

A DESCRIPTION AND LINGUISTIC ANALYSIS OF THE TAI KHUEN WRITING SYSTEM

R. Wyn Owen
Payap University
rwynowen@gmail.com

Abstract

This article provides a description and linguistic analysis of the Tai Tham script-based orthography of Tai Khuen, a Tai-Kadai language spoken in Eastern Shan State, Myanmar. The language has a long history of writing flowing out of the literary and religious culture nurtured by the Lan Na Kingdom from the 13th Century onwards. Comparison of phoneme and grapheme inventories shows that the orthography is well able to represent the spoken language as well as ancient Pali religious texts. Apart from spelling conventions reflecting the etymology of borrowed Pali and Sanskrit morphemes, sound changes over time have also decreased the phonological transparency of the orthography, notwithstanding the need of some more conservative varieties which still preserve distinctions lost in other varieties. Despite the complexities of the script, the literacy rates in Khuen are remarkably high for a minority language not taught in the government school system.

Keywords: Phonology; Orthography; Tai Languages.

ISO 639-3 codes: kkh, pli, tha, nod, khb, khm

1. Introduction

Tai Khuen – hereafter Khuen /k^hũn/ – is a Southwestern Tai language with around 100,000 (Diller 1994) speakers in Myanmar. Lewis et al. (2016) cite a Khuen population in Thailand in 2000 of 6,280 while Schliessinger (2003:90) reports one Khuen village of around 600 inhabitants in Laos. In Myanmar, Khuen <Hkun> is one of the 135 ethnic groups officially recognised by the government (Ananda Travel 2009). The Khuen capital, Kengtung was established by Mangrai as a northern fortress in 1267 (Ongsakul 2005). The Khuen were introduced to the Lan Na orthography in the 13th century (Sai Kham Mong, 2004). Since that time the orthography used by the Khuen has developed distinctively from the Lan Na orthography. Literacy has developed a high value in Khuen culture. This paper describes the orthography and analyses its capacity to represent the language and ancient Pali religious texts as well as assessing its user-friendliness. Both Egerod (1959) and Petsuk (1978:110-117) have described Khuen orthography in relation to the dialect spoken in Kengtung. This article is written from a new vantage point in time and incorporates ways both the language and orthography have changed since those accounts were written as well as knowledge of contemporary dialectal variation (Owen 2008; 2012). Furthermore, at a time when use of the orthography in electronic media is growing, aspects of the current state of development of Khuen in this domain are also addressed.

After a brief overview of the typological features of the language and the literary culture, Khuen phonology is summarised with attention given to a key theoretical issue in Tai phonology, namely the phonemic status of the glottal stop. Historical development of the Khuen language and orthography are briefly described and the defining characteristics of the Khuen writing system considered. Section 6 lays out the grapheme-phoneme correspondences and Section 7 gives a brief assessment of the user-friendliness of the orthography. I would like to thank the monks of Wat Tha Kradas, Chiang Mai and Wat Pa Daek, Mae Sai for their patience in answering my many questions in the preparation of this paper.

General terminology in this paper follows Cook & Bassetti (2005:4). A writing system is ‘the overall way in which written symbols connect to the language (e.g. alphabetic, syllabic writing systems)’. The term ‘script’ refers to ‘the physical implementation of a writing system e.g. the Roman and Cyrillic alphabets for

alphabetic writing systems)’. We note that the same script may be used to write many different languages. Finally an ‘orthography’ is defined as, ‘the rules for using a script in a particular language (e.g. the English or Italian orthography for the Roman alphabet).’ So the Khuen orthography is based on Tai Tham script which is also used by Northern Thai and Tai Lue whose orthographies differ slightly from Khuen.

