

Mon-Khmer Studies

Volume 42

Author: Michelle MILLER
Title: A Description of Kmhmu' Lao Script-Based
Orthography
Pages: 12-25
Date Received: 26/6/2012
Revised Text Accepted: 14/5/2013
Copyright for this paper vested in the author
Released under Creative Commons Attribution License

Volume 42 Editors:
Paul Sidwell
Brian Migliazza

ISSN: 0147-5207
Website: <http://mksjournal.org>

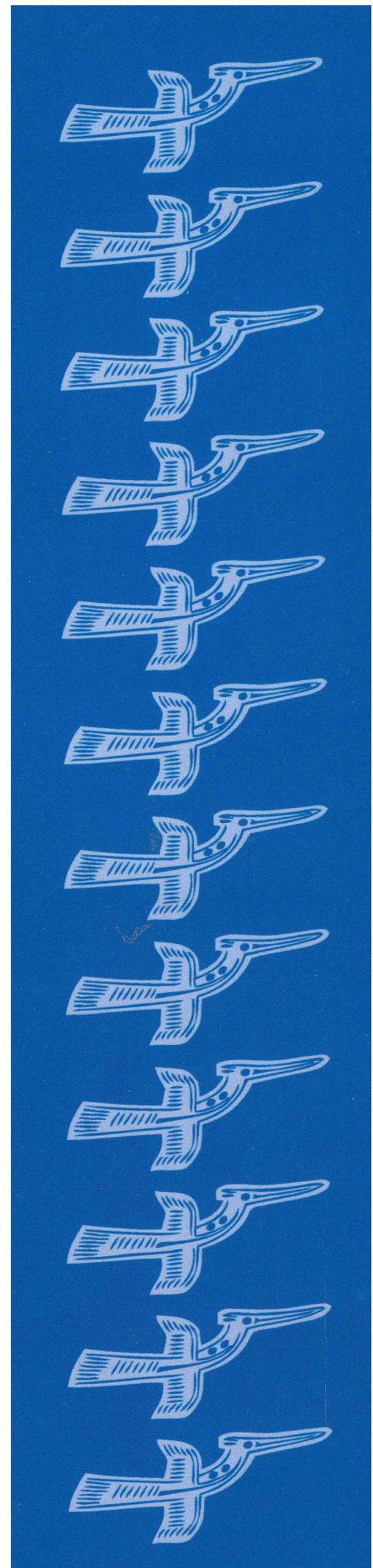
Published by:



Mahidol University (Thailand)



SIL International (USA)



A Description of Kmhmu' Lao Script-Based Orthography

Michelle MILLER

Linguistics Institute, Payap University

Abstract

Kmhmu' is a language of the Mon-Khmer language family. Extensive linguistic research and analysis of the varieties of Kmhmu' spoken in Southeast Asia has led to the grouping of Kmhmu' into three dialect categories, generally referred to as Northern, Western and Southern (Svantesson 1989). The orthography described in this paper was developed for the Southern dialect and utilizes a Lao-based script. Suksavang and Preisig (Suksavang et al 1994) were instrumental in refining this orthography. This description of the Southern Kmhmu' orthography explains how the Lao script is used and/or adapted to represent the Kmhmu' phonemes, presents orthographic conventions for writing words of various structural types and summarizes teaching/learning experiences observed in mother-tongue Kmhmu' speakers.

Keywords: Austroasiatic, orthography, sesquisyllables

ISO 639-3 language codes: kjg

1. Introduction

Kmhmu' is a language of the Mon-Khmer language family. There has been extensive linguistic research and analysis of the varieties of Kmhmu' spoken in Southeast Asia, which has led to grouping of Kmhmu' dialects into three major dialect categories, generally referred to as Northern, Western and Southern (Svantesson 1989). The orthography described in this paper was developed for the Southern dialect, which is spoken in Phongsali, eastern Udomsay, Luang Prabang, Hua Phan, Xieng Khouang, Sayabuli, Vientiane and Bolikhamsay provinces in Lao PDR (Suksavang et al. 1994, Svantesson 1989), in Diên Biên Phủ, Sơn La and Nghệ An provinces in North Vietnam and some villages of Sipsongpanna in China (Suwilai 2002)". Preisig proposes that this orthography can also be used by speakers of other dialects with some explanation on how alternate pronunciation could be associated with the various graphemes and orthographic conventions of this orthography (Suksavang et al 1994), an assertion that requires further testing.

Systematic linguistic analysis and an effort to develop a written form of Kmhmu' began in the 1950s with the work of William Smalley. According to an account by Preisig¹, he was the first to use the Lao script to write Kmhmu', though this script did not come into wide-spread use at that time. Perhaps the earliest attempt at writing Kmhmu' was made by Mrs. C.H. Crooks who used the northern Thai script to print a translation of the Gospel of Mark (Svantesson 1983:1, Smalley 1963:75). Another of the first pieces of literature produced in Kmhmu' utilized a Roman script, a Kmhmu'-French dictionary developed in 1964 by a French teacher for use in the classroom (Suksavang et al 1994). A French priest, Father Bonometti, also used a Roman script-based orthography in his translation of Scripture portions printed during the 1960s.

