

Proposal for a Lao Script Root Zone LGR

LGR Version: 2

Date: 31 January, 2017

Document version: 3.1

Authors: Lao Script Generation Panel

Contents

1. General Information/ Overview/ Abstract.....	2
2. Script for which the LGR is proposed.....	2
3. Background on Script and Principal Languages Using It.....	2
4. Overall Development Process and Methodology.....	3
5. Repertoire	3
5.1. Consonants.....	3
5.2. Vowels.....	5
5.3. Tone Marks.....	6
5.4. Signs	6
5.5. Overall Lao Repertoire	7
6. Variants.....	9
6.1. Within-script variants.....	9
6.2. Cross-script variants.....	10
6.2.1. Lao and Thai.....	10
6.2.2. Lao and Khmer	11
7. Whole Label Evaluation Rules (WLE)	11
8. Contributors	12
9. References.....	12
Appendix A: Initial Analysis for Confusion at Code Point Level	14
Appendix B: Reference Figures for Code Points.....	16
Appendix C: Phoneme and Glyph Mapping Table.....	23

1. General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Lao LGR in the XML format and the rationale behind the design decisions taken. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document:

- Proposed-LGR-LaoScript-20170131.xml

Labels for testing can be found in the accompanying text document:

- Labels-LaoScript-20170131.txt

2. Script for which the LGR is proposed

ISO 15924 code: Laoo

ISO 15924 no.: 356

ISO 15924 English Name: Lao

Native name of the script: ອັກສອນລາວ

Maximal Starting Repertoire (MSR) version: MSR-2

3. Background on Script and Principal Languages Using It

The Lao script is used to write the Lao language, which is the official language of Laos. The Lao language belongs to the Tai-Kadai language family and is spoken by approximately 30 million people mainly in Laos, Isan - the north-eastern part of Thailand, neighboring Cambodia, China, Myanmar and Vietnam. The written form of the Lao language has been derived from Pali and Sanskrit languages. With the passage of time, adjustments have been made in the script and through various orthographic reforms; the Lao script has developed continuously. These reforms, in 1930s and 1960s, re-aligned Lao spelling of the “words according to their pronunciation. This especially concerned the spelling of Indic loanwords whose original pronunciation (and thus, their spelling) included a far greater range of syllable-final consonants than were found in the phonology of spoken Lao” [208] (pg. 29).

The Lao syllables are written around the main consonant. Vowels occur above, below, before, after or around this consonant. Lao is a tonal language, with six tones: high normal, low normal, mid, high falling, mid falling, and low rising. Combination of consonants and tone marks are used to indicate some of these tones¹.

¹ See <http://www.seasite.niu.edu/lao/LaoLanguage/LaoAlphabet/Laotones.htm> for more details on the tonal system.

One of the main characteristics of Lao writing is that words are written without spaces between them. The spaces are used to mark the end of sentences, paragraphs or clauses. Lao is written from left to right in horizontal lines.

Lao script and Thai script have many characters which have some similarity (see Appendix A), but Lao has fewer letters which are in a more rounded form, making them visually different from Thai.

The Lao dialect is differentiated into five main areas in Laos - Vientiane, Luang Prabang, Xieng Khuang, Khammuan and Champassak provinces - while Isan also has different dialects.

4. Overall Development Process and Methodology

The Lao GP has been formed by a working group with members having experience in linguistics, computational linguistics, Lao linguistics and the IDNs. The draft LGR proposal was initially developed at each stage by members of the Lao GP together with consultation from additional experts as needed. Their linguistic and technical feedback was incorporated. The proposal was finalized based on community feedback through an open-to-the-public workshop held in February 2016, where other members from academia, media, newspapers and government were also invited.

The Lao GP also undertook consultations with Thai GP and Khmer GP to discuss cross-script variants, and with Integration Panel for feedback on initial version of the work. The work has been facilitated by ICANN, as needed.

The proposal developed by the Lao GP was finally published for public comment by ICANN.

As a matter of notation, tags used in this document for defining classes of code points may use upper-case, but are in lower-case in the accompanying XML document.

5. Repertoire

The Lao grapheme cluster is based on the Lao phonological system, with the following basic *phonological* template:

C V (T) (Cf) (S⁺)

Where C represents the initial/onset cluster of consonant(s) in a syllable referred to as the “main consonant” cluster. The syllable requires a vowel (V) following the onset. However, it may be orthographically represented in different ways around the main consonant, as discussed in more detail below. If needed, the tone is also represented by a tone mark (T) and a final consonant (Cf) may also be included for closed syllables. A syllable may include one or more signs (S) at the end. Further details and constraints are discussed in the following sections.

5.1. Consonants

The following constraints on the **Consonants** in the syllable apply.

1. **Main Consonant (C).** Main consonant represents a cluster of consonants and is required for a syllable in the Lao script, with the following cases: (a) It contains at least one consonant; (b) It may also contain two consonants together, in which case second consonant (C2) is restricted to the set {ງ (0E87), ມ (0E8D), ນ (0E99), ມ (0EA1), ຮ (0EA3), ລ (0EA5), ວ (0EA7)}, e.g. ແຂວງ (province) or ກຮາມ (gram) and ຫລາຍ (many); (c) In addition the sequence of three consonants ຫງວ (0EAB 0E87 0EA7) is also possible, e.g. for writing Vietnamese names; (d) Finally, a consonant may also be followed by a semi-consonant mark ູ (0EBC). For example, ໂປຼແກຼມ (program) and ເອເລັກໂຕຼນິກ (electronic).

In summary, C is formed by the following way, however, read further below:

Consonant (C2 | 0EBC) or the sequence ຫງວ which is 0EAB 0E87 0EA7

where Consonant can be any of the consonants, | indicates “or” and () indicate optionality.

Consonants may also formulate sequences of arbitrary length in case of acronyms. For example: ສນຍ (PMO, for Prime Minister’s Office) and ບທສ (MPT, for Ministry of Post and Telecommunications). As any arbitrary sequence of consonants is possible due to the acronyms, the constraint on consonants discussed earlier can be further simplified to limiting the semi-consonant 0EBC to occur after a consonant. Other sequences in the earlier constraint are subsumed in the consonant sequences allowed possible for the acronyms. Thus, the C rule is simplified to any arbitrary sequence of consonants, optionally with a semi-consonant after it, as follows:

Consonant¹⁺ (0EBC)

where Consonant can be any of the consonants and “1+” indicates one or more in a sequence, or alternatively, 0EBC can only occur after a consonant.

Allowable combinations or consonants are listed in Appendix C, though the Lao GP considers these spelling rules which are not coded to simplify the label rules.

2. **Final Consonant (Cf).** A consonant may also occur at the end of a syllable. Some, but not all, consonants can occur in this position, as given in the following set: {ກ (0E81), ງ (0E87), ຊ (0E8A), ມ (0E8D), ດ (0E94), ທ (0E97), ນ (0E99), ບ (0E9A), ພ (0E9F), ມ (0EA1), ຮ (0EA3), ລ (0EA5), ວ (0EA7), ສ (0EAA)}. However, not all syllables are closed; open syllables are those that do not have a consonant in this final position. There can be a maximum of one final consonant in a syllable.

It has also been noted by the Lao GP that even more than three consonants can be arbitrarily combined to form an abbreviation.

5.2. Vowels

Most Lao vowels are combining marks and, as with all combining marks, placement should be restricted to contexts expected and supported by rendering engines, to avoid misleading or ambiguous presentation.

As to diphthongs, the Lao GP considers the full encoding restrictions (as given in item 5 below) will make the rules too complex to implement. Therefore, such limitations on diphthongs, although set out here, have not all been implemented. The Lao GP realizes that this “over-produces” Lao words; however, it considers this inadequacy to be a spelling issue. Allowing the additional sequences would not cause variant or other structural issues to arise. Therefore, control of correct spelling can be adequately managed by the applicant in the process of TLD label applications.

