

Language Specific Peculiarities Document for

Lao as Spoken in Laos

1. Dialects

There are several dialects identified in Lao. The number varies according to sources, ranging from three to six. The dialects of Lao have been principally defined in terms of tones, including number of tones, pitch levels and contours and vocabularies (cf. Suzuki, 2004; Person, 1998).

Although different dialects are said to be mutually intelligible, spoken Lao displays considerable variation between the dialects. There are marked differences in pronunciations and/or tones as well as substantial differences in regular vocabularies, and subtle idiomatic and grammatical differences (Enfield, 2007, p. 19; Suzuki, 2004)¹. Among the dialects, North, Vientiane and South dialects are often described as major dialects; there is little documentation on Northeastern and Central dialects. Vientiane and North dialects, although geographically closely situated, exhibit distinctive characteristics in their tonal structures, pitch contours, and pronunciations (Enfield, 2007). For example, there are two vowel letters (e.g. ວ vs. ດ) that are pronounced the same in Vientiane dialect but pronounced differently in Luangprabang. In other cases, when speakers of different dialects pronounce certain words differently, they may also spell the words differently.

Region	Districts or Cities
Vientiane	Vientiane, Vientiane Municipality
Northern	Xaignabouli, Bokeo, Oudomxai, Phongsali, Luangnamtha, Luangphabang
Northeastern	Xiangkhouang, Houaphan
Central	Khammouan, Bolikhamxai
Southern	Champasak, Savannakhet, Salavan, Attapu, Xekong ²

The dialect spoken in the capital city Vientiane and surrounding central districts is the most widely understood and is generally regarded as the unofficial spoken standard. Standard written Lao is based on Vientiane Lao.

Taking into consideration the populations of dialect speakers and dialectal variations, both spoken and written, we propose to collect the Vientiane dialect for the speech database.

¹Some of the examples include a distinctive dipping (falling-rising) tone in North dialect that is not present in other dialects, and different realization of some consonants and vowels observed in North and South dialects (e.g. /f/ vs. /p^h/, diphthongs vs. long vowels), which in turn leads to orthographic variants (see Enfield, 2007, pp. 19-20 for details).

²Place names are sometimes transliterated differently, e.g. Oudomxai is also spelt as Oudomxay or Oudomsai, or Udomxai, and Xaignabouli as Xaignabouri, Xayabury, Sayaboury, or Sainyabuli etc (Permanent Committee on Geographical Names for British Official Use, 2005).

2. Deviation from native-speaker principle

No special deviation as only native speakers of Lao will be recruited for voice recording in this project.

3. Special handling of spelling

Spelling reform has been taking place in Lao since the 20th century. The Lao script underwent a number of reforms in the 20th century, particularly in 1975. Since 1975 the spelling system has become much simpler and more phonetic with both the letter ຣ “r” and the inherent vowel being removed (See Section 4 for details). This resulted in the Lao alphabet having fewer duplicate letters and being more phonetic than the languages which use similar scripts, such as Thai and Khmer. As of 2000, ຣ has been reintroduced and can be used to represent the ‘r’ sound in recently borrowed words of foreign origin; in words borrowed from Pali and Sanskrit, however, ມ ‘l’ should be used in place of ຣ. This convention serves to make the words from Pali and Sanskrit look more Lao, while words from other sources are more markedly foreign (e.g. ປາຣີ ‘Paris’, ຣັດຊະຍ ‘Russia’).

The fact that there is a fairly well-standardized orthography for Lao, however, does not mean that spelling in the language is highly standardized. People still spell words differently according to their pronunciation of the words, and there are some variations in spelling among the different generations due to phonological changes over time and their education systems.

Regarding foreign and loan words, there is no official transliteration system for Lao. Although historically French-based systems are used, different systems have been used for transliterating foreign loan words, and they vary widely and considerably in the spelling of foreign loan words, particularly of vowels. The spelling reform is said to have also simplified the spelling of foreign loan words by transliterating all foreign loan words according to their sound only, and not their orthographic form or etymological spelling. Nevertheless, as described above, phonetic spelling will still result in orthographic variances as speakers may vary in the extent to which they approximate the original sounds. The word ‘French’, for instance, can be transcribed as ມຣັ່ງ, ມະລັ່ງ, or ມລັ່ງ.

