



# Lao Syllabification

---

Phonpasit PHISSAMAY

Nadir DURRANI

Lao Localization Project Team



# Outline

---

- Structure of Lao language
- Why do we need the syllabification?
- Syllabification algorism
- Research output



# Definition:

---

- Lines are broken as result of one of two conditions. The first condition is the **presence of an explicit line breaking character**. The second conditions result from a formatting algorithm having selected among available line breaking opportunities.
- The line breaking methodologies:
  - *Western* – spaces and hyphens are used to determine breaks
  - *East Asian* – line can break anywhere, unless prohibited
  - *South East Asian* – line breaks required morphological analysis



# Lao Script

---

- Laos has its own Lao letter alphabet where the root come from Sanskrit currently used for communicative language, and Buddhist language which is from the Pali alphabet used in writing Buddhism prayers.
- The Lao language is composed of 78 characters and signs:
  - 33 Consonants
  - 28 Vowels
  - 4 Tone marks
  - 3 Special symbols, and
  - 10 Lao digits



# Word in language

---

- Syllables: Syllable is a unit of spoken language, which may have a common meaning or not. A unit of spoken language may have one single syllable or many syllables
- One syllable may consist of:
  - Consonant and vowels
  - Consonant, vowel and alternate consonant
  - Consonant, vowel and tone mark
  - Consonant, vowel, alternate consonant and tone mark
- The orthography:
  - Calligraphy 1: C-V-altC-T, appear to be a right writing method in terms of linguistics such as pronunciation.
  - Calligraphy 2: C-V-T-altC, is natural from the standpoint of writing from left to right



# Structure of Lao language

---

- The structure of Lao language is completely different from English structure. The Lao characters appear at 4 different levels of one line, which may be grouped as follows:
  - Level 1: The character appearing in level 1 is of diacritic type. There are 5 diacritics
  - Level 2: Level 2 is occupied by superscript vowels only. There are 7 vowels
  - Level 3: This level is the main level of Lao word. There is always a character at level 3 at each position in a Lao word. All 33 consonants as well as 12 normal vowels and 2 special symbols are also at level 3.
  - Level 4: The characters appearing in level 4 are two lowered script vowels and one mixed consonant.



# Lao sentences

---

- The continuum in Lao Script leaves no clue for word processor to decide where to put break when end of line is reached.
- For Latin and several other scripts presence of space helps the word processor to decide where to break the line.



# Lao sentences

---

- Example for Latin

Lao characters originate from two sources Pali and Sanskrit and written Buddhist prayer from Tong Pali. Lao alphabets have their roots with Sanskrit currently used as communicative and Buddhist language from the Pali alphabet used in written Buddhism prayers.

Lao characters originate from two sources Pali and Sanskrit and written Buddhist prayer from Tong Pali. Lao alphabets have their roots with Sanskrit currently used as communicative and Buddhist language from the Pali alphabet used in written Buddhism prayers.





# Contd..

---

- Example for Lao

ກອມມູນິດພາຍໃນປະເທດຫວຽດນາມສ້າງຕັ້ງເປັນພັກກອມມູນິດຫວຽດນາມນັບແຕ່ນັບແຕ່ມີສ້າງຕັ້ງຂຶ້ນມາພັກກອມມູນິດຫວຽດນາໄດ້ພົບກັບສະພາບການອັນທຸລະການດານແລະກາ.



# Why is segmentation so Important?

---

- Segmentation lies at the heart of Script Processing
  - To Properly Display the Text.
  - To Sort Lao Words.
  - For Spell / Grammar Checker.
  - For all the advanced Localization applications like Speech Synthesizer, Machine Translation etc.



# Lao Collation

---

- Lao Collation is purely based on syllables i.e. comparison is done syllable by syllable if we find that first syllables are same we move on to next syllable.
- The Line Breaking Algorithm finds out the syllable boundaries accurately, moreover it provides information about what role a character is playing within syllable because some characters like 'ຸ' are hybrid in nature and play different roles when they act as consonants, combining consonant, vowels or consonantal.