Vowels and diphthongs are enumerated in Appendix C. Even though phonologically a vowel always follows the main consonant, vowel can be represented by one or more graphemes and typed in one of the following ways, based on visual placement of the graphemes:

1. **A vowel-before**, typed before the main consonant C.
2. **A vowel-after**, typed after the main consonant C.
3. **A vowel-above**, typed after the main consonant C.
4. **A vowel-below**, typed after the main consonant C.
5. **A diphthong**, which is realized in one of the following ways:
 - i. A combination of two or three vowel graphemes, to represent vowels and diphthongs, as enumerated in Appendix C. For example, for three vowel graphemes, only the following combinations are possible: {Cxᵛ, C[̣], C[̇]}, where x or a dotted circle represents the location of the main consonant. Vowel-before, vowel-above and vowel-after are typed separately respectively and as per constraints in relation to the main consonant given in 1-4 above. Vowel-after is typed after the vowel-above. A tone mark, if needed, would occur after the vowel-above. Such syllables do not have Cf but may be followed by ໘ sign. This results in the structure: V-before C V-above (T) V-after (໘⁰⁻³). Note that the ໘⁰⁻³ indicates limiting the sign to occur no more than three times due to practical reasons.
 - ii. A combination of two vowel graphemes (vowel-before and vowel-above) with the consonant grapheme ໐ (0EAD). In this case the final consonant is required, in addition, and must be from the subset, Cf-sub: {ᵛ (0E81), ໓ (0EA1), ໐ (0E94), ໒ (0E99), ໕ (0E9A), ໘ (0E87), ໙ (0E8D) and ໐ (0EA7)}. For examples, diphthongs Cx[̇]໐, Cx[̣]໐; A tone mark may come after

- the vowel-above, if needed. Finally, such a syllable may be followed by ໘ sign (and not the ໐ sign). This results in the structure: V-before C V-above (T) ໐ Cf-sub (໘⁰⁻³).
- iii. Either of the consonants {໑ (0EAD), ໒ (0EA7)} may optionally combine with the vowel ໐ (0EB1) before it (without vowel-before or vowel-after). In this case a final consonant from the subset, Cf-sub: {໓ (0E81), ໔ (0EA1), ໕ (0E94), ໖ (0E99), ໗ (0E9A), ໘ (0E87), ໙ (0E8D) and ໐ (0EA7)} is required to close the syllable. For example, indicating the main consonant as x the following diphthongs are possible: x໒ (forming, e.g., x໒໓) and x໑ (forming, e.g., x໑໓); alternatively they may be preceded by the vowel-above x໐໑ (forming, e.g., x໐໑໓) and x໐໒ (forming, e.g., x໐໒໓). A tone mark may come after the vowel-above, if needed. In both cases, a tone mark can occur before the consonants {໑ (0EAD), ໒ (0EA7)} (after the main consonant and the (optional) vowel-above). Finally, such a syllable may be followed by ໘ sign (but not the ໐ (0ECC) sign). This results in the structure: C (໐) (T) ໑/໒ Cf-sub (໘⁰⁻³).
- iv. The vowel-above ໑ (0EBB) followed by the consonant ໒ (0EA7) and the vowel ໓ (0EB0). It does not allow a tone mark or final consonant but may be followed by ໘ sign (and not the ໐ sign). This results in the structure: C ໑ ໒ ໓ (໘⁰⁻³).
- v. The vowel-before ໔ (0EC0) and the vowel-after ໕ (0EB2) can also combine with the vowel ໓ (0EB0). It does not allow a tone mark or final consonant but may be followed by ໘ sign (and not the ໐ sign). This results in the structure: ໔ C ໕ ໓ (໘⁰⁻³). This limitation is implemented in Rule 5 later.

To formulate these vowels, specific combinations of vowel-before and/or vowel-above and/or vowel-after are used. But as noted at the beginning of the current section, the specific restrictions on the combinations of vowels and consonants valid for diphthongs have not been implemented, to avoid excess complexity in WLE rule-system, and these spelling issues can be managed by the applicant of the TLD label.

5.3. Tone Marks

Being a tonal language, Lao uses marks to represent some tones. A tone mark is not required for each syllable and there can be a maximum of one tone mark in a syllable. Tone marks used in the Lao language are listed in Appendix C. The tone mark occurs on the syllable, but is generally typed after the main consonant (in case of vowel-before or vowel-after with the consonant) or after the vowel-above or vowel-below, if these are used in the syllable.

5.4. Signs

A sign always comes at the end of syllable or can be followed by another sign. However, if there are multiple syllables in a label, a sign may occur in the middle of the label.

When the sign ໆ (0EC6) occurs after a syllable, its function is to repeat that syllable. For example, ອີ້ນອີ້ນ is written as ອີ້ນໆ. This sign can be repeated multiple times, to indicate multiple repetitions, e.g. ອີ້ນໆໆໆ. However, for practical considerations, it is suggested to limit the sign to three times (ໆ⁰⁻³) as more occurrences will be extremely rare. Moreover, as discussed, the repetition mark can occur at the end of a syllable (syl), but to simplify its use further, the Lao GP suggests that it be limited to end of a label at this time. Lao GP notes that this will not allow labels like syl1ໆ¹⁻³syl2syl3 or syl1syl2ໆ¹⁻³syl3 but only syl1syl2 syl3ໆ⁰⁻³, even though the former are possible and used sometimes. However, this would be an acceptable compromise to reduce complexity of WLE rules.

The Lao GP considered public comment response regarding the repetition mark and did not think it would create confusability. For example: Lao script community is able to differentiate the two domain labels with/without the repetition mark like ລາວລາວ and ລາວໆ (ລາວ = Lao) as these are visually distinct and can act as two different domain names. Lao GP decided to keep the repetition mark based on its active use by the community.

5.5. Overall Lao Repertoire

The table below illustrates in detail all the Lao code points to be used for the root zone. These are categorized based on the system discussed above.

S. No.	Unicode Code Point	Glyph	Name (aka other Lao Name)	Unicode General Category (gc)	Category	Language, with EGIDS value	Reference
1	0E81	ກ	LAO LETTER KO	Lo	Consonant, Cf	Lao, 1	[201] [204]
2	0E82	ຂ	LAO LETTER KHO SUNG	Lo	Consonant	Lao, 1	[201] [204]
3	0E84	ຄ	LAO LETTER KHO TAM	Lo	Consonant	Lao, 1	[201] [204]
4	0E87	ງ	LAO LETTER NGO	Lo	Consonant, C2, Cf	Lao, 1	[201] [204]
5	0E88	ຈ	LAO LETTER CO	Lo	Consonant	Lao, 1	[201] [204]
6	0E8A	ຊ	LAO LETTER SO TAM	Lo	Consonant, Cf	Lao, 1	[201] [204]
7	0E8D	ຍ	LAO LETTER NYO	Lo	Consonant, C2, Cf	Lao, 1	[201] [204]
8	0E94	ດ	LAO LETTER DO	Lo	Consonant, Cf	Lao, 1	[201] [204]
9	0E95	ຕ	LAO LETTER TO	Lo	Consonant	Lao, 1	[201] [204]
10	0E96	ຖ	LAO LETTER THO SUNG	Lo	Consonant	Lao, 1	[201] [204]
11	0E97	ທ	LAO LETTER THO TAM	Lo	Consonant, Cf	Lao, 1	[201] [204]
12	0E99	ນ	LAO LETTER NO	Lo	Consonant, C2, Cf	Lao, 1	[201] [204]
13	0E9A	ບ	LAO LETTER BO	Lo	Consonant, Cf	Lao, 1	[201] [204]
14	0E9B	ປ	LAO LETTER PO	Lo	Consonant	Lao, 1	[201] [204]
15	0E9C	ຜ	LAO LETTER PHO SUNG	Lo	Consonant	Lao, 1	[201] [204]