2. Khuen language and culture

In common with other Tai peoples, the Khuen are valley-dwelling wet rice growers. Khuen cultural history is closely connected to that of the Lan Na kingdom founded by Mang Rai in the late 13th Century and based in Chiang Mai (Wyatt 2003:70). Not least in this was the introduction of the Lan Na (Tai Tham) script (Penth 1994:13). The prominent role of the Buddhist clergy in the intellectual life of the community remains a pillar of support in the maintenance of Khuen language and literacy (Peltier 1996:20). Although not taught in government schools, Khuen literacy is taught in Buddhist temples since Khuen is the main language of Buddhist written materials in Kengtung and neighbouring townships. Furthermore, Owen’s (2008) study of literacy in five Khuen villages shows that men spent on average around 7 years of their lives as members of a temple community and almost 100% developed high literacy proficiency in Khuen. Although not members of the temple community, many women (around 50% in Owen’s study) also become literate in Khuen, some through classes held in local temples by the Khuen Cultural and Literature Committee and some through informal instruction from family members.

Tai languages are isolating tonal languages with a ‘propensity for monosyllabic morphemes’ (Morey 2005:208). Many polysyllabic morphemes in modern varieties are borrowed from other languages – in the case of Khuen from Burmese, Pali, Sanskrit and English (Petsuk 1978:26). In common with other Tai languages – see for example Iwasaki and Ingkaphirom (2005:25-48) for Standard Thai – compounding is the most productive derivational process in Khuen (Petsuk 1978:56-67). Reduplication is often used to derive an adverb from an adjective (Bhikku 2008) or as an intensifying device (Petsuk 1978:63). Affixation is quite limited in Khuen with Petsuk (1978:65) listing just five bound morphemes.

Petsuk (1978:26) posits the typical phonological syllable in Khuen monosyllabic morphemes or stressed syllables in polysyllabic morphemes as $C(C)V(C)^T$, that is, a compulsory initial consonant followed by a highly restricted optional second consonant, an obligatory vowel nucleus carrying an obligatory tone followed by an optional consonant final. Note also that the vowel can be the syllabic nasal [ṁ]. This structure provides a framework for interpretation of individual segments as described below. Unstressed syllables in polysyllabic morphemes conform to a CV^T template where the vowel is normally, but not always, [a] and the tone is also highly restricted but not enough to be considered toneless (Petsuk 1978:28-29). Whether or not the onset is obligatory is open to debate and will be discussed in Section 3.1 below.

3. Khuen phonology

Various authors have described the phonemes of Khuen in Myanmar (Egerod, 1959; Hudak, 1994 based on fieldwork in 1964 by Gedney; Petsuk, 1978; Owen, 2008). There is a high degree of commonality between the accounts, but also differences attributable to dialect and time. Diachronic changes and contemporary variation are discussed in more detail in Owen (2008; 2012). Generally speaking, the direction of change is towards fewer contrasts resulting in smaller phonemic inventories. Different dialects are at different stages in the process, and more conservative varieties preserve some contrasts that have been lost in others. For simplicity, this paper will use the more conservative descriptions (Egerod 1959; Hudak 1994) as a basis for discussing grapheme-phoneme correspondences, but segments whose phonemic status is questionable are discussed.

3.1 Status of glottal stop in Khuen

The status of the glottal stop in SW Tai languages has been much discussed with two basic positions. Taking the non-phonemic position, Morey (2005:112) argues against transcribing an initial phonemic /ʔ/ for the Tai languages of Assam because it is often omitted in connected speech, whereas other initial unaspirated stops are not omitted. Noss (1964:9) argues that although the glottal stop occurs phonetically in both initial and final positions in Standard Thai, it is not phonemic in either because its presence is determined respectively by the onset or end of either loud or normal stress phonemes. On the other hand for the phonemic position, Egerod gives phonemic status to the glottal in initial position, but in final position its occurrence is