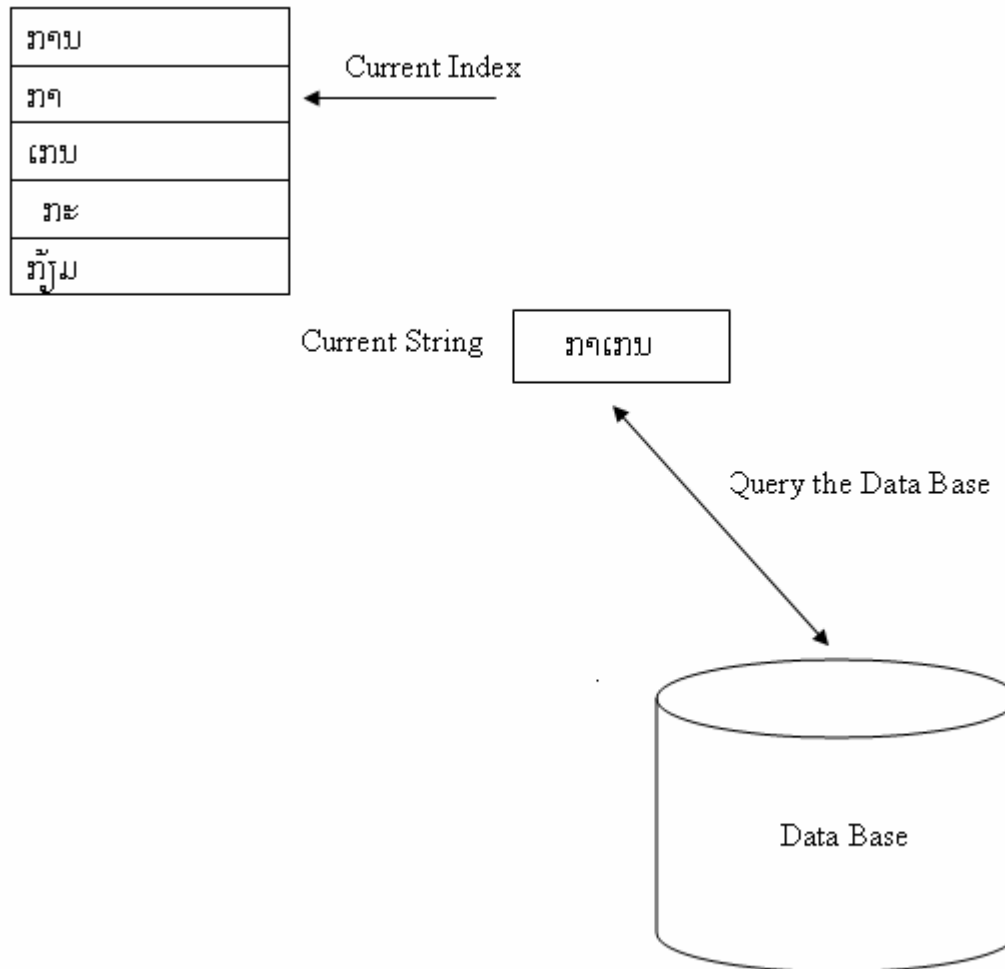


# Word Identification

---

- Detecting Word Boundaries
- Due to unending continuum of Lao Script it becomes very complicated to identify the word boundaries.
- We have to use brute force computing combining 1 character each time to the string and check if it is a valid word in dictionary.

# Contd.





## Contd.

---

- However if we already know the syllable boundaries we can improve the solution by appending one syllable each time rather than one character. This reduces the query time.
- This procedure could be further improved by employing syllable based statistical and probabilistic model.



# Spell Checking

---

- In order to do spelling check we need to detect word boundaries first.
- Syllable by syllable approach not only does that but also helps detecting the boundaries of erroneous word.
- We can do spell checking based on syllable rules initially and find out which part of word is error and then compare it with lexicon where we can store each syllable of a word separately. This helps finding the suggestion closest to error word.



# Lao syllabification algorithm.

---

- Lao Syllable Structure

		$t_{0,1}$		
		$v_{0,1}$		
$v_{0,1}$	$c_{0,1}$	$C$	$v_{0,1}$	$c_{0,4}$
		$c_{0,1}$		
		$v_{0,1}$		





# Combining 3 solutions

---

- Linguistic approach but less than 20%
  - Ex: Before vowels character is always defined as the beginning of syllable.
  - Ex: “<sup>ㄹ</sup>ㅅ” and “<sup>ㄹ</sup> ” is always defined as end of syllable.
- Statistic observation but less than 20%
  - Ex: ㄷ, ㄱ, ㅈ, ㅊ, ㅌ, ㅍ : never exist in the same syllable
  - Vowel in level 2 and level: never exist in the same syllable.
  - Tone marks: never exist in the same syllable.
- Input sequences, but less than 20%
  - The position of tone mark can defined the position of main consonant and the beginning of syllable
  - The position of some vowels can defined the position of main consonant and the beginning of syllable