16	0E9D	ຳ	LAO LETTER FO TAM (aka LAO LETTER FO SUNG)	Lo	Consonant	Lao, 1	[201] [204]
17	0E9E	ຸ	LAO LETTER PHO TAM	Lo	Consonant	Lao, 1	[201] [204]
18	0E9F	ູ	LAO LETTER FO SUNG (aka LAO LETTER FO TAM)	Lo	Consonant, Cf	Lao, 1	[201] [204]
19	0EA1	ມ	LAO LETTER MO	Lo	Consonant, C2, Cf	Lao, 1	[201] [204]
20	0EA2	ຮ	LAO LETTER YO	Lo	Consonant	Lao, 1	[201] [204]
21	0EA3	ຮ	LAO LETTER LO LING (aka LAO LETTER LO RADA)	Lo	Consonant, C2, Cf	Lao, 1	[204]
22	0EA5	ູ	LAO LETTER LO LOOT (aka LAO LETTER LO LING)	Lo	Consonant, C2, Cf	Lao, 1	[201] [204]
23	0EA7	ວ	LAO LETTER WO	Lo	Consonant, C2, Cf	Lao, 1	[201] [204] [205]
24	0EAA	ຮ	LAO LETTER SO SUNG	Lo	Consonant, Cf	Lao, 1	[201] [204]
25	0EAB	ຫ	LAO LETTER HO SUNG	Lo	Consonant	Lao, 1	[201] [204]
26	0EAD	ອ	LAO LETTER O	Lo	Consonant	Lao, 1	[201] [204] [205]
27	0EAE	ຮ	LAO LETTER HO TAM	Lo	Consonant	Lao, 1	[201] [204]
28	0EB0	ຸ	LAO VOWEL SIGN A	Lo	Vowel-after	Lao, 1	[201] [205] [206]
29	0EB1	ຸ	LAO SEMIVOWEL SIGN MAI KAN	Mn	Vowel-above	Lao, 1	[201] [205] [206]
30	0EB2	ຸ	LAO VOWEL SIGN AA	Lo	Vowel-after	Lao, 1	[201] [205] [206]
31	0EB4	ຸ	LAO VOWEL SIGN I	Mn	Vowel-above	Lao, 1	[201] [205] [206]
32	0EB5	ຸ	LAO VOWEL SIGN II	Mn	Vowel-above	Lao, 1	[201] [205] [206]
33	0EB6	ຸ	LAO VOWEL SIGN Y	Mn	Vowel-above	Lao, 1	[201] [205] [206]
34	0EB7	ຸ	LAO VOWEL SIGN YY	Mn	Vowel-above	Lao, 1	[201] [205] [206]
35	0EB8	ຸ	LAO VOWEL SIGN U	Mn	Vowel-below	Lao, 1	[201] [205] [206]
36	0EB9	ຸ	LAO VOWEL SIGN UU	Mn	Vowel-below	Lao, 1	[201] [205] [206]
37	0EBB	ຸ	LAO SEMI VOWEL SIGN MAI KONG	Mn	Vowel-above	Lao, 1	[205]
38	0EBC	ຸ	LAO SEMI VOWEL SIGN LO (aka LAO SEMI CONSONANT LO)	Mn	Semi-Consonant	Lao, 1	[201] [205] [206]
39	0EBD	ຸ	LAO SEMI VOWEL SIGN NYO (aka LAO SEMIVOWEL IA)	Lo	Vowel-after	Lao, 1	[201] [205]

40	0EC0	Ꞩ	LAO VOWEL SIGN E	Lo	Vowel-before	Lao, 1	[201] [205] [206]
41	0EC1	ꞩ	LAO VOWEL SIGN EI	Lo	Vowel-before	Lao, 1	[201] [205] [206]
42	0EC2	Ɦ	LAO VOWEL SIGN O	Lo	Vowel-before	Lao, 1	[201] [205] [206]
43	0EC3	Ɜ	LAO VOWEL SIGN AY	Lo	Vowel-before	Lao, 1	[201] [205] [206]
44	0EC4	Ɡ	LAO VOWEL SIGN AI	Lo	Vowel-before	Lao, 1	[201] [205] [206]
45	0EC6	Ɬ	LAO KO LA (aka LAO MAY SAM)	Lm	Sign	Lao, 1	[203]
46	0EC8	◌◌̆	LAO TONE MAI EK	Mn	Tone-mark	Lao, 1	[202]
47	0EC9	◌◌̇	LAO TONE MAI THO	Mn	Tone-mark	Lao, 1	[202]
48	0ECA	◌◌̈	LAO TONE MAI TI	Mn	Tone-mark	Lao, 1	[202]
49	0ECB	◌◌̉	LAO TONE MAI CATAWA (aka LAO TONE MAI JATTAWA)	Mn	Tone-mark	Lao, 1	[202]
50	0ECC	◌◌̊	LAO CANCELLATION MARK (aka LAO CANCELLATION MARKMAI KA LAN)	Mn	Sign	Lao, 1	[207]
51	0ECD	◌◌̋	LAO NIGGAHITA (aka LAO VOWEL SIGN OR)	Mn	Vowel-above	Lao, 1	[201] [205] [206]
52	0EB2 0EB0	ꞨꞨ	LAO VOWEL SIGN AA LAO VOWEL SIGN A			Lao, 1	[205]

Note, all code points are defined in Unicode 1.1 [0]. Further, note that category C2 is listed for information only, it does not form part of the formal definition of the LGR, as discussed in Section 5.1.

6. Variants

6.1. Within-script variants

The Lao code point repertoire does not contain within-script variants. The only potential variant sequence is ◌◌̋ (0ECD 0EB2) which can also be written as ◌◌̋ (0EB3). However, ◌◌̋ (0EB3) has been excluded from MSR-2 so this case will not occur.

Vowel-above, vowel-below, semi-consonant and tone mark may form visually identical label even when typed in different order in a label in some fonts.

1. Vowels

- a. Vowel-above { ◌̋ 0EB1 ◌̆ 0EB4 ◌̇ 0EB5 ◌̈ 0EB6 ◌̉ 0EB7 ◌̊ 0EBB ◌̋ 0ECD }

- b. Vowel-below { ◌̣ 0EB8 and ◌̤ 0EB9 }
- 2. Semi-consonant { ◌̥ 0EBC }
- 3. Tone marks { ◌̇ 0EC8 ◌̈̇ 0EC9 ◌̉ 0ECA ◌̊ 0ECB }

For example, ັ (0E99 0EB5 0EC9) can also be written as ັ (0E99 0EC9 0EB5). Therefore, 0EB5 0EC9 can be considered as a variant of 0EC9 0EB5. This variable sequencing is not consistently supported by all fonts and systems. After due consideration and discussion in the public workshop, the Lao Generation Panel recommends that only the valid sequence should be allowed using WLE rules and other possible sequences should not be valid, and not considered variant sequences.

The valid sequence is semi-consonant, then vowel-above or vowel-below, and finally the tone mark.

6.2. Cross-script variants

There are some similarities between Lao and other languages in South East Asia like Thai and Khmer. The Lao GP did a detailed analysis of such cross-script variants, which is listed in the mapping in Appendix A. A second pass of analysis was done, focused on more stringent homoglyphic criteria, based on which the following sets were derived.

6.2.1. Lao and Thai

The Lao script is similar with Thai. As per GP consideration, there are some code points which are similar with each other, most of which are vowels and tone marks. These are given below.