predictable following short vowels in Khuen (1959:125-6) or with three particular tones in Shan (1957:123-5). Hudak’s approach also posits phonemic /ʔ/ for Khuen (1994:978; 1022). Reconstructions of Proto-Southwestern Tai invariably posit a phonemic glottal stop (Pittayaporn 2009a; Li 1977; Jonsson 1991; Sarawit 1973; Brown 1985) which patterns with other glottalized or implosive consonants rather than other voiceless stops in relation to tonal splits and mergers. In initial position, [ʔ] contrasts with other oral stops in Khuen (Petsuk 1978:7) but, as for the Tai languages of Assam, ‘no minimal pair can be found to distinguish a glottal initial from a vowel initial’ (Morey 2005:113), so the contrast could be due to the absence of an initial rather than the presence of a glottal stop. The fact that what are regarded as vowel-initial syllables by proponents of the non-phonemic glottal position are restricted to certain tonal categories is not sufficient to establish the phonemic status of the glottal.

Duanmu (2010:57-8) points out that although an initial glottal stop is often pronounced when a word is spoken in isolation, in connected speech the glottal is not pronounced. He concludes that the glottal initial is an ‘unintended gesture: the vocal tract cannot assume the vowel gesture all of a sudden, and the glottal stop reflects an unintended state before the vowel is pronounced’ (2010:73).

In Khuen, as in southwestern Tai languages in general, the final glottal stop is limited to short vowels and predictable in this environment, whereas other stops can occur after both long and short vowels. This presupposes that vowel length is phonemic, the position taken in this paper, although as Li points out, the functional load of vowel length in Tai languages is low (1964:12). This will be discussed in more detail in Section 3.3 below. In summary, then, this paper takes the glottal stop to be non-phonemic.

3.2 Khuen Consonant Phonemes

Egerod (1959:125) identified 20 initial consonant phonemes in naturally spoken Khuen. These are shown in Table 1, with the exception of glottal stop which in this paper it is not considered phonemic – see Section 3.1. For conciseness, /c/ is used to represent the voiceless laminal pre-palatal affricate [tɕ].

Table 1: *Khuen initial consonant phonemes (adapted from Egerod 1959)*

Manner/Place of articulation	Labial	Alveolar	Aveolopalatal	Velar	Glottal
Oral stop	b p p ^h	d t t ^h	c [tɕ]	k k ^h	
Nasal stop	m	n		ŋ	
Fricative	f	s			h
Liquid		r l			
Approximant	w		j		

Subsequent authors describing spoken Khuen have largely agreed with Egerod’s phonemes, though some differences have been identified. Petsuk (1978:11-15) identified only 16 phonemes, analysing /b/, /d/ and /f/ as occasional reflexes of the phonemes /w/, /l/ and /p^h/ respectively. Petsuk’s informant was from near the state capital Kengtung, where Khuen is in contact with and heavily influenced by Shan (Egerod 1959:123-4; Owen 2012:19-20; 29). Hartmann’s comparison of modern reflexes of Proto-Southwestern Tai consonants neatly explains how contact with Shan might lead to Khuen /b d f/ shifting to /w l p^h/ (Hartmann 2008). Owen (2004) identified the same phonemes as Egerod in a dialect from an area more remote from the capital.

In addition to the phonemes of naturally spoken Khuen in Table 1, Egerod (1959:125) lists six phonemes used in ‘reading the alphabet aloud and in learned pronunciation of Pali words,’ viz. /g, g^h, c^h, ɲ, d^h, β/. So in addition to the inventory of phonemes in spoken Khuen, some speakers have an additional inventory used for reading Pali or pronouncing the alphabet. As can be seen by comparing this list to the full list of consonant graphemes in Section 6 below, not all graphemes used for Pali words have their traditional Pali pronunciation. In line with her general trend of recognising fewer phonemes in Khuen, Petsuk (1978:14) reported only one extra phone [tɕ^h] in reading pronunciation. The phoneme /r/ was absent from Petsuk’s data but present in Owen (2004) both in native Khuen morphemes and those derived from Pali. We note in passing that SWT varieties that preserve /r/ also often preserve /b d/ (see comparison of modern reflexes in Hartmann 2008.) In other words, /b d/ tend to pattern together and are likely to be retained for longer than /r/ and the evidence from different authors on Khuen supports this pattern.