[illegible]



## Contd..

---

- X0 represents a vowel which occurs before the nuclear consonant. It is can always defined the beginning of syllable.
- X1 is a combination consonant which comes before the nuclear consonant, only if nuclear consonant is one of {င, ခ, ခ, ဓ, မ, ဃ, ဣ}
- X is represents the nuclear consonants.



## Contd..

---

- X2 is a combination consonant which comes after the nuclear consonant, which placing under or next to the nuclear consonant.
- X3 is represents a vowel which occurs under the nuclear consonant.
- X4 is represents a vowel which occurs upper the nuclear consonant. However it is never exists with X3.



## Contd..

---

- X5 is represents a tone marks which occurs upper the nuclear consonant or upper vowels.
- X6 is represents consonant vowel, which occurs after nuclear consonant. It functions when the syllable doesn't have any vowels. And it always exists with X8.
- X7 is represents an after vowels. However X71 is always represents the end of syllable and it is never exist with tone mark.



## Contd..

---

- Based on this classification Lao Syllable structure can be more precisely defined as:

		X5					
		X4					
X0	X1	X	X6	X7	X8	X9	X10
		X2					
		X3					



# Syllabification Rules

---

- By using a set of conditions we can find out syllable boundaries.
- There rules are based on vowels.
- For Example: Rule for  $x_{02} = \llcorner X$ 
  - $\llcorner X = x_{02} (x_1)X(x_2)(x_5)(x_6)(x_8) (x_9:x_{10})$
  - $\llcorner X_{\text{z}} = x_{02} (x_1)X(x_2)x_{71}$
  - $\llcorner X_{\text{ö}} = x_{02} (x_1)X(x_2) x_{47} (x_5) x_8 (x_9:x_{10})$



For  $X_0 = x$

1.1  $x = \mathbf{X01} (X_1)\mathbf{X}(X_2)(X_5)(X_8)(X_9:X_{103})(X_{102})(X_{101})$

---

- $X_0(X_1)X = \text{ເຫງ}$
- $X_0X(X_2) = \text{ເຫວ}$
- $X_0X(X_2)(X_5) = \text{ເຫວ}^{'}$
- $X_0X(X_2)(X_5)(X_8) = \text{ເຫວ}^{'}$
- $X_0X(X_2)(X_5)(X_8)(X_9:X_{103}) = \text{ເຫວ}^{'}$
- $X_0X(X_2)(X_5)(X_8)(X_9:X_{103})(X_{102}) = \text{ເຫວ}^{'}$
- $X_0X(X_2)(X_5)(X_8)(X_9:X_{103})(X_{102})(X_{101}) = \text{ເຫວ}^{'}$





1.2  $\vec{X}$ ,  $\vec{X} = \mathbf{X01} (\mathbf{X1})\mathbf{X}(\mathbf{X2})\mathbf{X4}_{1-2}(\mathbf{X5})(\mathbf{X8}) (\mathbf{X9:X103})(\mathbf{X102})(\mathbf{X101})$

---

- $\mathbf{X01X(X2)X4}_{1-2}$  = ເຫລັ້
- $\mathbf{X01X(X2) X4}_{1-2}(\mathbf{X5})$  = ເຫລັ້
- $\mathbf{X01X(X2) X4}_{1-2}(\mathbf{X5})(\mathbf{X8})$  = ເຫລັ້
- $\mathbf{X01X(X2) X4}_{1-2}(\mathbf{X5})(\mathbf{X8})(\mathbf{X9:X10}_3)$  = ເຫລັ້
- $\mathbf{X01X(X2) X4}_{1-2}(\mathbf{X5})(\mathbf{X8})(\mathbf{X9:X10}_3)(\mathbf{X102})$  = ເຫລັ້
- $\mathbf{X01X(X2) X4}_{1-2}(\mathbf{X5})(\mathbf{X8})(\mathbf{X9:X10}_3)(\mathbf{X102})(\mathbf{X101})$  = ເຫລັ້