Business names, foreign and loan words from other languages will be spelt in Lao script and standardized to the furthest extent possible using automated methods to detect variant and inconsistent spellings of the same words, and standardized spelling guidelines provided to transcribers. These guidelines will be in accord with the current spelling system, while giving guidance with regard to common patterns of variable or inconsistent spelling which occur within the system. For example, the word ‘French’ will be standardized as ມຣັ່ງ by using the letter ຣ for foreign words (except for Pali-Sanskrit words) and without an epenthetic vowel (inherent vowels are avoided for Lao and Pali-Sanskrit words, but not for recent foreign words).

ວັດຈະນານຸກົມລາວ (Lao Dictionary) by Onmanisone (2008) will be used as a reference.

4. Description of character set used for orthographic transcription

The Lao script will be used for the orthographic transcription of Lao. The Unicode range is U+E80-0EFF.

The script consists of independent consonant and dependent vowel glyphs and ligatures, and various signs. There are three consonant ligatures i.e. ຫນ [n], ຫມ [m], and ຫຼ [l] which are also rendered as digraphs ຫນ, ຫມ, and ຫລ respectively. While a single Unicode code point is available for the first two ligatures, the letter ຫຼ is represented as the sequence of the letters ຫ and ູ. A ligature ອຳ [am] is also available to represent a vowel digraph ອຳ. The consonant ligatures are regarded as standard rendering by the Ministry of Education and therefore the ligatures will be used in the orthographic transcription.

The letter ຣ (U+0EA3) was used to represent [r] in foreign words, but it was officially removed from the script originally by the Lao government (the Ministry of Education) in 1975. Since then, it has been substituted by the letter ລ (U+0EA5); however, the letter was re-introduced to transliterate foreign words (except for Pali-Sanskrit) into Lao in 2000. The letter ງ (U+0EBD) was used in older writing conventions as an alternate form of non-initial ມ (U+0E8D); used after ຫ (e.g. ຫງັ), syllable-finally (e.g. ບາງ), and as a short form for ຄ-ມ when followed by a final consonant (e.g. ວງງຈັບ). Now the letter ງ is only used to represent a diphthong [iə] with a final consonant (as a short form for ຄ-ມ). When a syllable does not end with a consonant, the full form of ຄ-ມ is primarily used.

The letter ຫ, in addition to representing the [h] sound, is used to change the default tone associated with the consonant letters ງ, ນ, ມ, ລ, ວ, ມ to high tone. The character ອ is described as a ‘final nasal’, however, in modern Lao it is used to represent a long [ɔ] in open syllables.

Special characters which might be found in a Lao text are: ໑ ‘ellipsis’ to indicate ellipsis (...); abbreviation (໑໑໑ “et cetera”), ງ ‘word repetition marker’ to indicate repetition of the following word or syllable, and ອ ‘cancellation mark’ which is now obsolete (i.e. used in the old spelling system) to indicate silenced final consonants for foreign words (e.g. ຄູີ້ ບູ້). In modern writing, Western punctuation marks such as ! ‘exclamation mark’, ? ‘question mark’, () ‘parentheses’, ", «» ‘quotation marks’, - ‘hyphen’, and ... ‘ellipsis’ are also commonly found. These special characters, however, do not appear in transcription.

The alphabet is split into vowels and consonants. However, there are no independent vowels; all vowels need to be written in conjunction with a base consonant (i.e. the preceding consonant). Where there is no consonant to support a vowel letter, the character ອ is used as the vowel carrier. Tone marks ອ, ອ, ອ and ອ are placed over a consonant or vowel diacritics which can appear before, after, above, below or around the consonant. Note however, that there are some constraints on the combination of tone marks and vowel diacritics and on the co-occurrence of different vowel diacritics, as well as on the position of vowel diacritics in relation to the base consonant. For example, a combination of vowel diacritics above and below is not allowed in the same syllable; for a consonant cluster, tone marks or vowel diacritics appear over the second consonant; each vowel diacritic has a fixed position with respect to the base consonant and cannot occur at any random position around the base consonant.