Preisig and Suksavang were instrumental in refining the Kmhmu' Lao script-based orthography. The first description of this orthography was prepared by Preisig in 1990 as an unpublished manuscript entitled *The Kmhmu' Orthography Dialect of Xieng Khouang, Luang Prabang and Sam Neua*, and the orthography has not been changed since that paper was written (Preisig 1990). The Kmhmu' Lao script orthography was first used in an official publication in 1994, the *Kmhmu'-Lao-French-English Dictionary* (Suksavang et al 1994). This orthography statement is an effort to formalize the description prepared by Preisig (1990). Section 2 describes how the Lao script is used to represent the Kmhmu' phonemes. Special attention is given to explaining the rationale behind the various solutions employed for representing Kmhmu' phonemes that are not found in Lao. Section 3 contains a description of Kmhmu' word structure, an

¹ personal communication 2010

explanation of the orthographic conventions for writing words and a description of conventions that have been adopted for punctuation. Section 4 contains a description of how the Kmhmu' orthography is being used followed by a review of teaching/learning experiences with this orthography. Conclusions and outlook are in Section 5. The language data used in this description comes from Osborne (2013) and Suksavang et al (1994).

2. Kmhmu' orthography

The Southern dialect of Kmhmu' has 36 consonant phonemes. All of them occur in initial position and 16 of them form codas. There are 17 initial consonant clusters. The Kmhmu' orthography uses the Lao graphemes to represent the same sound-symbol correspondence in Kmhmu' and Lao where possible. Because the sound inventories of Lao, a Tai-Kadai language, and Mon-Khmer Kmhmu' differ considerably, it was necessary to adapt the use of the Lao graphemes and writing conventions to accommodate a Kmhmu' orthography. Specific adaptations are described in the sections below, grouped according to initials, clusters, codas, and vowels.

2.1 Initial consonants

Table 1 below presents a summary of all 36 initial consonant phonemes. Various solutions were employed to represent these phonemes, and explanations of these solutions are presented in this section.

Table 1: The 36 Initial Kmhmu' consonant phonemes and graphemes; consonants not found in Lao are in shaded cells.

Manner of articulation	Point of articulation									
	bilabial		alveolar		palato-alveolar		velar		glottal	
voiceless unasp. stops	p	ປ	t	ຕ	ʈ	ຈ	k	ກ	ʔ	ອ
voiceless asp. stops	p ^h	ພ	t ^h	ທ	ʈ ^h	ຊ	k ^h	ຄ		--
voiced stops	b	ບ	d	ດ	ɖ	ຜ	g	ຖ		--
pre-glottalized nasals	ʔm	ມ່	ʔn	ນ່	--	--	ʔŋ	ງ່		--
voiceless nasals	m̥	ຫມ	n̥	ຫນ	ɲ̥	ຫຍ	ŋ̥	ຫງ		--
voiced nasals	m	ມ	n	ນ	ɲ	ຍ	ŋ	ງ		--
pre-glottalized approx.	ʔw	ວ່			ʝ	ຢ່				
voiceless approximants	w̥	ຫວ		---	ʝ̥	ຫຢ	--	--		--
voiced approximants	w	ວ			ʝ	ຢ				
fricatives	--		s	ສ	--	--	--	--	h	ຫ
voiceless lateral	--		l̥	ຫລ	--	--	--	--		--
voiced lateral	--		l	ລ						
voiceless trill	--		r̥	ຫຣ	--	--	--	--		--
voiced trill	--		r	ຣ						

There are 17 initial Kmhmu' consonant phonemes that are not found in Lao. The solutions used to represent these non-Lao sounds can be grouped into three categories, namely 1) creation of a new grapheme, 2) assignment of a new value to an existing grapheme or diacritic and 3) the use of special characters. The sections below describe how these solutions were applied.

2.1.1 Creation of a new grapheme

Voicing is a Kmhmu'-specific contrastive feature not found in Lao. Lao does not have a voiced velar stop, /g/, or a voiced palato-alveolar stop, /gʲ/. In the Lao grapheme inventory there are not any available graphemes for representing additional velar phonemes, and thus a completely new grapheme was created to represent the voiced velar stop in Kmhmu'. The new grapheme was

created to resemble the Lao grapheme for the consonant-initial voiceless velar stop /k/ which is {ກ}. The grapheme created for /ŋ/ in Kmhmu' is {ກ̣}. The representation of the voiced palato-alveolar stop was derived via a different solution (see section 2.1.1.2).

Example word in Kmhmu': /ŋul/ ກ̣ ລ 'to be fat'

2.1.2 Assigning new value to an existing grapheme or diacritic

Existing Lao graphemes and diacritics with phonemic values foreign to Kmhmu' were used to symbolize preglottalisation, another Kmhmu'-specific contrastive feature, as well as several other non-Lao phonemes, namely the voiced and voiceless palato-alveolar stops, and the voiceless sonorants.

2.1.2.1 Voiceless palato-alveolar stop /ç/

The Lao grapheme {ຊ} is used to represent the voiceless aspirated palato-alveolar stop. In the Lao orthography this grapheme represents the low-class consonant phoneme /s/. Kmhmu' also has the phoneme /s/, but there is another Lao grapheme available to represent /s/ in Kmhmu', the high-class grapheme {ສ}.

Example word in Kmhmu': /çççç/ ສຊຊຊ 'to put away in a corner'

2.1.2.2 Voiced palato-alveolar stop /j/

The Lao grapheme {ຢ} is used to represent the voiced palato-alveolar stop. In the Lao orthography this grapheme represents the diphthong /ia/. The grapheme {ຢ} was likely chosen to represent initial consonant /j/ because the diphthong /ia/ approximates a palatal place of articulation. Admittedly, its use as a consonant in the Kmhmu' orthography may be somewhat confusing for new readers of this orthography, though its use as an initial consonant grapheme is not ambiguous in any context.