1.3 ເື້ອ, ເື້ອ = **X01** (X1)X(X2)**X4**<sub>3-4</sub>(X5) **X62** (X8)  
(X9:X103)(X102)(X101)

---

- **X01XX4**<sub>3-4</sub> **X62** = ເກື້ອ
- **X01XX4**<sub>3-4</sub> (X5) **X62** = ເກື້ອ
- **X01XX4**<sub>3-4</sub> (X5) **X62**(X8) = ເກື້ອນ
- **X01XX4**<sub>3-4</sub> (X5) **X62** (X8)(X9:X103) = ເກື້ອນສ໌
- **X01XX4**<sub>3-4</sub> (X5) **X62** (X8)(X9:X103)(X102) = ເກື້ອນສ໌ໆ
- **X01XX4**<sub>3-4</sub> (X5) **X62** (X8)(X9:X103)(X102)(X101) = ເກື້ອນສ໌ໆໆ



1.4 ឃេ, ឃាង = **X01 (X1)X(X2)(X7<sub>2</sub>)X71**

---

- **X01XX71=សេង**
- **X01X(X72)X71=សោង**
- **X01X(X2)X71=រោន**
- **X01X(X2)(X72)X71=រោនាង**



1.5 ເຮົາ = **X0<sub>1</sub>** (X1)**X**(X2) **X46** (X5) **X72**(X9:X103)(X102)(X101)

---

- **X01X(X2)X46X72**=ເຫລົ່າ
- **X01XX46X72**=ເປົ່າ
- **X01XX46(X5)X72**=ເປົ່າ
- **X01(X1)XX46X72**=ເຫງົ່າ
- **X01(X1)XX46(X5)X72**=ເຫງົ້າ



$$1.6 \text{ } \check{X}(x8) = \text{X01 (X1)X(X2) X47 (X5) X8 (X9:X103)(X102)(X101)}$$


---

- $\text{X0}_1 (\text{X1})\text{X X4}_7 (\text{X5})\text{X8}=\text{ເຫລັ້ນ}$
- $\text{X01 X X47 X8}=\text{ເສັ້ງ}$
- $\text{X01 X X47(X5) X8}=\text{ເສັ້ງ}$
- $\text{X01 (X1)X X47 X8(X9:X103)}=\text{ເຫລັ້ນສ໌}$
- $\text{X01 (X1)X X47 X8(X9:X103)(X102)}=\text{ເຫລັ້ນສ໌ໆ}$
- $\text{X01 (X1)X X47 X8(X9:X103)(X102)(X101)}=\text{ເຫລັ້ນສ໌ໆໆ}$



1.7 ເຊ, ເຊ້ = **X01** (X1)**X**(X2) (X47)(X5)**X63**(X9:X103)(X102)(X101)

---

- **X01(X1)XX63**=ເຫນຽ
- **X01(X1)(X47)XX63**=ເຫນຽ
- **X01X(X5)X63**=ເນຽ
- **X01X(X47)(X5)X63**=ເນຽ



For  $X_{02} = \mathbb{X}$

2.1  $\mathbb{X} = \mathbf{X}_{02} (X_1) \mathbf{X} (X_2)(X_5)(X_8) (X_9:X_{103}) (X_{102})(X_{101})$

---

- $X_{02}(X_1) \mathbf{X} = \text{ແຫງ}$
- $X_{02}\mathbf{X}(X_2) = \text{ແຫລ}$
- $X_{02}\mathbf{X}(X_2)(X_5) = \text{ແຫລ}'$
- $X_{02}\mathbf{X}(X_2)(X_5)(X_8) = \text{ແຫລ}''$
- $X_{02}\mathbf{X}(X_2)(X_5)(X_8)(X_9:X_{103}) = \text{ແຫລ}'''$
- $X_{02}\mathbf{X}(X_2)(X_5)(X_8)(X_9:X_{103})(X_{102}) = \text{ແຫລ}''''$
- $X_{02}\mathbf{X}(X_2)(X_5)(X_8)(X_9:X_{103})(X_{102})(X_{101}) = \text{ແຫລ}'''''$



## 2.2 ແຊ = **X02 (X1)X(X2)X71**

---

- **X02(X1)XX71**= ແຫວະ
- **X02X(X2)X71**= ແກວະ
- **X02XX71**= ແປະ





2.3 ແໜ້ນ(x8) = **X02** (X1)**X**(X2) **X47** (X5) **X8** (X9:X103) (X102)(X101)