The consonants, vowels and tone marks can be found in the Romanization Table in the following section.

5. Description of Romanization scheme

There is no official Romanization system for current Lao orthography. Several Romanization schemes have been used for Lao, in which representation of some sounds differ. These include schemes developed by the United States Board on Geographic Names and the Permanent Committee on Geographical Names for British Official Use (BGN/PCGN), the Royal Thai General System of Transcription (RTGS), and the Romanization scheme by American Library Association and Library of Congress (ALA-LC).

The following is Appen Butler Hill's Romanization scheme. This Romanization system is based on the ALA-LC, which more closely resembles the phonetic representation of Lao, and is fully reversible. However, it should be pointed out that, unlike ALA-LC and the other Romanization schemes mentioned, the ABH system is a simple mapping of code points to ASCII characters or character pairs, and therefore, the following are true for ABH Romanization but not for any of the other systems:

- Vowels which in Lao are written to the left of a consonant are Romanized to the left of the corresponding Romanized consonant.
- Vowels and diphthongs composed of multiple graphemes are transliterated according to their constituent parts rather than as whole units, unless the constituent parts are represented in Unicode by a single code point (as with ອຳ /am/).
- Consonants are not transliterated differently in syllable-final position than in syllable-initial position, even when they are pronounced differently in those two positions
- Tone is not indicated as such, but tone marks are transliterated in whichever position they occur in the Unicode representation of the word.

Thus, where other Romanization schemes prioritize pronunciation, ABH Romanization prioritizes faithfulness to Lao orthography (more precisely, to the Unicode representation of that orthography). These differences are apparent if one contrasts ALA-LC and ABH representations of a string of Lao characters such as ມັດເຂົ້າ ‘grain of rice’. The Lao spelling consists of

- ມ /m/
- ົ /e/ (composed of ົ and ົ, which would represent /e:/ and /a/, respectively, in isolation)
- ົ /t/ in syllable-final position; /d/ in syllable-initial position
- ົ /k^h/
- ົ /ao/ (the constituent parts represent /e:/, /o/, and /a:/, respectively, in isolation)
- ົ tone mark; indicates falling tone (here, low falling tone)

The ALA-LC Romanization of this phrase would be *met khao*; the ABH equivalent would be *ema1deKo4A*.

Note that transcription work is done by Lao speakers working with the Lao script and there will be no Romanization used in transcription; the Romanization scheme is primarily used as a reference for those unfamiliar with the Lao script.

5.1 Lao Romanization Scheme

In the table below, ‘-’ or a dotted circle is used to denote (the position of) a consonant where applicable.

UNICODE	LANGUAGE	ROMAN	DESCRIPTION
U+0E81	ກ	k	LAO LETTER KO
U+0E82	ຂ	K	LAO LETTER KHO SUNG
U+0E84	ຄ	K1	LAO LETTER KHO TAM
U+0E87	ງ	N	LAO LETTER NGO
U+0E88	ຈ	c	LAO LETTER CO
U+0E8A	ຊ	s1	LAO LETTER SO TAM
U+0E8D	ຍ	J	LAO LETTER NYO
U+0E94	ດ	d	LAO LETTER DO
U+0E95	ຕ	t	LAO LETTER TO
U+0E96	ຖ	T	LAO LETTER THO SUNG
U+0E97	ທ	T1	LAO LETTER THO TAM
U+0E99	ນ	n	LAO LETTER NO
U+0E9A	ບ	b	LAO LETTER BO
U+0E9B	ປ	p	LAO LETTER PO
U+0E9C	ຜ	P	LAO LETTER PHO SUNG
U+0E9D	ຝ	f	LAO LETTER FO TAM ³
U+0E9E	ຟ	P1	LAO LETTER PHO TAM
U+0E9F	ຟ	f1	LAO LETTER FO SUNG ⁴
U+0EA1	ມ	m	LAO LETTER MO
U+0EA2	ຍ	y	LAO LETTER YO

^{3,4} The Unicode names are misleading and these should have been “FO SUNG” and “FO TAM” respectively.