It should be noted that using the grapheme {ຢ} as a consonant will necessitate special considerations for font design. Because the 'tail' of this grapheme hangs below the baseline it will collide with the vowel grapheme the /u/ which is written below the consonant grapheme unless adjustments are made in the font.

Example word in Kmhmu': /jççç/ ຢຊຊຊ 'great grandchildren'

2.1.2.3 Pre-glottalized initial sonorants /m̥/ /n̥/ /ŋ̥/ /j̥/ /w̥/

The Kmhmu' orthography uses the Lao tone diacritic ◌̣ to indicate pre-glottalization of three nasals and two approximants shown below. The southern Kmhmu' dialect does not have phonemic tone, so this diacritic is not needed for the purpose of representing tone in Kmhmu'.

nasals: /m̥/ - {ມ̣}, /n̥/ - {ນ̣}, /ŋ̥/ - {ງ̣}

approximants: /j̥/ - {ຢ̣}, /w̥/ - {ວ̣}

Example words in Kmhmu':

/m̥/ → /m̥m̥ŋ̥/ ມ̣ ມ̣ ງ̣ 'to be hidden'

/n̥/ → /n̥n̥n̥/ ນ̣ ນ̣ ນ̣ 'month'

/ŋ̥/ → /sịm̥ŋ̥k̥/ ສີ̣ ມ̣ ດຽ ກ 'black toucan'

/j̥/ → /j̥j̥j̥/ ຢ̣ ຢ̣ ຢ̣ 'round basket for carrying poultry'

/w̥/ → /w̥w̥t̥w̥t̥/ ວ̣ ວ̣ ຕ ວ̣ ຕ 'loud booming sound'

2.1.2.4 Voiceless initial sonorants /m□/ /n□/ /□□/ /ŋ□/ /w□/ /j□/ /l□/ /r□/

Kmhmu' has eight voiceless sonorant consonants. Lao does not have voiceless sonorant consonants therefore some innovation was required to represent this feature. In the Kmhmu' orthography the Lao grapheme for /h/, {ຫ}, is written preceding the grapheme of the voiced form of the consonant to form a digraph as shown below. In the Lao orthography, the same grapheme, {ຫ}, is used as a diacritic with six of the low class consonants to indicate tone (Becker 2003), therefore new readers of Kmhmu' who can read Lao need to learn the new meaning of the diacritic {ຫ}.

nasals: /m□/ - {ຫມ}, /n□/ - {ຫນ}, /□□/ - {ຫຍ}, /ŋ□/ - {ຫງ}

approximants: /w□/ - {ຫວ}, /j□/ - {ຫຍ}

trill: /r□/ - {ຫຣ} lateral: /l□/ - {ຫລ}

Example words in Kmhmu':

/m□/ → /m□□ŋ/ ຫມັ່ງ 'old'

/n□/ → /n□□m/ ຫນົມ 'young'

/□□/ → /□□□□m/ ຫຍາມ 'to be used'

/ŋ□/ → /si□m ŋ□□□k/ ຍສົມ ຕຫງກ 'black toucan'

/w□/ → /w□□t/ ຫວັດ 'to throw'

/j□/ → /j□□□ŋ/ ຫຍາງ 'female' (animal)

/l□/ → /l□□□ŋ/ ຫລາງ 'classifier for traps'

/r□/ → /r□□□ŋ/ ຫຣາງ 'tooth'

2.1.3 Use of special characters

In Lao there is a no voiced trill /r/, but because there are many borrowed words particularly from Thai that have a syllable-initial voiced trill, a special character that is no longer part of the official Lao alphabet (Becker 2003) is used to represent the voiced trill in borrowed words, namely {ຣ}. The Kmhmu' phoneme inventory includes both a syllable-initial and syllable-final voiced trill, and the special character {ຣ} is used for this sound in Kmhmu' also.

Example word in Kmhmu': /r□□ŋ/ ຣາງ 'flower'

2.2 Consonant Clusters

Lao does not have consonant clusters, but Kmhmu' has them in syllable-initial position. Consonant clusters are found only in major syllables; minor syllables do not have consonant clusters (see sections 3.2 and 3.3 for discussion of major and minor syllables). Consonant clusters in Kmhmu' are written with the two consonant graphemes of the cluster in a sequence. There is a restricted set of consonant phonemes found in both the first (C₁) and second (C₂) elements of the cluster forming a total of 16 clusters.

Table 2: Consonant cluster (C₁ and C₂) phonemes and graphemes

C ₁		C ₂					
		w	ວ	l	ລ	r	ຣ
p	ປ			pl	ປລ	pr	ປຣ
p ^h	ພ					p ^h r	ພຣ
b	ບ			bl	ບລ		
t ^h	ທ					t ^h r	ທຣ
d	ດ					dr	ດຣ
t ^h	ຊ					t ^h r	ຊຣ
ɖ	ຜ					ɖr	ຜຣ
k	ກ	kw	ກວ	kl	ກລ		
k ^h	ຄ	k ^h w	ຄວ	k ^h l	ຄລ	k ^h r	ຄຣ
g	ກ	gw	ກວ			gr	ກຣ
s	ສ					sr	ສຣ

2.3 Codas

Of the 16 Kmhmu' final consonant phonemes six are not found in Lao: /ɰ/, /h/, /r/, /l/, /ɰ/ and /jɰ/. Table 3 below presents a summary of all 16 Kmhmu' final consonant phonemes. The solutions used to represent these non-Lao sounds can be grouped into three categories, namely 1) use of the corresponding Lao consonant-initial grapheme, 2) creation of a new grapheme and 3) assignment of a new value to an existing grapheme. Detailed explanations of these solutions are presented below.