---

- **X02 X X47 X8**= ແບັດ
- **X02XX47(X5)X8**=ແກ້ງ
- **X02XX47(X5)X8(X9:X103)**=ແກ້ງສ໌
- **X02XX47(X5)X8(X9:X103)(X102)**=ແກ້ງສ໌ໆ
- **X02XX47(X5)X8(X9:X103)(X102)(X101)**=ແກ້ງສ໌ໆໆ

For  $X03 = \text{ໂx}$

3.1  $\text{ໂx}, \text{ໂx໑} = \text{X03 (X1)X(X2)(X5)(X8) (X9:X103)(X102)(X101)}$

---

- $\text{X03(X1)X} = \text{ໂຫງ}$
- $\text{X03X(X2)} = \text{ໂຫລ}$
- $\text{X03X(X2)(X5)(X8)} = \text{ໂຫລ່ນ}$
- $\text{X03X(X2)(X5)(X8)(X9:X103)} = \text{ໂຫລ່ນສ໌}$
- $\text{X03X(X2)(X5)(X8)(X9:X103)(X102)} = \text{ໂຫລ່ນສ໌ໆ}$
- $\text{X03X(X2)(X5)(X8)(X9:X103)(X102)(X101)} = \text{ໂຫລ່ນສ໌ໆໆ}$



## 3.2 ໂຊຊະ = **X03** (X1)**X**(X2)**X71**

---

- **X03(X1)XX71**= ໂຫວະ
- **X03X(X2)X71**= ໂກວະ
- **X03XX71**= ໂປະ



### 3.3 ໂ໊໋ວ, ໂ໊໋ຍ = **X03** (X1)**X**(X2)**X47**(X5) **X8**<sub>3:8</sub>

---

- **X03(X1)XX47X83**=ໂຫງ້ຍ
- **X03(X1)XX47X88**=ໂຫງ້ວ
- **X03XX47(X5)X83**=ໂອ້ຍ
- **X03XX47(X5)X83**= ໂກ້ວ

For X04

$\text{X} = \text{X04} (\text{X1})\text{X}(\text{X2})(\text{X5})(\text{X61}) (\text{X9:x103})(\text{X102})(\text{X101})$

- ♦  $\text{X04} (\text{X1})\text{X} = \text{ໄຫວ}$
- ♦  $\text{X04} \text{X}(\text{X2}) = \text{ໄກວ}$
- ♦  $\text{X04} \text{X}(\text{X5}) = \text{ໄກ່}$
- ♦  $\text{X04} \text{X}(\text{X5})(\text{X61}) = \text{ໄກ່ວ}$
- ♦  $\text{X04} \text{X}(\text{X5})(\text{X61}) (\text{X9:x103}) = \text{ໄກ່ວສ໌}$
- ♦  $\text{X04} \text{X}(\text{X5})(\text{X61}) (\text{X9:x103})(\text{X102}) = \text{ໄກ່ວສ໌ໆ}$
- ♦  $\text{X04} \text{X}(\text{X5})(\text{X61}) (\text{X9:x103})(\text{X102}) (\text{X101}) = \text{ໄກ່ວສ໌ໆໆ}$



For  $X0_5$

$\text{ໃ}x = \text{X05} (X1)\text{X}(X2)(X5) (X61) (X9:X103)(X102)(X101)$

---

- ◆  $\text{X05} (X1)\text{X} = \text{ໃຫລ}$
- ◆  $\text{X05} \text{X}(X2) = \text{ໃຫລ}$
- ◆  $\text{X05} \text{X}(X5) = \text{ໃຂ່}$
- ◆  $\text{X05} \text{X}(X5) (X61) = \text{ໃຂ່ວ}$
- ◆  $\text{X05} \text{X}(X5) (X61) (X9:X103) = \text{ໃຂ່ວສ໌}$
- ◆  $\text{X05} \text{X}(X5) (X61) (X9:X103) )(X102) = \text{ໃຂ່ວສ໌ງ}$
- ◆  $\text{X05} \text{X}(X5) (X61) (X9:X103) )(X102) (X101) = \text{ໃຂ່ວສ໌ງຍ}$



For X3

$$X_{\text{9}} \& X_{\text{3}} = (X1)\mathbf{X}(X2)\mathbf{X3}_{1-2}(X5)(X8) (X9:X103)(X102)(X101)$$