UNICODE	LANGUAGE	ROMAN	DESCRIPTION
U+0EA3	ຮ	r	LAO LETTER LO LING
U+0EA5	ລ	l	LAO LETTER LO LOOT
U+0EA7	ວ	w	LAO LETTER WO; When preceded and followed by consonants, realised as [uɔ̌].
U+0EAA	ສ	s	LAO LETTER SO SUNG
U+0EAB	ຫ	h	LAO LETTER HO SUNG
U+0EAD	ອ	O	LAO LETTER O; When preceded and followed by consonants, realised as [ɔː]. Not realised as a phoneme elsewhere.
U+0EAE	ຮ	h1	LAO LETTER HO TAM
U+0EB0	າ	a	LAO VOWEL SIGN A
U+0EB1	ອາ	a1	LAO VOWEL SIGN MAI KAN; Used for closed syllable
U+0EB2	າ	A	LAO VOWEL SIGN AA
U+0EB3	ອາ	A1	LAO VOWEL SIGN AM
U+0EB4	ິ	i	LAO VOWEL SIGN I
U+0EB5	ີ	i1	LAO VOWEL SIGN II
U+0EB6	ຶ	Y	LAO VOWEL SIGN Y
U+0EB7	ຶ	Y1	LAO VOWEL SIGN YY
U+0EB8	ຸ	u	LAO VOWEL SIGN U
U+0EB9	ຸ	u1	LAO VOWEL SIGN UU
U+0EBB	ອ	o	LAO VOWEL SIGN MAI KON; Used for closed syllable
U+0EBC	ົ	11	LAO SEMIVOWEL SIGN LO ⁵
U+0EBD	ຶ	j	LAO SEMIVOWEL SIGN NYO; Used for closed syllable
U+0EC0	ເ	e	LAO VOWEL SIGN E
U+0EC1	ເ	e1	LAO VOWEL SIGN EI

⁵ See section 4.

UNICODE	LANGUAGE	ROMAN	DESCRIPTION
U+0EC2	ໄ	o1	LAO VOWEL SIGN O
U+0EC3	໋	I	LAO VOWEL SIGN AY
U+0EC4	໌	I1	LAO VOWEL SIGN AI
U+0EC8	໊	3	LAO TONE MAI EK
U+0EC9	໋	4	LAO TONE MAI THO
U+0ECA	໌	5	LAO TONE MAI TI
U+0ECB	ໍ	6	LAO TONE MAI CATAWA
U+0ECD	໎	O1	LAO NIGGAHITA; Used in the old convention for final nasal; Used as a vowel [ɔ:] in the current system in open syllables.
U+0EDC	໐	n1	LAO HO NO
U+0EDD	໑	m1	LAO HO MO

5.2 Vowel Diacritics

Lao has abandoned the use of an inherent vowel (i.e. a default vowel /a/ carried by a consonant letter), and all vowels are explicitly indicated by vowel diacritics (e.g. ຕາວ, ຖຸນອມ should now be spelt as ຕະຫຼາວ, ຖະໜອມ, respectively).

The vowel letters found in the above table are the basic vowel letters and often two or more vowels letters are combined to form a vowel sound. The combination of letters to represent vowel sounds can be found in the phone chart in a later section. Additional diphthong- and triphthong-like vowels are formed with a final ຍ, ງ or ວ, which are treated as semi-vowels here.

6. Description of method for word boundary detection

Word boundaries are not marked in Lao text. Texts are written in continuum and white space is reserved for sentence or clause boundaries.

For the purpose of this database, white space will also be used to delimit word boundaries. Words will be determined with reference to their part of speech and dictionaries. Onmanisone (2008) will be used as a reference.

Note that acronyms are transcribed as a word if they conform with Lao syllable structure (e.g. ອາຊຽນ ‘ASEAN’). Otherwise, acronyms are transcribed letter by letter with underscores between the letters (e.g. ວ_ມ_ຕ ຈຸນຮຸ່ນສຳນັກງານລັດຖະມົນຕີ ‘Minister of the Government’).