Lao Code Points

0E88 ◌̣ LAO LETTER CO
 0EB0 ◌̥ LAO VOWEL SIGN A
 0EB1 ◌̈̇ LAO VOWEL SIGN MAI KAN
 0EC2 ◌̉ LAO VOWEL SIGN O
 0EC3 ◌̊ LAO VOWEL SIGN AY
 0EB8 ◌̣ LAO VOWEL SIGN U
 0EB9 ◌̤ LAO VOWEL SIGN UU
 0EC8 ◌̇ LAO TONE MAI EK
 0EC9 ◌̈̇ LAO TONE MAI THO
 0ECA ◌̉ LAO TONE MAI TI
 0ECB ◌̊ LAO TONE MAI JATTAWA
 0ECC ◌̋ LAO CANCELLATION MARK

Thai Code Points

0E08 ◌̣ THAI CHARACTER CHO CHAN
 0E30 ◌̥ THAI CHARACTER SARA A
 0E31 ◌̈̇ THAI CHARACTER MAI HAN-AKAT
 0E42 ◌̉ THAI CHARACTER SARA O
 0E43 ◌̊ THAI CHARACTER AI MAIMUAN
 0E38 ◌̣ THAI CHARACTER SARA U
 0E39 ◌̤ THAI CHARACTER SARA UU
 0E48 ◌̇ THAI CHARACTER MAI EK
 0E49 ◌̈̇ THAI CHARACTER MAI THO
 0E4A ◌̉ THAI CHARACTER MAI TRI
 0E4B ◌̊ THAI CHARACTER MAI CHATTAWA
 0E4C ◌̋ THAI CHARACTER THANTHAKHAT

However, these are still not strict homoglyphs and have different shapes. A Lao reader can tell Thai and Lao apart. Further, as there is only one consonant which is similar (and marks cannot form a label without a consonant), it may not be possible to create any significant number of TLD labels

which are confusable. Thus, the Lao GP has agreed not to define any cross-script variants (considering that Thai and Lao code points cannot be mixed within a label).

6.2.2. Lao and Khmer

The Lao GP has only found two marks which are similar to those in Khmer.

Lao Code Points

0EB8 ◌ LAO VOWEL SIGN U
0EB9 ◌ LAO VOWEL SIGN UU

Khmer Code Points

17BB ◌ KHMER VOWEL SIGN U
17BD ◌ KHMER VOWEL SIGN UU

As there are no consonants which are similar, these marks are not considered as variant because they cannot form a label by themselves.

7. Whole Label Evaluation Rules (WLE)

The aim of the WLE rules in Lao is to manage the context of the combining marks and special combination of other characters, e.g. the repetition sign. Though the rules are written to preserve the inherent structure of the written syllable, the motivation of these rules is that if the contexts are not managed, the resulting labels may not be rendered properly causing unpredictable behavior. Where unpredictable behavior is not expected, the over-production is not controlled or structure is not managed, as these are considered language-based spelling constraints, which are not relevant for label formulation.

For consonants, there is a single rule, which enforces the ordering of the semi-consonant mark.

1. The semi-consonant 0EBC follows a Consonant. A sequence of consonants with the optional semi-consonant form the main consonant cluster C.

For vowels, the following rules are needed to ensure the grapheme structure and proper ordering. Variation in this ordering is not licensed for TLDs, as discussed in the section on within-script variants.

2. A vowel-before precedes the main consonant cluster C.
3. A vowel-above or a vowel-below follows the main consonant C.
4. A vowel-after follows the main consonant C or a tone mark or a vowel-above.
5. The sequence ๓ (0EB2 0EB0) follows the main consonant C and ๐ (0EC0) to form ๐๓.

For tone marks, again a rule is needed to enforce the correct ordering.

6. A tone mark follows the main consonant C or a vowel-above or a vowel-below.

Additional rules are needed to manage the context of signs.

7. The sign ໐ (OECC) follows Cf.
8. The sign ໑ (OEC6) can only occur 0-3 times at the end of a label. See relevant discussion in Section 5.4.

8. Contributors

The Lao Generation Panel included the following members:

1. Mr. Phonpasit Phissamay, Chair
2. Mr. Khamphanh Souvannakha, Co-chair
3. Mr. Valaxay Dalaloy
4. Prof. Dr. Bualy Paphaphanh
5. Mr. Thonglor Duangsavanh
6. Mr. Bounmy Kongmany
7. Mr. Sengfa Holanouphab
8. Mrs. Chittaphone Chansylilath
9. Mr. Phouthong Sisavath
10. Mr. Khamphay Inthara
11. Mr. Phousana Silivong
12. Mr. Saysomvang Souvannavong
13. Ms. Phavanhna Douangboupha

ICANN Staff:

14. Sarmad Hussain
15. Rida Hijab Basit

9. References

[0] The Unicode Standard 1.1, The Unicode Consortium, Mountain View, CA. 1993

[101] Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013) <http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf>

[102] Integration Panel "Requirements for LGR Proposals from Generation Panels" available online as <https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-20150424.pdf>

[103] The Unicode Consortium, Unicode Character Database, available online as <http://www.unicode.org/Public/UCD/latest/>

[201] Lao grammar book published by the Ministry of Education in 1967, see Appendix B, Figure 1.

[202] Lao grammar book published by the Ministry of Education in 1967, see Appendix B, Figure 2.

[203] Lao grammar book published by the Ministry of Education in 1967, see Appendix B, Figure 3.

[204] Lao grammar book published by the Ministry of Education in 2000, see Appendix B, Figure 4.

[205] Lao grammar book published by the Ministry of Education in 2000, see Appendix B, Figure 5.

[206] Lao grammar book published by the Ministry of Education in 2000, see Appendix B, Figure 6.

[207] Lao grammar 1935, see Appendix B, Figure 7.

[208] Enfield, N. J. A Grammar of Lao. Mouton de Gruyter, Berlin and New York, 2007.

Appendix A: Initial Analysis for Confusion at Code Point Level

No	Lao script	Thai	Khmer
1	ງ	-	-
2	ຈ	จ	-
3	ຊ	-	-
4	ຍ	ย	-
5	ດ	ด	-
6	ຕ	ต	-
7	ບ	บ	-
8	ປ	ป	-
9	ຝ	ฝ	-
10	ພ	พ	-
11	ຟ	ฟ	-
12	ວ	ว	-
13	ຫ	-	-
14	ະ	ะ	-
15	๕	๕	໕
16	ັ	ັ	-
17	๘	๘	-
18	າ	า	ា
19	ໂ	โ	-
20	ໍາ	ໍာ	-
21	ໃ	ใ	-
22	ໄ	ไ	-

23	ງ	າ	ງ-
24	ອ	-	-
25	ຸ	ຸ	ຸ
26	ອຸ	ອຸ	ອຸ
27	ໂ	ໂ	ໂ
28	ອ້	ອ້	-
29	ອຶ	ອຶ	-
30	ອ່	ອ່	ອ່
31	ອຸ້	ອຸ້	-
32	ອຸ່	ອຸ່	ອຸ່
33	ງ	-	-

Appendix B: Reference Figures for Code Points

In the following figures, code points and sequence from the repertoire are identified with an annotation that shows their serial number from the table in Section 5.