---

- ◆  $\mathbf{X}(X2)\mathbf{X32} = \text{ຫລຸ}$
- ◆  $\mathbf{XX32}(X5) = \text{ຫຼ່}$
- ◆  $\mathbf{XX32}(X5) (X8) = \text{ຫຼ້ນ}$
- ◆  $\mathbf{XX32}(X5) (X8) (X9:X103) = \text{ຫຼ້ນສ໌}$
- ◆  $\mathbf{XX32}(X5) (X8) (X9:X103) )(X102) = \text{ຫຼ້ນສ໌ໆ}$
- ◆  $\mathbf{XX32}(X5) (X8) (X9:X103) )(X102) (X101) = \text{ຫຼ້ນສ໌ໆໆ}$

For  $X4_{1-4}$

$$\bar{X} \ \& \ \bar{X} \ \& \ \bar{X} \ \& \ \bar{X} = (X1)X(X2)X4_{1-4}(X5)(X8) (X9:X103)(X102)(X101)$$

- ◆  $(X1) X X4_{1-4} = \text{ຫລີ}$
- ◆  $X (X2)X4_{1-4} = \text{ຫວີ}$
- ◆  $X X4_{1-4}(X5) = \text{ຊີ}$
- ◆  $X X4_{1-4}(X5) (X8) = \text{ພິນ}$
- ◆  $X X4_{1-4}(X5) (X8) (X9:X103) = \text{ພິນສ໌}$
- ◆  $X X4_{1-4}(X5) (X8) (X9:X103) )(X102) = \text{ພິນສ໌ໆ}$
- ◆  $X X4_{1-4}(X5) (X8) (X9:X103) )(X102) (X101) = \text{ພິນສ໌ໆໆ}$





For X45

$$\overset{\circ}{X} = (X1)\textcolor{red}{X}(X2)\textcolor{red}{X45}(X5)(X72)(X9:X103)(X102)(X101)$$

---

- ◆  $(X1)\textcolor{red}{X} \textcolor{red}{X45} = \text{ຫວໍ}$
- ◆  $\textcolor{red}{X} (X2)\textcolor{red}{X45} = \text{ຂວໍ}$
- ◆  $\textcolor{red}{X} \textcolor{red}{X45}(X5) = \overset{!}{\text{ຂໍ}}$
- ◆  $\textcolor{red}{X} \textcolor{red}{X45}(X5) (X72) = \overset{!}{\text{ຂໍ້}}$
- ◆  $\textcolor{red}{X} \textcolor{red}{X45}(X5) (X72)(X9:X103) = \overset{!}{\text{ຂໍ້າສີ່}}$
- ◆  $\textcolor{red}{X} \textcolor{red}{X45}(X5) (X72)(X9:X103) (X102) = \overset{!}{\text{ຂໍ້າສີ່ໆ}}$
- ◆  $\textcolor{red}{X} \textcolor{red}{X45}(X5) (X72)(X9:X103) (X102) (X101) = \overset{!}{\text{ຂໍ້າສີ່ໆ}}$



For  $X4_6 = \hat{x}$

9.1  $\hat{x}_{(x8)} = (X1)X(X2)X46(X5)X8(X9:X103) (X102)(X101)$

---

- ◆  $(X1)X \ X46 = \text{ຫົວ}$
- ◆  $X (X2) \ X46 \ x8 = \text{ໂງ່}$
- ◆  $X \ X46 \ (X5) \ x8 = \text{ພີ່}$
- ◆  $X \ X46 \ (X5) \ x8 \ (X9:X103) = \text{ພີ່ສີ່}$
- ◆  $X \ X46 \ (X5) \ x8 \ (X9:X103) \ (X102) = \text{ພີ່ສັງ}$
- ◆  $X \ X46 \ (X5) \ x8 \ (X9:X103) \ (X102) \ (X101) = \text{ພີ່ສັງຍຸ}$