7. Table containing all phonemes in the stipulated notation

The phonemic transcription of the words in this database uses X-SAMPA symbols, which can be found at <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>. The total number of phonemes is 43. There are 19 consonants, 2 semi-vowels, 22 vowels (18 monophthongs including 9 long vowels, and 4 diphthongs⁶), and 6 tones⁷ (mid, high, rising, high falling, low falling, and low).

In the table, ‘-’ or dotted circle is used to denote a (position of) consonant.

LAO PHONE CHART

TYPICAL LAO CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
CONSONANTS					
ບ	U+0E9A	b	b	b	
ປ	U+0E9B	p	p	p	
ດ	U+0E94	d	d	d	syllable initially; /t/ syllable finally
ຕ	U+0E95	t	t	t	
ຊ	U+0E88	c	c	c	/c/ syllable initially and ranges between alveolar-palatal and palatal; /t/ syllable finally
ກ	U+0E81	k	k	k	
ອ	U+0EAD	O	ʔ	ʔ	Used to support vowels at syllable onset and used as part of some vowels
ຜ	U+0E9C	P	p ^h	p_h	/p ^h / syllable initially; /p/ syllable finally
ຝ	U+0E9E	P1	p ^h	p_h	/p ^h / syllable initially; /p/ syllable finally
ຖ	U+0E96	T	t ^h	t_h	/t ^h / syllable initially; /t/ syllable finally
ຜ	U+0E97	T1	t ^h	t_h	/t ^h / syllable initially; /t/ syllable finally
ຮ	U+0E82	K	k ^h	k_h	

⁶Long and short distinctions are not made here as “long” and “short” diphthongs are in complementary distribution where they always appear with different tones, and are thus subphonemic.

⁷There are 5 and 6 tone analyses of Lao tones. See Section 10.

TYPICAL LAO CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
ຄ	U+0E84	K1	k ^h	k_h	/k ^h / syllable initially; /k/ syllable finally
ຝ	U+0E9D	f	f	f	/f/ syllable initially; /p/ syllable finally
ຜ	U+0E9F	fl	f	f	/f/ syllable initially; /p/ syllable finally
ຮ	U+0EAA	s	s	s	/s/ syllable initially; /t/ syllable finally
ຊ	U+0E8A	s1	s	s	/s/ syllable initially; /t/ syllable finally
ຫ	U+0EAB	h	h	h	This is used (silently) before the following initial consonants to change their default tone value
ຮ	U+0EAE	h1	h	h	/h/ syllable initially; /n/ syllable finally
ມ	U+0EA1	m	m	m	
ຫມ	U+0EDD	m1	m	m	
ນ	U+0E99	n	n	n	
ຫນ	U+0EDC	n1	n	n	
ຍ	U+0E8D	J	ɲ	ɲ	/ɲ/ syllable initially; /j/ syllable finally
ງ	U+0E87	N	ŋ	ŋ	
ຣ	U+0EA3	r	l	l	
ລ	U+0EA5	l	l	l	/l/ syllable initially; /n/ syllable finally
ຢ	U+0EA2	y	j	j	
ວ	U+0EA7	w	w	w	
VOWELS					
ິ	U+0EB4	i	i	i	

TYPICAL LAO CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
ꨀ-ꨁ	U+0EC0 U+0EB0	ea	e	e	Used for open syllable
ꨀꨁ-	U+0EC0 U+0EB1	ea1			Used for closed syllable
ꨀꨂ-ꨁ	U+0EC1 U+0EB0	e1a	ɛ	E	For open syllable
ꨀꨂꨁ-	U+0EC1 U+0EB1	e1a1			For closed syllable
-ꨁ	U+0EB0	a	ɑ	A	Also used as an indication of shortness for syllable final vowels
ꨀ-	U+0EB1	a1			Used when a syllable has a coda; Also as a vowel shortener when forming complex vowels
ꨀꨃ	U+0EB6	Y	ɯ	M	
ꨀꨄ	U+0EB8	u	u	u	
ꨀꨅ-	U+0EC0 U+0EB4	ei	ɤ	7	For closed syllable
ꨀꨅꨁ	U+0EC0 U+0EB4 U+0EB0	eia			For open syllable
ꨀꨆ-ꨁ	U+0EC2 U+0EB0	o1a	o	o	
ꨀ-	U+0EBB	o			When a syllable has a coda
ꨀ-ꨇꨁ	U+0EC0 U+0EB2 U+0EB0	eAa	ɔ	O	For open syllable
ꨀꨇ-	U+0EB1 U+0EAD	a1O			For closed syllable
ꨀꨈ	U+0EB5	i1	i:	i:	
ꨀ-	U+0EC0	e	e:	e:	
ꨀꨉ-	U+0EC1	e1	ɛ:	E:	
-ꨇ	U+0EB2	A	ɑ:	A:	