Previous authors agree that there are two initial consonant clusters in ordinary Khuen words, namely /kw/ and /k^hw/, but they feature in only a relatively small number of words, with the second element often

subject to elision in natural speech. In addition to the above, Petsuk (1978:16-17) reported the following clusters in loan words from Burmese, Sanskrit or Pali: /pj p^hj mj kj tw sw/. Egerod reported /mj, kj/ in ‘Burmese-Shan loanwords’ as well as /tw t^hw sw/ and /tr t^hr sr/ (1959:123).

Syllable codas in Khuen are limited to final nasal and oral stops /p t k m n ŋ/ and the approximants /w j/. As noted above, in final position the glottal stop is predictable – limited to short vowels – and is thus not phonemic. The other three oral stops have no audible release phonetically. We note that the inventory of finals is less than half the size of the initials. Discussing a similar pattern for his reconstruction of Proto-Tai, Pittayaporn (2009b:193-4) points to neutralisations of position and phonation type which cause the reduction of the large onset inventory to the highly limited coda inventory. Such asymmetries between onset and coda inventories are common in Southeast Asia (Rhee 2003:140-141).

3.3 Khuen vowel phonemes

The Khuen vowel phonemes as identified by Egerod (1959) and Hudak (1994) are presented in Table 2. The familiar nine Tai vowel positions are attested with length distinctions in each position.

Table 2: *Khuen vowel phonemes*

	Front	Central	Back
High	i i:	ɯ ɯ:	u u:
Mid	e e:	ɤ ɤ:	o o:
Low	ɛ ɛ:	a a:	ɔ ɔ:

Egerod (1959:125) commented that there was some tendency to confuse long and short versions of the high vowels. Petsuk (1978:6) identified phonetic length differences in all nine positions for spoken Khuen but only found evidence of contrast for the low vowels. Owen (2008:307) observed that the large difference in frequency of occurrence of the different Khuen vowels meant that the functional load of the length distinction varied greatly from position to position. Furthermore, contemporary Khuen dialects show different patterns of phonetic length realisation (Owen 2008:307). Length contrasts do not exist for all syllable types. As mentioned above, in the absence of any other consonant final, short vowels are always followed by a (phonetic) glottal stop whereas long vowels never occur before a final glottal. Furthermore, since vowel length distinctions in checked syllables condition tone distinctions, contrast in identical environment is not found for checked syllables. Contrast is found in unchecked syllables, particularly those ending in a nasal or semivowel. All told, combined with the omission of glottal stop from the consonant inventory, positing phonemic length for all nine vowels positions provides a neat, perfectly symmetric solution with slight possible overdifferentiation in the case of less-common vowels. This will be discussed further in relation to the orthography in section 6.3 below.

3.4 Khuen tones

Older descriptions of Khuen phonology identified six phonemic tones (Egerod 1959; Hudak 1994) whereas more recent studies identified only five (Petsuk 1978; Owen 2003). Owen (2012) found both five-tone and six-tone systems in contemporary use. He presented evidence pointing to contact with 5-tone Shan as the cause of a loss of the low rising tone which had no phonetic counterpart in Shan. As will be shown in Table 4, the low rising tone (#3) merged with the low falling tone (#4). Both five- and six-tone systems are described in Table 3. Phonetic characterisations use a 5-point scale with ‘5’ denoting high pitch and ‘1’ low pitch. Phonemically the tones are labelled 1-6 as shown in the first column. Following Gedney (1973), tones are numbered according to the tones on unchecked syllables and the phonetic characterisation is for unchecked syllables as shown in the examples in the final column. Tones on checked syllables are associated with the nearest phonetic counterpart on unchecked syllables, although different phonetic environments often produce different pitch contours, not least because of vowel length differences. Petsuk (1978:24) reports that when the tone of the preceding syllable ends in high pitch, Tone 2 is realised as high rising [45].