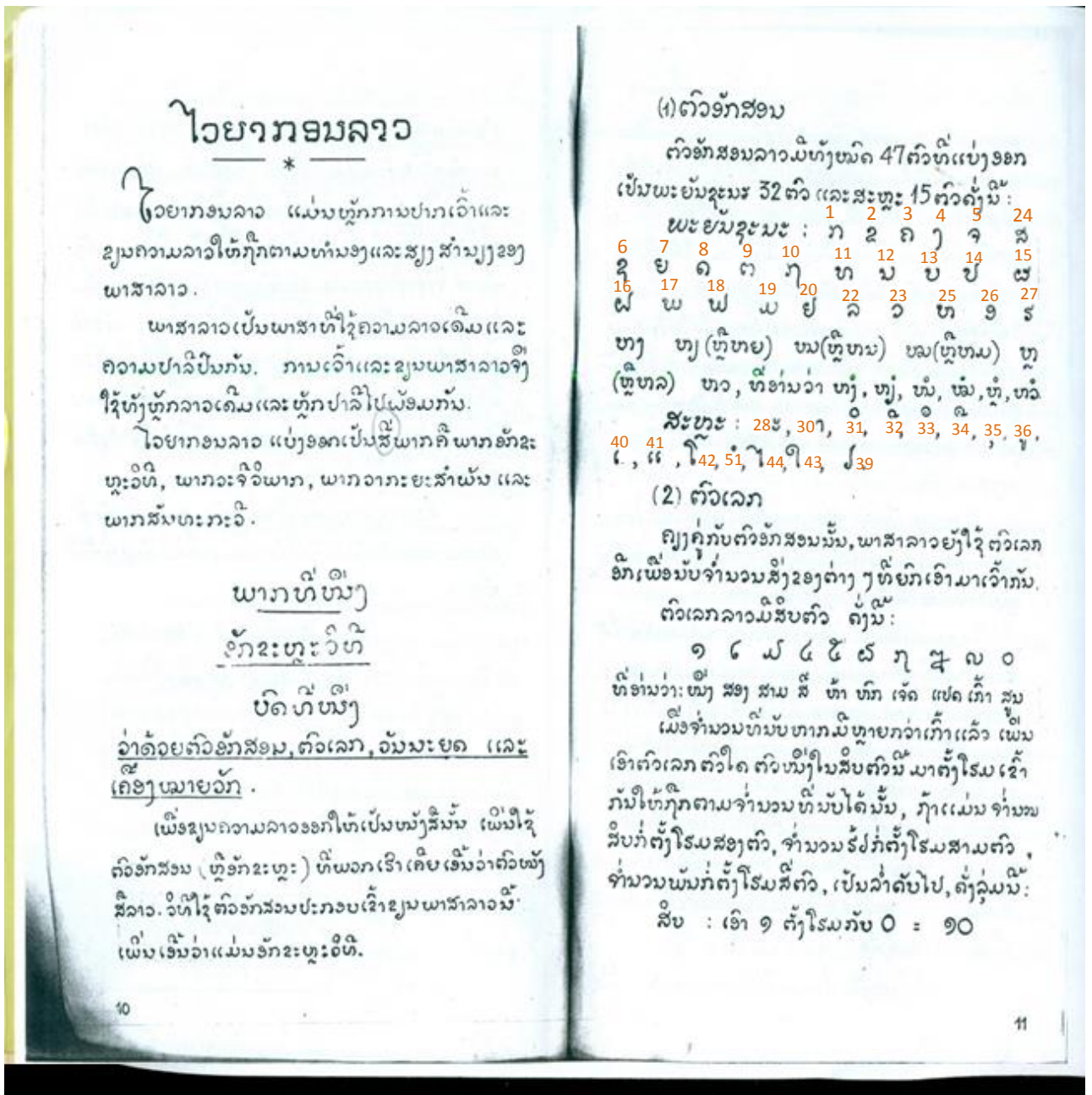


Figure1: Vowels and consonants illustrated in Lao grammar book published by the Ministry of Education in 1967

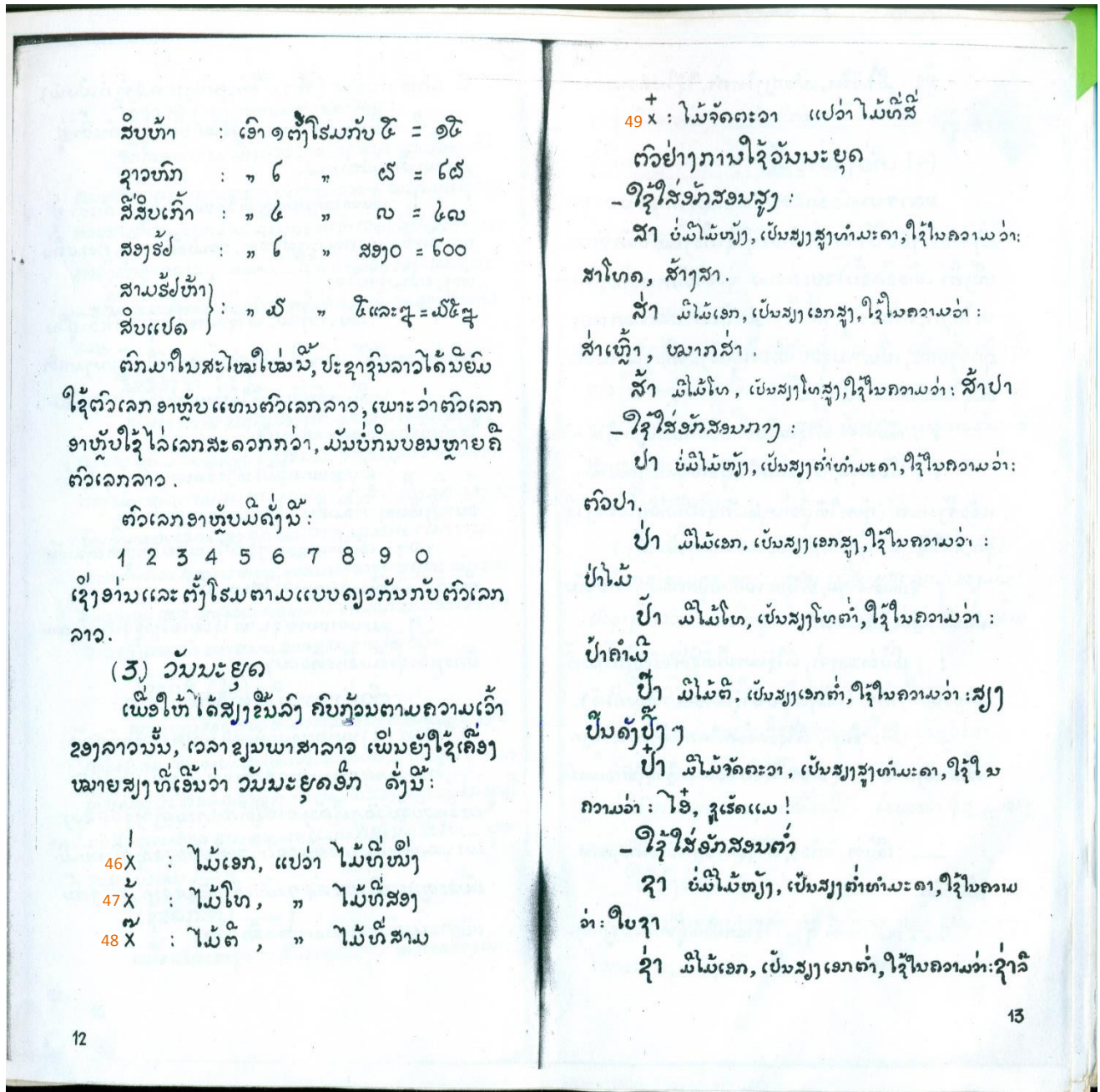


Figure 2: Tone marks illustrated in Lao grammar book published by the Ministry of Education in 1967