TYPICAL LAO CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
ອ	U+0EB7	Y1	ɯ:	M:	
ເ	U+0EB9	u1	u:	u:	
ເື	U+0EC0 U+0EB5	ei1	ɤ:	7:	For open syllable
ເື-	U+0EC0 U+0EB7	eY1			For closed syllable
໊	U+0EC2	o1	o:	o:	
໋	U+0ECD	O1	ɔ:	O:	
໋-	U+0EAD	O			When preceded and followed by consonants
DIPHTHONGS					
ເື	U+0EC0 U+0E8D	eJ	iə	i@	For open syllable (long vowel)
ເື	U+0EC0 U+0EBD	ej			rare usage
ເື	U+0EC0 U+0EB1 U+0EBD	ea1j			rare usage
ເື-	U+0EBD	j			When preceded and followed by consonants
ເື	U+0EC0 U+0EB1 U+0E8D	ea1J			For open syllable (short vowel)
ເື-	U+0EB1 U+0EBD	a1j			For closed syllable
ເື	U+0EC0 U+0EB1 U+0EBD U+0EB0	ea1ja			For open syllable; Old spelling convention
ເື	U+0EB1 U+0EBD U+0EB0	a1ja			Old spelling convention

TYPICAL LAO CORRESPONDENCE	UNICODE	ROMAN	IPA	SAMPA	COMMENTS
ົວ	U+0EBB U+0EA7	ow	uɤ	u@	
-ວ-	U+0EA7	w			When preceded and followed by consonants
ົວະ	U+0EBB U+0EA7 U+0EB0	owa			For open syllable
້ວ-	U+0EB1 U+0EA7	alw			For closed syllable
ເືອ	U+0EC0 U+0EB7 U+0EAD	eY1O	uɤ	M@	
ເືອ-	U+0EC0 U+0EB6 U+0EAD	eYO			For closed syllable
ເືອະ	U+0EC0 U+0EB6 U+0EAD U+0EB0	eYOa			For open syllable
ເົາ	U+0EC0 U+0EBB U+0EB2	eoA	ɔo	ʌo	Variant pronunciation for /aw/

OTHER SYMBOLS	
.	syllable break
#	word boundary

Unlike other tonal languages such as Vietnamese, tone marks in Lao do not directly correspond to individual tones. Tones in Lao are rather determined in reference to consonant classes and syllable structure. Details of the tones are described in Section 10.

The following table provides examples of six tones in IPA and SAMPA (in which tones are indicated with numbers).

Vientiane Tone Representation

TONE	IPA	SAMPA	Example	Transcription
mid level	˥	_1	ຄ່າ	k_h A: _1
high rising		_2	ຄາ	k_h A: _2
low rising		_3	ຂາ	k_h A: _3
high falling		_4	ຄ້າ	k_h A: _4
low falling		_5	ຂ້າ	k_h A: _5
low	˩	_6	ກາ	k A: _6

8.1 List of rare phonemes

Information regarding rare phonemes will need to be updated once data is available.

Some phonemes have positional restrictions, such that only a limited set of consonants /p, t, k, ʔ, m, n, ŋ, w, j/ can occur in the coda position.

Since /ɛŋ, ia, ɛt, ɛp/ as syllable final phonemes are observed to occur primarily in loan words and onomatopoeia (Ueda, 1994), we expect these to be rare in the database.

8.2 List of foreign phones

Foreign names and loan words are phonologically nativized into Lao.