Table 3: The 16 Kmhmu' final consonant phonemes and graphemes

Manner of articulation	Point of articulation									
	bilabial		alveolar		palato-alveolar		velar		glottal	
voiceless unasp. stops	p	ປ	t	ດ	t̪	ຈ	k	ກ	ʔ	ຮ
voiced nasals	m	ມ	n	ນ	ɲ	ຮ	ŋ	ງ	--	--
voiced approx.	w	ວ	--	--	j	ຍ	--	--	--	--
fricatives	--	--	--	--	ʃ	ຢຫ	--	--	h	ຫ
voiced lateral	--	--	l	ລ	--	--	--	--	--	--
voiced trill	--	--	r	ຣ	--	--	--	--	--	--

2.3.1 Use of the corresponding Lao consonant-initial grapheme

Smalley (1963) espouses adherence to the “phonemic principle”, namely that “every distinctive sound is represented by one symbol and only one in the writing system” (Smalley 1963:38). This principle was followed in determining graphemes for the final consonants in Kmhmu' that are not found in Lao where possible. The following sections describe how this was applied for four consonant-final Kmhmu' phonemes.

2.3.1.1 Palato-alveolar stop /ɰ/

The grapheme chosen to represent /ɰ/ is {ຈ}.

Example word in Kmhmu': /hoɰɰ/ ໂຫຈ ‘to be finished’

2.3.1.2 Voiceless glottal fricative /h/

The grapheme chosen to represent /h/ is {ຫ}.

Example word in Kmhmu': /tuh/ ຕຸຫ ‘old rice field’

2.3.1.3 Voiced lateral approximate /l/

The grapheme chosen to represent /l/ is {ᦟ}.

Example word in Kmhmu': /kᦑᦑᦟ/ កកᦟ 'before'

2.3.1.4 Voiced trill /r/

The grapheme chosen to represent /r/ is {ᦞ}.

Example word in Kmhmu': /kᦑᦑᦞ/ កកᦞ 'to grill'

2.3.2 Creation of a new grapheme

The initial consonant phoneme /ᦑ/ in Lao is represented by the grapheme {ᦑ̃}. This grapheme could not be used for the final consonant phoneme /ᦑ/, however, because in the Lao orthography the grapheme {ᦑ̃} in syllable-final position represents the phoneme /j/, not /ᦑ/ as it does in the initial position. It was therefore necessary to find another character to represent /ᦑ/ in syllable-final position. The solution was to design a new character, the grapheme {ᦑ̂}.

Example word in Kmhmu': /tᦑᦑᦑ̂/ តកᦑ̂ 'to weave'

2.3.3 Assignment of a new value to an existing grapheme

The voiceless palatal approximant /jᦑ/ in syllable-final position is represented by the digraph {ᦑᦞ}. The voiceless palatal approximant in initial position is represented as {ᦞᦑ} and therefore some discussion is warranted to explain why different graphemes are used for this phoneme in initial and final position.

The Lao orthographic convention marks word-final /j/ as {ᦑ}, and it is thus logical to use the {ᦑ} in this grapheme. The digraph {ᦑᦞ} adds devoicing to the word-final palatal semi-vowel. However, there is a phonetic difference between syllable-initial and -final /jᦑ/. There is a voiced transition from a syllable-initial /jᦑ/ into the following vowel whereas the syllable-final /jᦑ/ has a voiced transition caused by the preceding vowel before it becomes fully voiceless. The grapheme {ᦞᦑ} for syllable-initial /jᦑ/ and {ᦑᦞ} for final /jᦑ/ thus reflect the pronunciation variants for this one phoneme in different positions within a word.

Example word in Kmhmu': /rᦑᦑᦑᦞ/ រកᦑᦞ 'to cease (of rain)'.

2.4 Kmhmu' vowels

Kmhmu' has 10 vowel qualities which all occur in both short and long forms (Osborne 2013). There are also three diphthongs (Osborne 2013). Vowel quality contrast is neutralized in open syllables where all vowels and diphthongs are long. Neutralization also occurs in syllables ending on /h/, /ᦑ/, /jᦑ/, and /j/ where all vowels are short (Osborne 2013). All of the Kmhmu' vowels except one and all three diphthongs are found in Lao. The Kmhmu' orthography can therefore use the Lao graphemes in the same sound-symbol correspondence, with the addition of two graphemes {ᦑᦑ̃} and {ᦑᦑ̂} representing the long and the short forms respectively of the near-open central vowel /ᦑ/. Table 4 presents a complete listing of all vowels and corresponding Kmhmu' graphemes. As in Lao, the representation of some vowels is dependent on syllable structure, with some vowels being written differently in open and closed syllables.

2.5 Special symbols representing vowel-consonant combinations

The Kmhmu orthography has incorporated the use of the Lao special characters representing CV combinations /ᦑm/, /am/, and /ᦑw/, realized as {ᦑᦞᦞ}, {ᦑᦞᦞ}, and {ᦑᦞᦞ} (for convenience these are also included in Table 4).