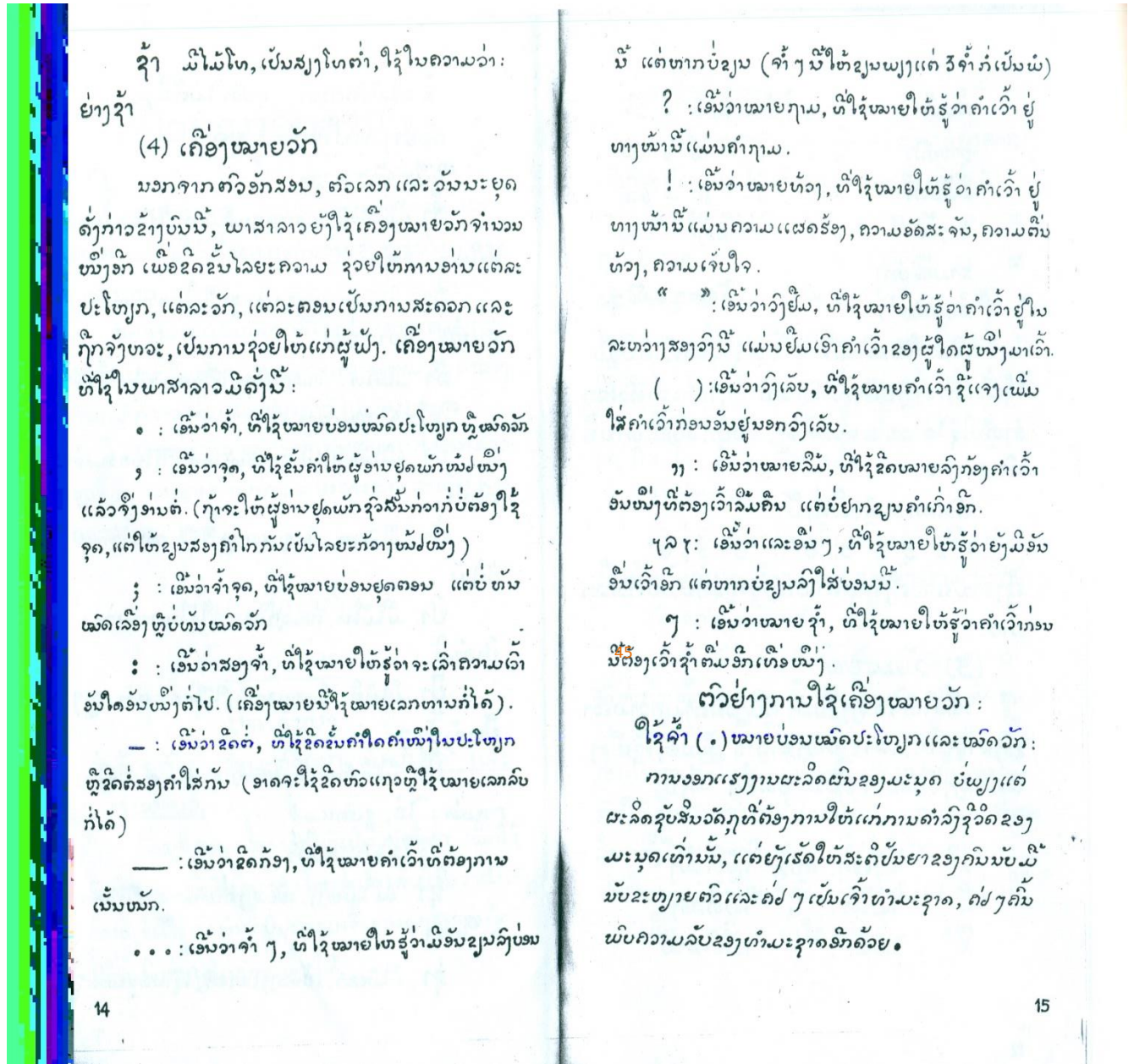
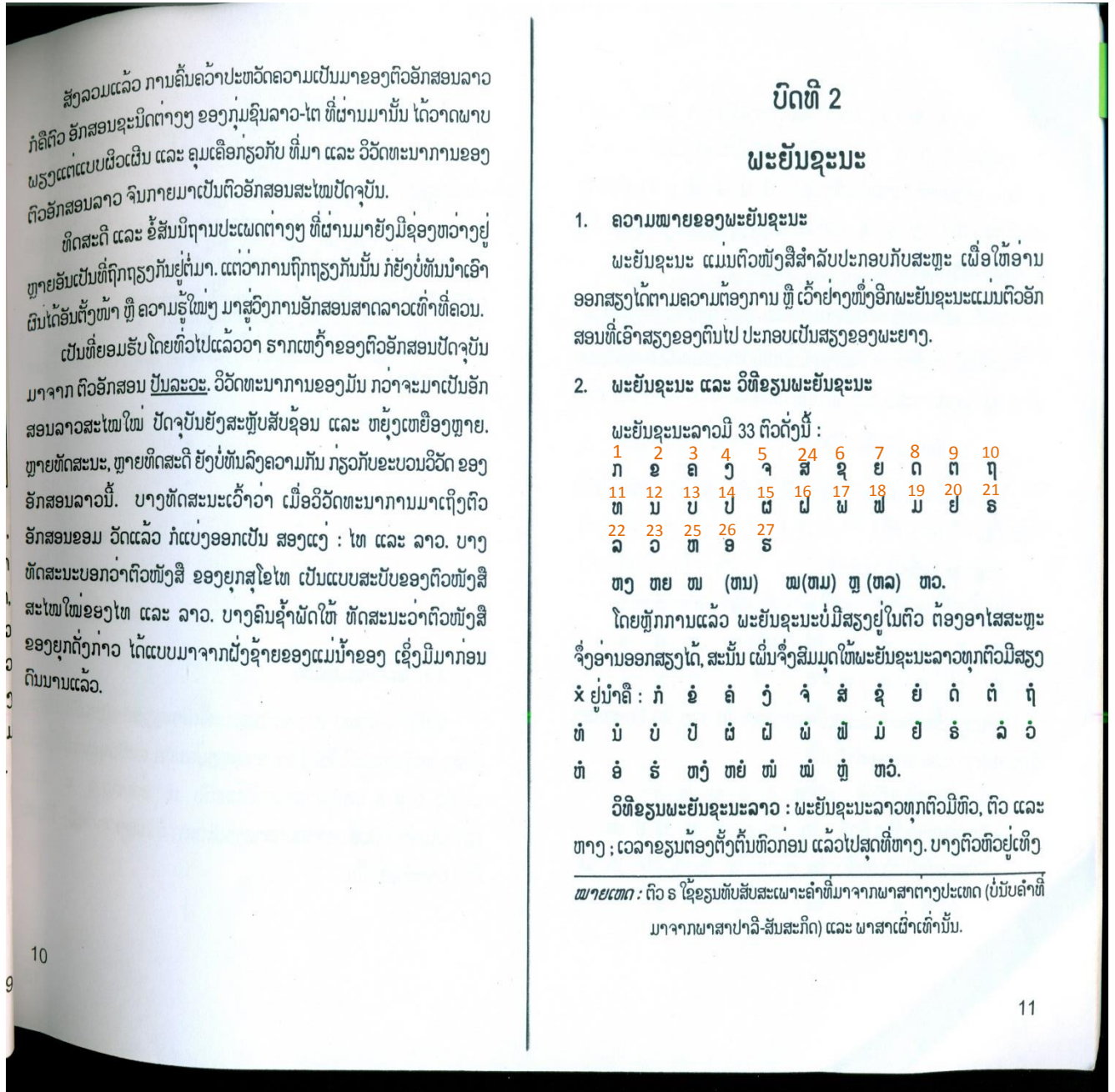


Figure 3: Lao Ko La sign illustrated in Lao grammar book published by the Ministry of Education in 1967



ບົດທີ 2

ພະຍັນຊະນະ

1. ຄວາມໝາຍຂອງພະຍັນຊະນະ
ພະຍັນຊະນະ ແມ່ນຕົວໜັງສືສຳລັບປະກອບກັບສະຫຼະ ເພື່ອໃຫ້ອ່ານ ອອກສຽງໄດ້ຕາມຄວາມຕ້ອງການ ຫຼື ເວົ້າຢ່າງໜຶ່ງອີກພະຍັນຊະນະແມ່ນຕົວອັກ ສອນທີ່ເອົາສຽງຂອງຕົນໄປ ປະກອບເປັນສຽງຂອງພະຍາງ.
2. ພະຍັນຊະນະ ແລະ ວິທີຂຽນພະຍັນຊະນະ

ພະຍັນຊະນະລາວມີ 33 ຕົວດັ່ງນີ້ :

1	2	3	4	5	24	6	7	8	9	10
ກ	ຂ	ຄ	ງ	ຈ	ສ	ຊ	ຍ	ດ	ຕ	ຖ
11	12	13	14	15	16	17	18	19	20	21
ທ	ນ	ບ	ປ	ຜ	ຝ	ພ	ຟ	ມ	ຢ	ຮ
22	23	25	26	27						
ລ	ວ	ຫ	ອ	ຮ						

ຫງ ຫຍ ຫໝ (ຫນ) ຫພ(ຫມ) ຫຫ (ຫລ) ຫວ.