For loan words in the process of nativization, phonemes which are not part of the Lao phoneme set are adjusted either by being replaced with native phonemes (e.g. lateral coda [l] changes to the nasal [n] in ໂອເດັນ ‘hotel’ [ʔoːŋ ten˩]), or by being dropped ລຳ ‘mail’ [meːl]). Similarly, consonant clusters do not appear in Lao (apart from Cw⁸) and for loan words they are reduced to singletons (e.g. ຟິມ ‘film’ [fiːm˩], ເປີເຊັນ ‘percent’ [pɛːŋ sen˩]), or are separated by a vowel (e.g. ເດີກະເລ ‘degree’ [dɛːŋ kaːl leːl]). These phonological processes also apply to foreign words (not yet lexicalized into Lao) which contain non-Lao phonemes, e.g. ກູໂກ ‘Google’ [kuː˩ koː˩] (/g/ being substituted by /k/ and syllable-final /l/ being deleted)

The letter **s** theoretically represents the alveolar trill [r] in foreign words. Nevertheless /r/ is not part of the sound system of Lao and is usually realized as /l/. Even when it is pronounced as a trill, it is perceived as marked usage (Enfield, 2007), and therefore it is not included as a foreign phone.

8.3 Tone representation in Lao

Each syllable in Lao (with some exceptions noted below) is pronounced with a specific tone. Standard Lao (Vientiane Lao) is considered to have 5 or 6 tones (Enfield 2007; Hoshino & Marcus, 1989, Suzuki, 2004; Ueda, 1994). 6 tone analyses have recognized the “low” tone,

⁸ Some consider labialized consonants as individual phonemes rather than as consonant clusters (Erickson, 1996).

which in 5 tone analyses is considered to be an allophone of the low rising tone (Ueda 1994; Higbie, n. d.). The “low” tone has been claimed to not be phonemic (Ueda, 1994) and its realization to be unstable (Higbie, n. d.). However, as the low tone is considered a canonical tone (it is one of the tone categories taught in schools), in the database we follow the 6 tone analysis to represent Vientiane Lao.

There are also phonological phenomena in Lao which can be described as tone sandhi, atonic syllable (Ueda, 1994; Suzuki, 2004) or ‘unstressed tone’ (Enfield, 2007) where underlying canonical tones are changed to mid or low tones when syllables are reduced (e.g. shortened) in a certain context. This seems to occur particularly when the syllable in question is the initial syllable of a polysyllabic phonological word, versus when it occurs as a monosyllabic word.

Enfield gives the example of the classifier for non-human animates, “which surfaces as *too3* in a numeral classifier position [...] but as unstressed *toø* when in a modifier classifier position” (Enfield, 2007:139). For example, *kuu3 siø kin3 paa3 toø nii4*, in this context, *toø* appears to be part of a larger phonological word together with the following modifier, hence the reduced tone.

There is also a phenomenon of ‘underspecified tone’ (Enfield, 2007) whereby the tone of sentence-final particles is realized differently depending on pragmatic contexts. Again, Enfield provides a clear example of this phenomenon: “the basis polar question particle *bøø3* is given here as being of Tone 3 (low rising) , but in fact it may be given other kinds of tones – high falling, low falling, high rising, level – with accordingly different nuances.” (2007: 73)

For the purposes of the present database, we will seek to collect the set of particles which may be affected by unstressed tone, in order to provide these with a clear tone assignment, or set of assignments, in the lexicon (e.g. no tone and tone 3 for *to*).

With respect to underspecified tone, only the default or principal tone will be noted in the lexicon (i.e., in the example of *bøø* given above, this will be tone 3).

In terms of tone grapheme to phoneme correspondence, unlike other tonal languages such as Vietnamese, tone mark is only one of the factors in indicating the tone of a syllable, and tone marks do not indicate one specific tone each in Lao. The tone of a syllable in Lao is also determined in reference to consonant classes (“high”, “mid” or “low” class) and syllable structure (“dead” or “live” syllables). There are 4 tone marks in Lao. Not all words have a tone mark coded on a syllable and the tone marks are used in conjunction with open syllables to indicate tones. $\acute{\circ}$ and $\check{\circ}$ are the two most common tone marks and produce a mid tone and a falling tone (either high falling or low falling) respectively. Two further tone marks, i.e. $\overset{\circ}{\circ}$ and $\overset{\cdot}{\circ}$ are rarely used in modern writing, and when they are used, they are used with a mid-class consonant to indicate a high falling tone and a low rising tone, respectively.