Table 4: Kmhmu' vowel phonemes and graphemes; shaded graphemes denote the two vowels not found in Lao

	Front Unrounded					
	IPA	Short		Long		
		cv	cvc	IPA	cv	cvc
Close	i	ິ	ິ່	i:	ິ່	ິ່
Mid	e	ເຂ	ເ້	e:	ເ	ເ່
Near- open	ɛ	ແະ	ແ້	ɛ:	ແ	ແ່
Open		--			--	
Diphthongs	ia	ເ້ຍ	້ຽ່	ia:	ເຍ	ຽ່
Special symbols	am	ໍ່າ		--		
	Central Unrounded					
	IPA	Short		Long		
		cv	cvc	IPA	cv	cvc
Close	ɨ	ື	ື່	ɨ:	ື	ື່
Mid	ə	ເິ	ເິ່	ə:	ເິ	ເິ່
Near- open	ɘ	ເິ່	ເິ່	ɘ:	ເິ່	ເິ່
Open	ɑ	ະ	້	ɑ:	າ	າ່
Diphthongs	ia	ເືອ	ເືອ	ia:	ເືອ	ເືອ
Special symbols	aj	ໄ່		--		
	Back Unrounded					
	IPA	Short		Long		
		cv	cvc	IPA	cv	cvc
Close	u	ຸ	ຸ່	u:	ຸ	ຸ່
Mid	o	ໄະ	ື່	o:	ໄ	ໄ່
Near- open	ɔ	ເາະ	້ອ່	ɔ:	ໍ່	ອ່
Open		--			--	
Diphthongs	ua	ືວະ	ືວ່	ua:	ືວ	ວ່
Special symbols	aw	ເົາ		--		

3 Orthography conventions

This section describes the conventions for writing Kmhmu' words, borrowed words and punctuation.

3.1 Kmhmu' word structure

A brief description of Kmhmu' word structure will be presented in this section because the rationale for various orthographic conventions for writing Kmhmu' words are related to features of the word structure. Kmhmu' words in terms of their syllabic composition are usually mono- or sesquisyllabic (Matisoff 1973). There are a small number of di- and tri-syllabic words.

Monosyllabic words consist of one syllable, referred to as a major syllable in multisyllable words, which may be open or closed, and can contain the full range of vowel qualities (Osborne 2012).

Sesquisyllabic words consist of a major syllable preceded by a minor syllable, or as described by Gafos as “a heavy syllable which is optionally preceded by a vowelless syllable made

up of one or two consonants” (1999:119). The use of the term ‘minor syllable’ in this paper is in accordance with the criteria described by Herr (2011) as 1) a syllable that cannot exist independently of the major syllable, 2) has vowel neutralization, 3) a reduced consonant inventory and 4) lacks prosodic features such as vowel length, tone and stress. According to Herr (2011) the term ‘minor syllable’ should not be used for the initial syllables in multi-syllable words in which there is merely reduced or limited vowel contrast (either in vowel quality or length). She prefers the term “major syllable with reduced vowel contrast” or “reduced syllable” (Herr 2011:25) for such syllables.

It could be argued that in some dialects of Kmhmu’, namely one or more of the ‘Northern’ tonal dialects (Svantesson 1983), the term ‘minor syllable’ is not completely appropriate according to the definition stated above because, according to Svantesson (1983), there are both tonal and non-tonal minor syllables. Svantesson and Karlsson (2004 cf Herr p. 25) site examples in ‘Northern’ Kmhmu’ dialects in which the tone of the minor syllable is contrastive, although minimally so in that only about ten minimal pairs have been identified (2004:2 cf Herr p. 26). The Southern dialect of Kmhmu’ for which this orthography has been designed, has no contrastive tone (Osborne 2012) and thus the term ‘minor syllable’ is appropriate.

Di- and tri-syllabic words are rare. Disyllabic words have two major syllables. There are a few trisyllabic words that contain two minor syllables preceding a major syllable, or a minor syllable followed by two major syllables in which the second major syllable is reduplicated (Suksavang et al. 1994).

In the remainder of this paper when transcribing syllable structure, a lowercase ‘v’ will be used to denote the vowel of the minor syllable, given the greatly reduced nature of this vowel in both vowel quality and length, and the uppercase ‘V’ will be reserved for transcribing vowels with full vowel characteristics. This distinction in transcription is being made in order to eliminate ambiguity in transcriptions of sesquisyllabic or multisyllabic words.

3.2 Monosyllabic words (or major syllables)

Spelling monosyllabic words (or the major syllable in sesquisyllabic or multisyllable words) is quite straightforward in that each phoneme in the word (or syllable) has explicit graphemic representation, i.e., every sound has a corresponding character. Table 5 presents the Kmhmu’ syllable patterns for monosyllabic words with examples.

Table 5: Single-syllable words or major syllables

Syllable structure	IPA	Kmhmu’	Meaning
CV	tu:	ຕູ	'to falsely accuse'
CCV	p ^h ria	ເຟຣີ້ອ	'fire'
CVC	kuŋ	ກຸງ	'village'
CCVC	plɔːŋ	ປລອງ	'calf' (of leg)

There is potential for ambiguity in pronouncing single syllable words that are of the CCV pattern in which the vowel or part of the vowel is written before the consonant cluster. There are two situations in which ambiguity may occur:

(1) The consonant cluster may be interpreted as two separate phonemes in words in which the part of the vowel symbol is written before the consonant cluster and part is written after (or above), such as is the case with the vowels {c□z}, {cc□z} or {ŋ□z}. For example, consider the word /ple’/ ‘fruit’, written in Kmhmu’ as {cUລະ} (Suksavang et al). This word could be pronounced as the two-syllable word /pe l□’/ {cU - ລະ}. However, the two-syllable pronunciation has no meaning. The reader must therefore rely on the meaning of the word in its context in order to determine the correct pronunciation in such situations, albeit rare.