Table 3: Khuen phonemic tones

#	Owen (2012)		Gedney [1964]	Egerod (1959)	Petsuk (1978)	Example
	6-Tone System	5-Tone system				
1	falling rising [215]	falling rising [325]	rising [35]	high rising	low rising [15]	ka: ²¹⁵ 'crow'
2	high [44]	mid [32]	high [44]	high level	low mid slightly rising [23]	ka: ⁴⁴ 'car'
3	low rising [13]	-	low with slight rise [12]	low slightly rising	-	ka: ¹³ 'charm'
4	low [21]	low [21]	low [22]	low falling	low falling [21]	ka: ²¹ 'price'
5	mid glottalized [33ʔ]	mid glottalized [33ʔ]	mid glottalized [33ʔ]	mid slightly falling with glottal stop	mid glottalized [32ʔ]	ka: ^{33ʔ} 'dance'
6	high falling [51]	high falling [41]	high falling glottalized [51ʔ]	high falling	high falling [51]	ka: ⁵¹ 'trade'

The descriptions of the tones in Table 3 show that there is general agreement between authors when comparing systems with the same number of tones. Apart from the number of tones, the most noticeable difference between the five- and six-tone systems is the different pitch height of Tone 2. In the five-tone system, Tone 2 is much closer in height to Tone 5, whose glottal constriction is a key auditory cue in differentiating the two tones (Owen 2012:26).

Table 4 uses Gedney's (1973) tone box to show how the tones in both the five- and six-tone systems are distributed over 20 possible tonal categories defined by proto-initial and proto-tone category. The distribution of tones is consistent for authors who report the same total number of tones. Egerod (1959), Gedney [1964] and Owen (2004) report the distribution of the six-tone system in Table 4 whereas Petsuk (1978) and Owen (2012) report the distribution of the five-tone system. A checked syllable is one that ends in one of the stops /p t k/ (short or long vowels) or [ʔ] (short vowels only). An unchecked syllable on the other hand ends in a long vowel, nasal /m n ŋ/ or semi-vowel /w j/.

Table 4: Gedney tone box showing distribution of Khuen phonemic tones (adapted from Owen 2012)

		Proto-Tai Tonal Categories									
		Six-tone system					Five-tone system				
		Checked			Unchecked		Checked			Unchecked	
		A	B	C	DS	DL	A	B	C	DS	DL
Initials at time of tone splits	Voiceless friction sounds *s, m̥, p ^h , etc.	1	3	5	2	3	1	4	5	2	4
	Voiceless unaspirated stops *p, t, k, etc.	1	3	5	2	3	1	4	5	2	4
	Glottalised sounds *ʔ, ʔb, etc.	2	3	5	2	3	2	4	5	2	4
	Voiced sounds *b, m, l, etc.	2	4	6	6	4	2	4	6	6	4

Comparing the distributions of tones in the two systems in Table 4, it can be seen that the only difference is caused by the merger of Tone 3 with Tone 4 in the five-tone system. In both systems, B=DL, that is, the B and DL columns show the same distribution pattern, a common feature of tone splits in Tai (Gedney 1989:216). The split in the A-column between rows 2 and 3 is not common in SWT as a whole, but is shared with Northern Thai and Lue Mong Yawng (Eastern Shan State, Myanmar) but not with Lue Jinghong (Xishuangbanna, China) (Owen 2008).

4. Khuen language development and history

The Khuen orthography is based on the Tai Tham script which was derived from Mon (which itself was ultimately derived from Brahmi) in the Lan Na kingdom of the 13th century (Hartmann 1986:9). Northern Thai and Tai Lue are closely related languages whose history shares the use of orthographies based on Tai Tham script, but despite strong maintenance of the spoken forms of those languages, their current use of their Tham-based orthographies is limited. In 2009 the Tai Tham script was incorporated into the Unicode Standard (version 5.2) (The Unicode Consortium 2009), however, use of Khuen Unicode is hampered by rendering problems in major computer and mobile applications as well as the need for a special keyboard. This manuscript for example was prepared with a combination of Unicode and legacy fonts to overcome rendering problems. Trager (2014) describes some of the challenges in creating fonts for Tai Tham. See also Figure 1 and Table 5 for a sense of how multiple graphemic elements can over crowd the space around a base symbol.