The following table summarizes the distribution of the six tones of Vientiane Lao in relation to consonant classes, syllable structure, and tone marks.

	Live Syllable (syllable ends with long vowel or sonorant)			Dead Syllable (syllable ends with short vowel or obstruent)	
	No Tone Mark	◌̎	◌̍	Short Vowel (No Tone Mark)	Long Vowel (No Tone Mark)
Consonant Class					
High	low rising	mid	mid falling	high rising	mid falling
Mid	low	mid	high falling	high rising	mid falling
Low	high rising	mid	high falling	mid	high falling

9. Other language specific items

9.1. Table of Digits

0	ສູນ	su1n
1	ໜຶ່ງ	n1Y3N
2	ສອງ	sON
3	ສາມ	sAm
4	ສີ່	si13
5	ຫ້າ	h4A
6	ຫົກ	hok
7	ເຈັດ	eca1d
8	ແປດ	e1pd
9	ເກົ້າ	eko4A

9.2. Other Numbers

1 (as a digit in numbers > 10, i.e., 21)	ເອັດ	eOa1d
10	ສິບ	sib
20	ຊາວ	sAw
100	ຫ້ິງ໋ອຍ	n1Y3N l4OJ
	ຫ້ິງຮ້ອຍ	n1Y3N h14OJ
	໋ອຍ	l4OJ
	ຮ້ອຍ	h14OJ
1,000	ຫ້ິງຜັນ	n1Y3N P1a1n
	ຜັນ	P1a1n
10,000	ຫ້ິງຫ້ິນ	n1Y3N m1Y13n
	ສິບຜັນ	sib P1a1n
100,000	ຫ້ິງແສນ	n1Y3N e1sn
1 million	ຫ້ິງລ້ານ	n1Y3N l4An
	ລ້ານ	l4An
10 million	ສິບລ້ານ	sib l4An

10. References

- Enfield, N. J. (2007). *A Grammar of Lao*. Mouton de Gruyter: Berlin.
- Erickson, B. (1996). On the origins of labialized consonants in Lao. In K. L. Adams & T. J. Hudak (Ed.), *Papers from the sixth annual meeting of the Southeast Asian Linguistics Society, 1996*, 135-148. Arizona State University Program for Southeast Asian Studies, Monograph Series Press.
- Higbie, J. (n. d.). Thai Language/Lao Language. <http://www.thailao.net/index.html>. Accessed on 14 September 2012.
- Hoshino, T., & Marcus R. (1989). *Lao for Beginners: An Introduction to the Spoken and Written Language of Laos*. Singapore: Tuttle Publishing
- Onmanisone, T. (2008). ວັດຈະນານຸກົມລາວ [Lao Dictionary]. Vientiane: National Library.
- Permanent Committee on Geographical Names for British Official Use (2005). Laos: Toponymy and the Quest for National Identity. Retrieved from <http://www.pcn.org.uk/Laos%20PCGN-November%202005.pdf>
- Person, K. R. (1998). Voice quality and tone in several Lao dialects. In S.L. Chellian and W. J. D. Reuse (Ed.), *Papers from the Fifth Annual Meeting of the Southeast Asian Linguistics Society*, 261-276. Arizona State University, Program for Southeast Asian Studies.
- Roffe, G. E. (1946). The Phonemic Structure of Lao. *Journal of the American Oriental Society*, 66 (4), 289-295.
- Suzuki, R. (2004). Laosugo [Lao]. In Y. Kawaguchi, K. Moriguchi & Y. Saito (eds.) *Gengo Johogaku Kenkyu Hokoku 4*, 255-276. Tokyo: Tokyo University of Foreign Studies.
- Ueda, R. (1994). Gendai Laosugo Vientiane Hogen no Onsei Taikai [The Phonology of the Vientiane Dialect of Lao]. *Gengo Kenkyu*, 106, 95-115.