(2) The consonant cluster may be interpreted as two separate phonemes (or conversely two separate phonemes may be interpreted as a consonant cluster) in words with long vowels in which the vowel grapheme is written before the consonant cluster such as with the vowels {ŋ□}, {c□}, or {cc□}. For example, the word /kool/, written in Kmhmu' as {ໂກລ}, could be read as /ko:l/ or as /klo:/. Again, the reader must rely on the meaning of the word in context in order to determine the correct pronunciation. In both of these situations it has been observed that new readers may hesitate when encountering such structures, but with practice they gain automaticity in word recognition in context.

3.3 Sesquisyllabic words

Sesquisyllabic words consist of one major syllable preceded by one minor syllable. As described previously, writing the major syllable of a sesquisyllabic word is quite straightforward because each phoneme is represented by a grapheme. The method for writing the minor syllable, particularly the vowel, is determined by whether the minor syllable is open (Cv) or closed (CvC), and is described in sections 3.3.1 and 3.3.2 below. There is a restricted set of consonants that is found in the initial and final consonant position of the minor syllable, and the inventory of the initial consonant set is also dependent on whether the minor syllable is open or closed. Initial and final consonant inventories of minor syllables are presented in Table 6.

Table 6: Minor syllable consonant phonemes and graphemes

Initial consonants: Cv	Initial consonants: CvC	Codas: CvC
p ປ	p ປ	p ປ
t ຕ	t ຕ	t ຕ
tʰ ຕ	tʰ ຕ	k ກ
tʃ ຕ	tʃ ຕ	m ມ
k ກ	k ກ	n ນ
g ກ	--	ŋ ງ
s ສ	s ສ	ŋ ງ
h ຫ	h ຫ	l ລ
l ລ	l ລ	r ຣ
r ຣ	r ຣ	--

3.3.1 Open minor syllables (Cv)

In minor syllables of the Cv structure *a vowel grapheme is not written*, only the initial consonant grapheme is written. The major syllable of the word is written connected to the minor syllable to form a single unit.

Example word: Cv.CVC → /□□□□.l□□η/ → ຈລອງ 'boat'

The vowel is not written in the minor syllable of the Cv structure because the vowel quality is reduced and difficult to distinguish, the length of the vowel is always extra short, and writing a vowel grapheme may inadvertently cause the reader to read the syllable as a full-length, accented syllable which it is not. In fact for Kmhmu', Shaw (1993 cf. Gafos 1999:120) proposes that minor syllables with a single consonant be assigned "no mora" in his moraic model of syllables.

The vowel is not written in the minor syllable of the Cv structure because the vowel quality is reduced and difficult to distinguish, the length of the vowel is always extra short, and writing a vowel grapheme may inadvertently cause the reader to read the syllable as a full-length, accented syllable which it is not. In fact for Kmhmu', Shaw (1993 cf. Gafos 1999:120) proposes that minor syllables with a single consonant be assigned "no mora" in his moraic model of syllables. There are variant pronunciations of the non-contrastive vowels in the pre-syllable based on regional accents, though generally the vowel quality is a central unrounded vowel area [□ - ə - a] (Svantesson and Karlsson 2004:1; Suksavang et al 1994), and if speakers tried to spell the word as it is pronounced a confusing variety of spellings would likely result.

Because the vowel is not written in the minor syllables of the Cv structure, it is necessary to address the question of potential ambiguities in the reading of multisyllable words of both the Cv.CV(C) and Cv.CCV(C) word structures.

First, with words of the Cv.CV(C) structure, how will the reader know that the two consecutive initial consonants are to be read with a vowel between them and not as a consonant cluster or as a consonant phoneme whose grapheme is a digraph (note the only digraphs are the voiceless sonorant consonants), i.e., how will the reader know the word should be read as Cv.CV(C) multi-syllable word and not as a single-syllable word of the CCV(C) pattern or CV(C) (where C is a digraph)?

Because there is a restricted set of consonants that can be in the initial consonant position and there is a limited set of consonant clusters and digraphs, we can determine potential ambiguities by looking at the possible consonant combinations that may occur in a Cv.CV(C) word. The combinations that would be ambiguous in terms of how they should be read are those that mimic a consonant cluster or a digraph. Table 7 below presents a summary of possible consonants that may appear in 1) the consonant position of the Cv minor syllable, 2) the C2 position of a consonant cluster 3) consonant clusters and 4) digraphs.

Table 7: Determination of potentially ambiguous consonant combinations in Cv.CV(C) word constructions; shaded consonant clusters and digraphs are potentially ambiguous.

Minor syllable Cv	Major syllable CCV(C)	Consonant clusters CC	digraphs ຫ + C	Summary of potentially ambiguous consonant combinations
p ປ	w ວ	kw ກວ	ຸ ຫມ	pl ປລ
t ຕ		k ^h w ຄວ	ຸ ຫມ	
t ^h ຊ		gw ງວ	ຸ ຫຍ	
t ຈ	l ລ	pl ປລ	ຸ ຫງ	
k ກ		bl ບລ	ຸ ຫມ	kw ກວ, kl ກລ
g ງ		kl ກລ	w ຫວ	gw ງວ, gr ງລ
s ສ		k ^h l ຄລ	j ຫຍ	sr ສລ
h ຫ	r ຣ	pr ປລ	r ຫລ	hw ຫວ, hl ຫລ, hl ຫລ
l ລ		p ^h r ພລ	l ຫລ	
r ຣ		t ^h r ທລ		
--		dr ດລ		
		t ^h r ຊລ		
		qr ວລ		
		k ^h r ຄລ		
		gr ງລ		
		sr ສລ		
		pr ປລ		

Of the potentially ambiguous consonant combinations summarized in Table 7, there is no data to support the existence of words with these consonant constructions. We can conclude, then, that there are likely no ambiguities for the reading of Cv.CV(C) words. When reading Cv.CV(C) the appearance of a ‘cluster’ may initially cause some confusion for new readers of Kmhmu’ because in the Lao orthography there are not any unwritten vowels in multisyllable words. However, with practice, Kmhmu’ speakers are able to read words of the Cv.CV(C) structure fluently.