Following an unpublished religious dictionary (Bhikku 2005), a Khuen-English pocket dictionary (Bhikku 2008) was published and then a large English-Khuen reference dictionary (Kana Pap Aphithan 2008). Since these dictionaries are relatively new and not known to the whole Khuen community, it is not surprising that the process of orthographic standardisation is not yet finished. Different temple schools have different conventions in the composition of certain complex vowel graphemes, in spelling, and in the dictionary order of vowels. Peltier (1996) and the primers published by the Committee for the Preservation of Native Language and Culture (Kana Anurak Phasa lae Wattanatham Puen Mueng 2011) were taken as the main sources for contemporary usage of the script.

Since the adoption of the present orthography for Khuen, the spoken language – along with all other Tai languages – has undergone many phonological changes. Pittayaporn (2009) proposed specific phonetic characterisations of the four tones of Proto-Tai using the four ‘dimensions’ of pitch height, voice quality, pitch contour and vowel duration. Subsequently all Tai varieties underwent what is termed a ‘binary register split’ conditioned by an onset voicing contrast which cut across all four tonal categories (see Pittayaporn 2009 for a description). Liao (2016) described two stages of post Proto-Tai tonal development. In the first stage, voiceless and glottalized initials developed a high register and voiced initials developed a low register. In the second stage, glottalized initials split from the high register to form a third register, although the combination of syllable types with registers created too many tonal distinctions so the glottalized initials often merged into the low register. This development is reflected in the orthography by the division of the consonant inventory into three ‘classes’, namely ‘High’; ‘Mid’; ‘Low’ – see Table 12 for details of which consonants are in which classes. These labels do not reflect current phonetic tones. The high and low classes represent the historic high and low registers respectively while mid class represents the register created by the splitting of the glottalized sounds. So Khuen consonant initials contain inherent tonal information which combines with syllable type and tone diacritics to determine the correct tone. Like the Thai orthography but unlike contemporary Lao, Khuen spelling still reflects the Indic origins of loanwords. This decreases the phonological transparency of the orthography, particularly where finals are concerned, meaning that many spellings have to be memorised because they cannot be predicted by their pronunciation alone.

5. Defining characteristics of the Khuen writing system

Khuen is written from left to right and starting from the top of the page. The basis of the Khuen orthography is the orthographic syllable, a template for which is shown in Figure 1. Each syllable has:

1. an obligatory base symbol with its inherent vowel and tone class;
2. (optionally) up to four vowel symbols;
3. an optional medial consonant (prescript, subscript or postscript);
4. an optional final consonant (subscript, postscript or superscript); and
5. an optional tone diacritic which interacts with syllable type and initial consonant class to determine the tone.

The superscript consonant is limited to final ^ᵃ <-r> and ^ᵃ <-ŋ> in syllables where a subscript vowel prevents the use of a subscript final consonant (Petsuk 1978:112-3). Superscript forms are mainly found in handwritten text, whereas regular forms of these consonants in postscript position are the norm for printed texts. We note that the Unicode Standard includes codes for both superscript /ŋ/ and /r/ so it is possible to create Unicode fonts that use them.

For the vast majority of syllables, the base symbol is a consonant grapheme which contains an inherent /a/ vowel. Khuen also has a class of initial or independent vowel graphemes which can also form the base. Although there are instances of some optional template elements being added to an initial vowel base, in the vast majority of syllables involving an initial vowel grapheme, the vowel grapheme alone comprises the whole syllable, often the first syllable of a polysyllabic loan from Pali.