For  $X_4 = \hat{x}$

$\hat{x} = (X1)X(X2)X46 X61 X71$

---

◆  $(X1)XX46 X61 X71 = \text{ຫວີວະ}$

◆  $XX46 X61 X71 = \text{ຕົວະ}$

For X47


$$\tilde{X} = (X1)X(X2)X47(X5)(X61)X8(X9:X103) (X102)(X101)$$

---

- ◆  $(X1)X X47X8 =$  ຫ້ວດ
- ◆  $X (X2)X47(X5) X8 =$  ຫ້ວ□
- ◆  $X X47(X5) (X61)X8(X9:X103) =$  ຫ້ວນສ໌
- ◆  $X X47(X5) (X61)X8(X9:X103) (X102) =$  ຫ້ວນສ໌ໆ
- ◆  $X X47(X5) (X61)X8(X9:X103) (X102) (X101) =$  ຫ້ວນສ໌ໆໆ



For X6 ຂວ & ຂອ & ຂງ = (X1)**X**(X2)(X5)

**X6<sub>1-3</sub>X8**(X9:X103)(X102)(X101)

---

- **XX61X8**=ກວນ
- **XX62X8**=ກອນ
- **XX63X8**=ກຽນ
- **X(X5)X61X8**=ກວນ'
- **X(X5)X62X8**=ກອນ'
- **X(X5)X63X8**=ກຽນ'



For X71

$$x\text{ບ} = (X1)\text{X}(X2)\text{X71}$$

---

- $(X1)\text{X}(X2)\text{X71}$ =ຫວະ
- $\text{XX71}$ =ຄະ
- $\text{X}(X2)\text{X71}$ =ຟຣະ

## For X72



$x_1 = (X1)X(x2)(X5)X72 (X8)(X9:X103)(X102)(X101)$

---

- $(X1)XX72$ =ຫລາ
- $X(X5)X72$ =ດ້າ
- $X(X5)X72(X8)$ =ດ້ານ
- $X(X5)X72(X8)(X9:X103)$ =ດານສ໌
- $X(X5)X72(X8)(X9:X103)(X102)$ =ດານສ໌ໆ
- $X(X5)X72(X8)(X9:X103)(X102)(X101)$ =ດານສ໌ໆໆ



For X73

$x_{\text{ຳ}} = (X1)\textcolor{red}{X}(X2)(X5)\textcolor{red}{X73} (X9:X103)(X102)(X101)$

---

- $(X1)\textcolor{red}{XX73}$ =ຫວຳ
- $X(X2)(X5)\textcolor{red}{X73}$ =ຫວຳ'
- $\textcolor{red}{XX73}(X9:X103)$ =ພຳສ໌
- $\textcolor{red}{XX73}(X9:X103)(X102)$ =ພຳສ໌ໆ
- $\textcolor{red}{XX73}(X9:X103)(X102)(X101)$ = ພຳສ໌ໆໆ





## Contd..

---

- Before testing the conditions we need to classify all the characters in the input string and load them with all the possibilities.
- For example character '□', can be X,X2,X6 or X8 in a syllable.



## Contd..

---

- Before testing the conditions we need to classify all the characters in the input string and load them with all the possibilities.
- For example character ‘ঐ’ can be X,X2,X6 or X8 in a syllable.



# Algorithm

---

- For instance while processing the following text:

ໂຄງການ L10N ພັດທະນາພາສາລາວ

- Steps Involved
  - Put boundaries where ever you find punctuation marks or characters that do not belong to Lao Script.

- Look for possible re-ordering of text. For example user might type ٠ before ١.

[illegible]

# Contd..

- Use rules and conditions to define syllable boundaries.

Rule 3.1				Rule 13			Rule 10				Rule 12		Rule 13		Rule 13		Rule 13		Rule 13							
ŋ	n	ɲ		n	ŋ	u		u	õ	'	o		u	ŋ		u	ŋ		u	ŋ		u				
A0	A1	A2		A3	A4	A5		A7	A8	A9	A10		A11	A12		A14	A15		A16	A17		A18	A19	A20	A21	A22

- Break point character
  - Zero width space (U+200b)



# Research Output

---

- Lao Editor: Lao fonts, Lao keyboard, Lao ASCII to Unicode Converter, Syllable and Word Count, Line Break, Sorting, Find/Replace, Lexical, Dictionary, Spell Check and Lao menu interface.
- CD of Lao localization project:
  - Documentation: Reports, Technical Documents, and User guide book;
  - Software: Lao fonts, Lao Keyboard, Lao Editor;
  - Pictures: workshop, research work, and team
- Project website: [www.laol10n.info.la](http://www.laol10n.info.la)



Please enjoy demonstration

---