Regarding the potential for ambiguities when reading Cv.CCV(C) words, there is no ambiguity as to how to read them because there is only one way to decode a word with three consecutive initial consonants, i.e. the word can only be read by inserting a vowel sound after the first consonant.

Example word of the Cv.CCV(C) pattern: /l□□□druŋ/ □□□□□ ‘to fall down’ (hands and legs spread).

3.3.2 Closed minor syllables (CvC)

In contrast with the Cv minor syllables, a vowel is written in the CvC minor syllable. The same vowel grapheme is written in all closed minor syllables (CvC). The vowel chosen for this syllable structure is the close central vowel /□/, realized as the grapheme {□}. There is variation in speaker pronunciation of this vowel sound in the minor syllable, and there is often vowel harmony between the vowel of the minor syllable and that of the main syllable, but the close central vowel was chosen because that is the vowel most frequently used by speakers (Suksavang et al 1994). The use of the same vowel in all CvC minor syllables helps to maintain consistency in spelling. The vowel cannot be left unwritten as in the Cv minor syllable because doing so would allow for an initial consonant sequence of four graphemes in a word, i.e. a two-syllable word could begin with CCCC, making de-coding too difficult. Because the vowel is marked in multi-syllable words with a CvC minor syllable, there is no ambiguity in reading words having CvC.CCV(C) or CvC.CV(C) structures.

Example word with CvC.CCV(C) structure: /□□n□dri□h/ ຈີ່ນດຣັຍັຍຫ ‘comb’

Table 8: Summary of Kmhmu’ word structure (adapted from Osborne 2013)

Frequency of occurrence	Examples			
	Syllable structure	IPA	Kmhmu’	Meaning
common structures	CV	tu:	ຕູ	to falsely accuse
	CVC	mar	ມ້ຣ	snake
	CCV	pʰria	ເພຣີອ	fire
	CCVC	plɔ:ŋ	ປລອງ	calf (of leg)
	Cv.CV(C)	ʰle:	ຈແລ	spade for sowing
	Cv.CCV(C)	sɪ'klo:p	ສໂກລປ	to clutch many things concurrently
	CvC.CV(C)	sir'ne:	ສິຣແນ	long iron stick for piercing a hole
rare structures	CvC.CCV(C)	ʰm'bra?	ຕຶມບຣະ	fireplace for cooking
	CvC.CvC.CVC	pinhin'ko?	ປິນຫິນໂກະ	to dress someone
	Cv.CVC.CVC	tə,lam'pa:m	ຕລຳປພມ	butterfly
	CV.CV	ʰɔ:rɔ:	ຈໍຣັ	to approach

3.4 Loan words, punctuation and word breaks

Words that are borrowed from Lao are written with a Kmhmu’ spelling which reflects the Kmhmu’ pronunciation. Following is an example of a borrowed word written in the Kmhmu’ orthography:

(Lao) /k□ml□ŋ ກຳລັງ / ‘in the process of’ vs. (Kmhmu’) /k□m'l□ŋ/ ກຳລັງ

The Kmhmu’ script utilizes a specially created half-space to separate words and a full space at the end of a sentence. Lao does not have word breaks, only clause and sentence breaks which serve the function of punctuation. Because the Kmhmu’ orthography does have word breaks it is necessary to introduce punctuation to identify clause breaks and sentence breaks, which are the comma and period respectively. Other punctuation markers were incorporated to indicate expression (question mark and exclamation mark) and to identify direct speech (double quotation mark).

4 The Kmhmu' orthography in use

As described in Section I, Preisig and Suksavang were the primary decision-makers in the development of the current orthography in the late 1980s. They were influential in bringing a degree of standardization to spelling through the production of the Kmhmu'-Lao-French-English dictionary (1994). These individuals continue to be influential in the promotion of the Kmhmu' orthography. The growing community of Kmhmu' that has knowledge of the orthography has found it acceptable and is using it. It is difficult to estimate how many people can read this orthography, but several hundred Kmhmu' can read this orthography with a moderate level of fluency.

The two unique Kmhmu' characters {ᵑ̃} and {ᵑ̃̄} were approved for inclusion in Unicode by iso/iec jtc1/sc2/wg2 in. This will aid in making the Kmhmu' font readily accessible.

4.1 Learning the Kmhmu' orthography

Because of the high degree of over-lap between the grapheme-phoneme correspondences in Lao and Kmhmu' a reading /writing guide that helps readers of Lao learn the Kmhmu' Lao-script based orthography was published in 2010. This guide begins by introducing the Kmhmu' graphemes that have the same sound-symbol correspondences in Lao and then leads the learners through the differences, providing reading practice and phonics exercises focused on to help the learner gain fluency. The contents of this 'transfer guide' can be taught adequately to readers of Lao in about 20 instructional hours, including practice. The elements of the orthography are introduced in the following order:

- (1) Consonants and vowels that are the same in Lao/ Kmhmu'
- (2) Final consonants in Kmhmu' not found in Lao
- (3) Initial consonants in Kmhmu' not found in Lao
- (4) Vowels in Kmhmu' not found in Lao
- (5) Consonant clusters
- (6) Use of the diacritics marking pre-glottalization and de-voicing
- (7) Multisyllable words

The most difficult elements of the orthography to learn for Kmhmu' speakers who can already read Lao are the *six final consonants in Kmhmu' that are not found in Lao*. Through experiences teaching the Kmhmu' orthography it has been found that learners need assistance developing phonemic awareness of final consonant sounds in general in the process of learning to read and write. Interestingly, the uniquely Kmhmu' initial consonants prove to be quite easy to learn. Other difficult elements to learn are *consonant clusters*.