ໂດຍຫຼັກການແລ້ວ ພະຍັນຊະນະບໍ່ມີສຽງຢູ່ໃນຕົວ ຕ້ອງອາໄສສະຫຼະ ຈຶ່ງອ່ານອອກສຽງໄດ້, ສະນັ້ນ ເພິ່ນຈຶ່ງສົມມຸດໃຫ້ພະຍັນຊະນະລາວທຸກຕົວມີສຽງ ຂຸ່ມນຳຄື : ກໍ ຂໍ ຄໍ ງໍ ຈໍ ສໍ ຊໍ ຍໍ ດໍ ຕໍ ຖໍ ທໍ ນໍ ບໍ ປໍ ຜໍ ຝໍ ພໍ ຟໍ ມໍ ຢໍ ຮໍ ລໍ ວໍ ຫໍ ອໍ ຮໍ ຫງໍ ຫຍໍ ຫໝໍ ຫພໍ ຫຫໍ ຫວໍ.

ວິທີຂຽນພະຍັນຊະນະລາວ : ພະຍັນຊະນະລາວທຸກຕົວມີຫົວ, ຕົວ ແລະ ຫາງ ; ເວລາຂຽນຕ້ອງຕັ້ງຕົ້ນຫົວກ່ອນ ແລ້ວໄປສຸດທີ່ຫາງ. ບາງຕົວຕົວຢູ່ເທິງ

ໝາຍເຫດ : ຕົວ ຮ ໃຊ້ຂຽນທັບສັບສະເພາະຄຳທີ່ມາຈາກພາສາຕ່າງປະເທດ (ບໍ່ນັບຄຳທີ່ ມາຈາກພາສາປາລີ-ສັນສະກິດ) ແລະ ພາສາເຜົ່າເທົ່ານັ້ນ.

Figure 4: Consonants illustrated in Lao grammar book published by the Ministry of Education in 2000

42	9.	ໂ	"	ໂອ
51	10.	ໂ	"	ອໍ
44	11.	ໂ	"	ໂອໂມ້ມາຍ
43	12.	ໃ	"	ໂອໂມ້ມ່ວນ
29	13.	ໂ	"	ໂມ້ກັນ
37	14.	ໂ	"	ໂມ້ກົງ
39	15.	ຮ	"	ຕົວ ຍ ເພື່ອງ
23	16.	ວ	"	ຕົວ ວ
26	17.	ອ	"	ຕົວ ອ

3. ສຽງສະຫຼະ

ສຽງສະຫຼະເປັນສຽງທີ່ເບິ່ງອອກມາຈາກຮູດໍຕາມທໍາມະຊາດໂດຍບໍ່ມີສິ່ງກົດຂວາງຫຼາຍເໝືອນສຽງພະຍັນຊະນະ. ໃນພາສາລາວ ມີສະຫຼະ 28 ຕົວແຕ່ມີສຽງຕ່າງກັນ 27 ສຽງຄື :

ເຄື່ອງໝາຍ	ສຽງ	ເຄື່ອງໝາຍ	ສຽງ
28 ະ	ອະ	40 ເຂະ	ເອະ
30 າ	ອາ	ເຂ	ເອ
31 ຶ	ອິ	41 ແຂະ	ແອະ
32 ື	ອີ	ແຂ	ແອ
33 ື	ອີ	42 ໂຂະ	ໂອະ
34 ື	ອີ	ໂຂ	ໂອ
35 ຸ	ອຸ	ເຂາະ 52	ເອາະ
36 ູ	ອູ	51 ັ	ອໍ

ເຄື່ອງໝາຍ	ສຽງ	ເຄື່ອງໝາຍ	ສຽງ
ເຂີ	ເອີ	ເຂີອ	ເອີອ
ເຂີ	ເອີ	ເຂີອ	ເອີອ
ເຂັຽ	ເອັຽ	ໂຂ	ໂອ
ເຂຽ	ເອຽ	ໂຂ	ໂອ
ເຂວະ	ເອວະ	ເຂົາ	ເອົາ
ເຂົວ	ເອົວ	ເຂົາ	ເອົາ

ສຽງສະຫຼະ 1 ສຽງ ສາມາດແທນໃຫ້ຮູບສະຫຼະໜ້ອຍ ຫຼື ຫຼາຍຮູບຕ່າງກັນດັ່ງນີ້ :

1) ເຄື່ອງໝາຍ 1 ຮູບ ແທນ 1 ສຽງຄື :

ເຄື່ອງໝາຍ	ສຽງ	ເຄື່ອງໝາຍ	ສຽງ
28 ະ	ອະ	ໂ 36	ອູ
30 າ	ອາ	ເ 40	ເອ
31 ຶ	ອິ	ໂ 42	ໂອ
32 ື	ອີ	ໂ 51	ອໍ
35 ຸ	ອຸ		

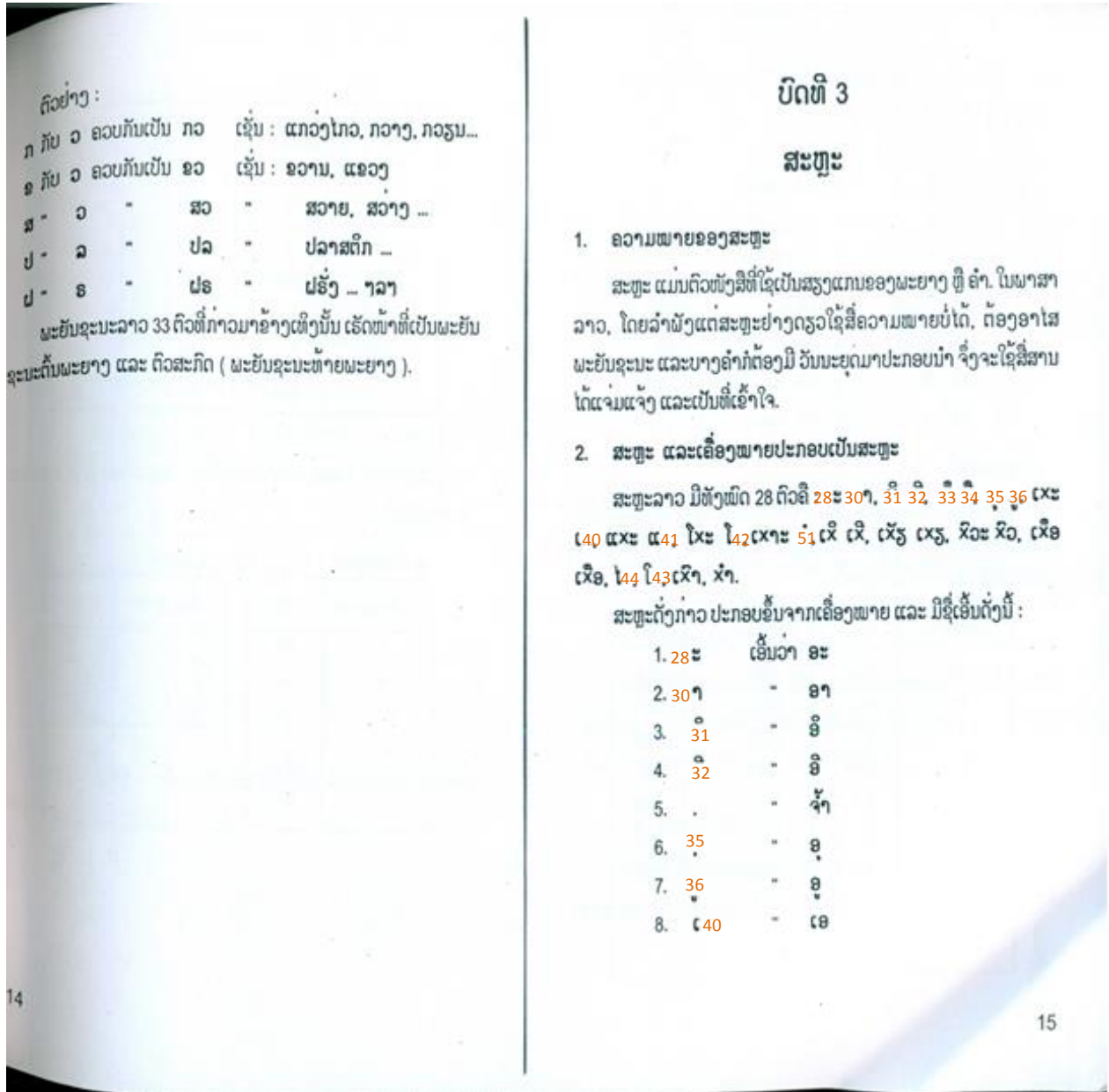


Figure 6: Vowels illustrated in Lao grammar book published by the Ministry of Education in 2000

