 0200 0020 0002	LAO LETTER LO LING
ຼ	0EBC	084C 0200 0020 0002	LAO SEMI VOWEL SIGN LO
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
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
ໜ	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→	
ະ	0EB0	0000 0202 0020 0002	LAO VOWEL SIGN A
ັ+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
ເXະ	0EC0+X+0EB0	0000 0216 0020 0002	LAO VOWEL SIGN E + MAIN

			CONSONANT + LAO VOWEL SIGN A
ເັ+X8/X9	0EC0+0EB1+X8/X9	0000 0218 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN MAIN KAN + CONSONANTAL
ເX	0EC0+X	0000 021A 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT
ແXະ	0EC1+X+0EB0	0000 021C 0020 0002	LAO VOWEL SIGN EI + MAIN CONSONANT + LAO VOWEL SIGN A
ແັ+X8/X9	0EC1+0EB1+X8/X9	0000 0220 0020 0002	LAO VOWEL SIGN EI + LAO VOWEL SIGN MAI KAN + CONSONANTAL
ແX	0EC1+X	0000 0222 0020 0002	LAO VOWEL SIGN EI + MAIN CONSONANT
ໂXະ	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
ໂX	0EC2+X	0000 0228 0020 0002	LAO VOWEL SIGN O + MAIN CONSONANT
ເXາະ	0EC0+X+0EB2+0EB0	0000 022A 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT + LAO VOWEL AA + LAO VOWEL SIGN A
ໍ	0ECD	0000 022C 0020 0002	LAO NIGGAHITA
Xອ+X8/X9	X+0EAD+X8/X9	0000 022E 0020 0002	MAIN CONSONANT + LAO LETTER O + CONSONANTAL
ເິ	0EC0+0EB4	0000 0230 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN I
ເີ	0EC0+0EB5	0000 0232 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN II
ເັຍ	0EC0+0EB1+0EBD	0000 0234 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN MAI KAN + LAO SEMIVOWEL SIGN NYO
ເXຍ	0EC0+X+0EBD	0000 0236 0020 0002	LAO VOWEL SIGN E + MAIN CONSONANT + LAO SEMI VOWEL SIGN NYO
ຍ+X8/X9	0EBD+X8/X9	0000 0238 0020 0002	LAO SEMI VOWEL SIGN NYO + CONSONANTAL
ົວະ	0EBB+0EA7+0EB0	0000 023A 0020 0002	LAO VOWEL SIGN MAI KON + LAO LETTER WO + LAO + VOWEL SIGN A
ັວ+X8/X9	0EB1+0EA7+X8/X9	0000 023C 0020 0002	LAO VOWEL SIGN MAI KON + LAO LETTER WO + CONSONANTAL
ົວ	0EBB+0EA7	0000 023E 0020 0002	LAO VOWEL SIGN MAI KON + LAO LETTER WO
ເຶອ	0EC0+0EB6+0EAD	0000 0240 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN Y + LAO LETTER O
ເືອ	0EC0+0EB7+0EAD	0000 0242 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN YY + LAO

			LETTER O
X໐+X8/X9	X+0EA7+X8/X9	0000 0244 0020 0002	MAIN CONSONANT + LAO LETTER WO + CONSONANTAL
ໄX	0EC4+X	0000 0246 0020 0002	LAO VOWEL SIGN AI + MAIN CONSONANT
ໃX	0EC3+X	0000 0248 0020 0002	LAO VOWEL SIGN AY + MAIN CONSONANT
ເ◌ົາ	0EC0+0EBB+0EB2	0000 024A 0020 0002	LAO VOWEL SIGN E + LAO VOWEL SIGN MAI KON + LAO VOWEL SIGN AA
◌ໍາ	0EB3	0000 024C 0020 0002	LAO VOWEL SIGN AM
◌ໍ+າ	0ECD+0EB2	0000 024C 0020 0002	LAO NIGGAHITA + LAO VOWEL SIGN AA
	← **Consonantal**→		
ກ	0E81	0000 0000 0022 0002	LAO LETTER KO
ງ	0E87	0000 0000 0024 0002	LAO LETTER NGO
ຈ	0E88	0000 0000 0026 0002	LAO LETTER CO
ສ	0EAA	0000 0000 0028 0002	LAO LETTER SO SUNG
ຊ	0E8A	0000 0000 002A 0002	LAO LETTER SO TAM
ຍ	0E8D	0000 0000 002C 0002	LAO LETTER NYO
ດ	0E94	0000 0000 002E 0002	LAO LETTER DO
ນ	0E99	0000 0000 0030 0002	LAO LETTER NO
ບ	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
ໆ	0EC6	0000 0000 0000 0000	LAO KO LA
ຯ	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
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
ຣ	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
ຫຼ	0EAB+0EBC	0866 0200 0020 0002	LAO LETTER HO SUNG + LAO SEMIVOWEL SIGN LO
ອ	0EAD	086A 0200 0020 0002	LAO LETTER O
ຮ	0EAE	086C 0200 0020 0002	LAO LETTER HO TAM
ຫນ	0EDC	0870 0200 0020 0002	LAO LETTER HO NO
ຫມ	0EDD	0872 0200 0020 0002	LAO LETTER HO MO
← Vowels→			
ະ	0EB0	0000 0202 0020 0002	LAO VOWEL SIGN A
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
ເ	0EC0	0000 0216 0020 0002	LAO VOWEL SIGN
ແ	0EC1	0000 0222 0020 0002	LAO VOWEL SIGN EI
ໂ	0EC2	0000 0224 0020 0002	LAO VOWEL SIGN O
◌ໍ	0ECD	0000 022C 0020 0002	LAO NIGGAHITA
ໄ	0EC4	0000 0246 0020 0002	LAO VOWEL SIGN AI
ໃ	0EC3	0000 0248 0020 0002	LAO VOWEL SIGN AY
◌ັ	0EB1	0000 024A 0020 0002	LAO VOWEL SIGN MAI KAN
◌ົ	0EBB	0000 024C 0020 0002	LAO VOWEL SIGN MAI KON
ຽ	0EBD	0000 0250 0020 0002	LAO SEMI VOWEL SIGN NYO
ວ	0EA7	0000 0252 0020 0002	LAO LETTER WO
ອ	0EAD	0000 0254 0020 0002	LAO LETTER O
		← Consonantal→	
ກ	0E81	0000 0000 0022 0002	LAO LETTER KO
ງ	0E87	0000 0000 0024 0002	LAO LETTER NGO
ຈ	0E88	0000 0000 0026 0002	LAO LETTER CO
ສ	0EAA	0000 0000 0028 0002	LAO LETTER SO SUNG
ຊ	0E8A	0000 0000 002A 0002	LAO LETTER SO TAM
ຍ	0E8D	0000 0000 002C 0002	LAO LETTER NYO
ດ	0E94	0000 0000 002E 0002	LAO LETTER DO
ນ	0E99	0000 0000 0030 0002	LAO LETTER NO
ບ	0E9A	0000 0000 0032 0002	LAO LETTER BO
ພ	0E9E	0000 0000 0034 0002	LAO LETTER PHO TAM
ຝ			

	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
ໆ	0EC6	0000 0000 0000 0000	LAO KO LA
ຯ	0EAF	0000 0000 0000 0000	LAO ELLIPSIS

Table 2: Collation Elements – Alphabetical Order