4.2 Use of the orthography by speakers of other dialects

Suksavang et al claim that this orthography can be used by speakers of the 'Northern' dialect by assigning new meaning to various elements of the orthography to correspond to the phonologic features of that dialect. No documentation of this having actually been done is available, so it is unknown how viable this approach is. I suspect, though, that given the large variety of sub-dialects it would be difficult to capture all the differences adequately.

However, speakers of other dialects have used the transfer guide with a couple of outcomes. Some speakers, when encountering a word that is pronounced differently in their dialect (for example with tone in stead of de-voicing) the speaker automatically reads the word the way he normally speaks it. Other people have used the transfer guide to *learn* the southern dialect, and are motivated to do so because it is the largest dialect and the one used in most radio broadcasts and available literature. There are significant vocabulary differences between dialects in addition to phonological differences.

4.3 Other Kmhmu' orthographies

There is a second Kmhmu' orthography that has been developed which uses a Roman script (Suksavang et al). Young people in particular are seemingly more motivated to learn the Roman script than the Lao script because it 1) is perceived as being more uniquely Kmhmu' and 2) is perceived to be more prestigious because of its association with French or English. The Roman script-based Kmhmu' orthography is significantly more difficult to learn, however.

Work has been done by Mahidol University and SIL International in conjunction with Kmhmu' communities in Thailand to develop a Thai-based orthography for the one of the northern dialects of Kmhmu' spoken in Chiang Rai province. The extent of its use in the community needs to be explored.

5. Summary and outlook

The Kmhmu' Lao script-based orthography is an efficient representation of the Southern Kmhmu' dialect and is being used by the Kmhmu' community as an effective tool for written communication. This orthography is particularly easy to learn for mother tongue speakers of the Southern Kmhmu' dialect who can also read and write Lao. The elements of the orthography that are not found in Lao, such as the uniquely- Kmhmu' final consonants and consonant clusters, require extra practice to learn in order to gain reading fluency. This orthography could also be used in a tother-tongue-first literacy program and would assist learners bridge into Lao language literacy.

Further study needs to be done to investigate whether or not this orthography can be used by speakers of the Northern Kmhmu' dialects by assigning a number of alternate phoneme-grapheme correspondences.

Acknowledgements

I want to express my sincere appreciation to Dr. Sigrid Lew, SIL International, for her assistance in helping me prepare this manuscript. Her insights and guidance were invaluable. I am also grateful to those who collected the data that is in this paper, particularly Suksavang Simana', Elizabeth Preisig and Anne Osborne. Additionally, it is a privilege to work with the Kmhmu' people of Southeast Asia.

References

- Becker, Benjawan P. 2003. Lao for beginners. Bangkok: Paiboon Pomsan Publishing.
- Gafos, Adamantios I. 1999. The articulatory basis of locality in phonology. New York: Garland Publishers.
- Herr, Kristen. 2011. The phonologic interpretation of minor syllables, applied to Lemi Chin. Chiang Mai, THN Payap University M.A. thesis.
- Matisoff, James A. 1973. Tonogenesis in Southeast Asia. Consonant types and tone, ed. by Larry Hyman. Southern California Occasional Papers in Linguistics. 1.71-95.
- National Statistics Center of the Lao PDR. 2007. <http://www.nsc.gov.la> (Accessed 19 October 2009).
- Osborne, R. Anne. 2013. Phonology Statement for Kmhmu'. To appear in Mon-Khmer Studies Journal.
- Preisig, Elisabeth. 1990. The Kmhmu' orthography: Dialect of Xieng Khouang, Luang Prabang and Sam Neua. Unpublished manuscript.
- Smalley, William. 1961. Outline of Khmu structure. An essay of the American Oriental Society No 2, ed. by Edward H. Schafer. New Haven: American Oriental Society.
- Smalley, William. 1963. Orthography Studies: Articles on new writing systems. London: United Bible Societies in cooperation with the North-Holland Publishing Company, Amsterdam.
- Suksavang Simana', Somseng Sayavong and Elisabeth Preisig. 1994. Kmhmu'-Lao-French-English dictionary. Vientiane, Lao P.D.R.: Ministry of Information and Culture, Institute of Research on Culture.

- Suwilai Premsrirat. 2002. Thesaurus and Dictionary Series of Khmu dialects in Southeast Asia. Mon-Khmer Studies-Mahidol University Special Publications No.1. Institute of Language and Culture for Rural Development, Mahidol University, Salaya.
- Svantesson, Jan-Olaf. 1983. Kammu phonology and morphology. Malmö, Sweden: CWK Gleerup.
- Svantesson, Jan-Olof. 1989. Tonogenetic mechanisms in Northern Mon-Khmer. *Phonetica* 46.60-79.
- Svantesson, Jan-Olaf & Anastasia Mukhanova Karlsson. 2004. Minor syllable tones in Kammu. International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages. Beijing, China, March 28-31, 2004, 1-4.