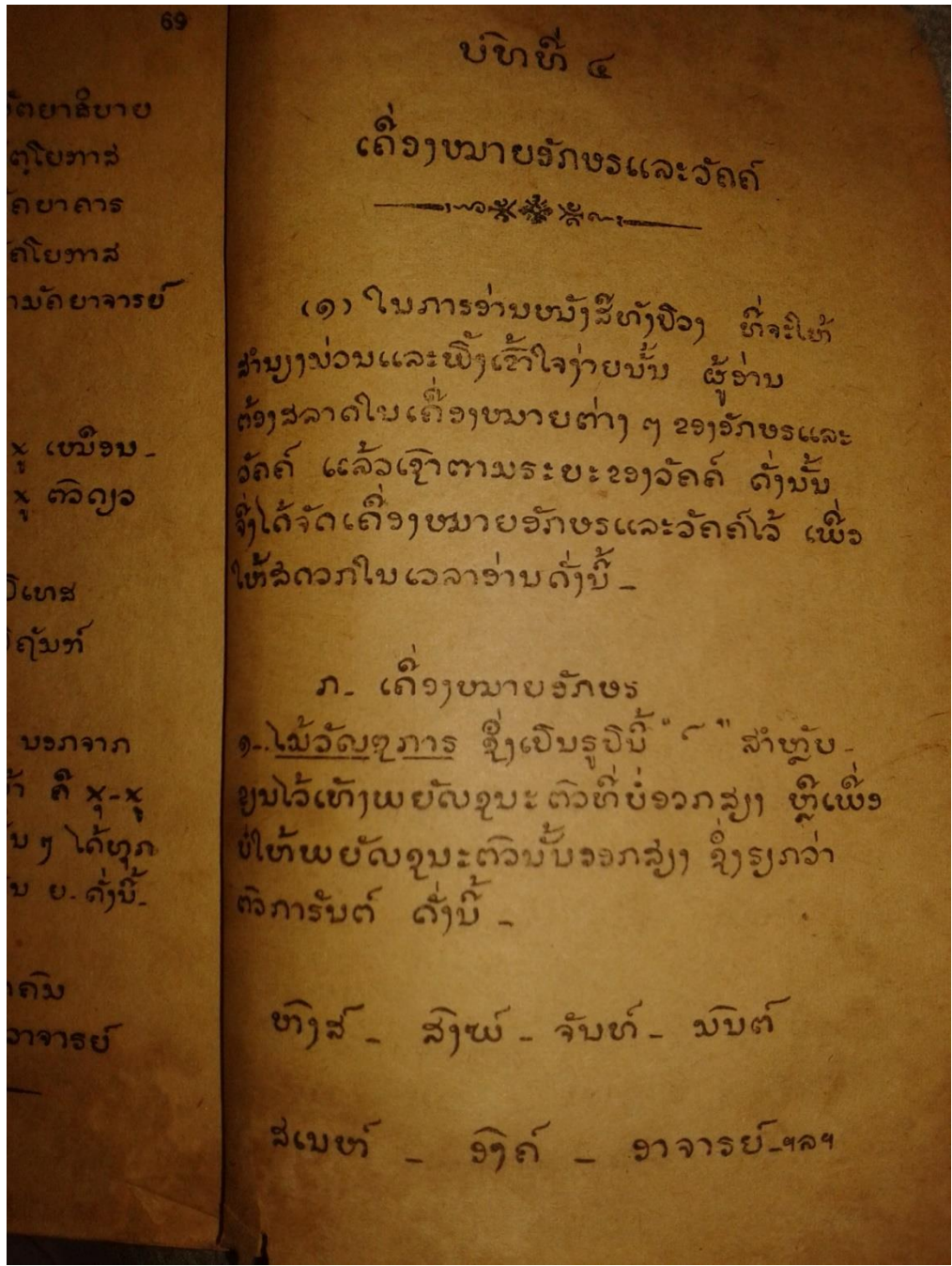


Figure 7: Cancellation mark illustrated in Lao grammar book published in 1935

Appendix C: Phoneme and Glyph Mapping Table

This appendix lists the consonants, vowels and their combinations which can be used to represent consonants, consonant clusters, vowels and diphthongs occurring at various positions in a syllable of Lao, as per details discussed in Section 5. This table has been developed by the Lao Text-to-Speech development team at the e-Government Center of Ministry of Posts and Telecommunications of Laos.

Some, but not all the, restrictions have been encoded in the WLE rules proposed, to manage the complexity of these WLE rules. The remaining restrictions are considered spelling rules to be managed by the registrant. See Section 5 for details.

Consonant _{initial}				Vowel				Consonant _{final}				Tone			
				Short		Long									
kh	ຂ	k	ກ	kh	ຄ	a	ະ	aa	າ	m [^]	ມ	k [^]	ກ	0	no tone
s	ສ	c	ຈ	ng	ງ	a	າ	ii	ີ	n [^]	ນ	d [^]	ດ	1	◌◌◌
th	ຖ	d	ດ	ng	ທງ	i	ີ	vv	ື	ng [^]	ງ	b [^]	ບ	2	◌◌◌
ph	ຜ	t	ຕ	ngw	ງວ	v	ື	uu	ູ	y [^]	ຍ	l [^]	ລ	3	◌◌◌
f	ຜ	b	ບ	ngw	ທງວ	u	ູ	ee	ເ	w [^]	ວ	s [^]	ສ	4	◌◌◌
h	ຫ	p	ປ	s	ຊ	e	ະ	xx	ຸ			f [^]	ຝ		
khw	ຂວ	j	ຢ	y	ຍ	e	ະ	oo	ໂ			s [^]	ຊ		
sw	ສວ	z	ອ	y	ຫຍ	x	ຸ	@@	◌			r [^]	ຣ		
fr	ຜຣ	kw	ກວ	th	ທ	o	ໂ	qq	ື			th [^]	ທ		
fr	ຜຼ	kr	ກຣ	n	ນ	o	ໂ	ia	ື						
khl	ຂລ	kl	ກລ	n	ຫນ	@	ຸ	ia	ຸ						
khr	ຂຣ	pl	ປລ	n	ຫນ	@	ື	ua	ໂ						
thw	ຖວ	pr	ປຣ	ph	ຟ	Q	ໂ	vva	ໂ						
		dr	ດຣ	f	ຝ	ia	ໂ	ia	ຸ						
		br	ບຣ	m	ມ	ia	ໂ	@@	ອ						
		bl	ບລ	m	ຫມ	Ua	ໂ	ua	ວ						
		tr	ຕຣ	m	ຫນ	Ua	ໂ								
		tw	ຕວ	r	ຣ	Va	ໂ								
		zw	ອວ	l	ລ	a-y [^]	ໂ								
				l	ຫລ	a-y [^]	ໂ								
				l	ູ	a-w [^]	ໂ								
				w	ວ	a-m [^]	ໂ								
				w	ຫວ	a-m [^]	ໂ								
				h	ຮ										
				khw	ຄວ										
				khr	ຄຣ										

				khl	ຄລ										
				thw	ທວ										
				thr	ທຣ										
				phl	ຟລ										
				phr	ຟຣ										
				fl	ຟລ										
				fr	ຟຣ										
				sw	ຊວ										