Figure 1: *Khuen orthographic syllable (adapted from Lew (2014:30))*

		Modifying Tone Mark		
		Superscript Vowel	Modifying Tone Mark (or Superscript Consonant)	
Prescript Vowel	Prescript Consonant	Base Symbol	Postscript Vowel	Postscript Consonant
		Subscript Vowel or Consonant	Subscript Vowel or Consonant	

The examples in Table 5 illustrate how the various positions can be filled. For each numbered example, the phonemic transcription is given underneath the orthographic. To show the symbol placement a schematic depiction is used in which C denotes a consonant; V denotes a vowel or vowel component; and T denotes a tone mark. The base symbol is given in boldface.

Table 5: *Illustrative examples for Khuen orthographic syllable*

Num.	Sample Word	Symbol Placement	Gloss	Num.	Sample Word	Symbol Placement	Gloss
(1)	၉က /e:k³/	V - C	'one'	(2)	၉[၉ /bro:t³/	V - C - C C	'help' (royal)
(3)	၉၉ /pɾ²/	Y V - C - V V	'mud'	(4)	က၉ /ka:n¹/	C - Y C	'work'
(5)	၉ /kiw⁶/	Y - T C C	'eye brow'	(6)	၉ /kʰaw³/	T Y V - C - V	'knee'
(7)	၉၉ /kwa:j²/	C - Y C C	'buffalo'	(8)	၉၉ /kwe:n¹/	C - C CV	'ox-cart'
(9)	၉ /cu:p³/	C - V C	'kiss'	(10)	၉ /r:¹/	Y V - C C - V	'be left over'
(11)	၉ /lɔ:t³/	C - C C V	'tube'	(12)	၉ /nɔ:ŋ¹/	C - C VC	'swamp'
(13)	၉ /cʰa²la:t⁴/	C - Y C C	'clever'	(14)	၉ /sa²de:ŋ²/	V - C - C C	'show'

Examples (1) and (2) illustrate respectively an independent vowel base and prescript vowel and consonant while the maximal number of components (four) in a compound vowel are shown in (3). (4) shows how a final consonant can be subscript to a postscript vowel. (5) and (6) illustrate the placement of tone markers in the presence of a superscript vowel component. (7) illustrates an initial cluster plus a postscript final consonant whereas (8) shows both subscript consonant and subscript vowel preceding a

postscript consonant final. (9) shows the allograph of the vowel /u:/ which is used when the subscript vowel position beneath the base symbol is occupied, in this case by the grapheme for final /-p/. Examples (10)-(12) illustrate the crowding that occurs when both a consonant and a vowel component occur in the subscript position (Trager 2014). (11) shows how the final consonant grapheme can be placed above the medial /ɔ:/ grapheme which is always subscript. Examples (13) and (14) both include phonological minor syllables. (14) shows that the vowel of the main phonological syllable can be placed in prescript position, even preceding the prescript consonant of the minor syllable.

As can be seen from Figure 1, the stacking of graphemes above and below the base symbol means that a line of Khuen text will have much greater depth than an orthography in which graphemes are placed in a horizontal sequence only. This is one of the many features that the Khuen writing system shares with other Brahmi-derived writing systems in South and Southeast Asia. Bright coined the term ‘alphasyllabary’ for such writing systems in which vowels are denoted by symbols that are subsidiary to consonants and not necessarily in linear temporal order of pronunciation (see Daniels & Bright 1996:xxxix for a definition of alphasyllabary and Bright 1999 for further discussion). Lew (2014:31) pointed out that the emphasis on graphic prominence in the definition of alphasyllabary narrows its scope and proposes the term ‘syllabet’ for a writing system ‘defined by the use of orthographic syllables constituted by phoneme-based graphemes in non-linear graphic arrangement.’ While the Khuen writing system clearly has many features in common with other writing systems classed as alphasyllabaries by Bright, the superior status of consonants over vowels, reflected in graphic prominence (size and/or positioning of graphemes), is not the most salient characteristic for Khuen. As can be seen from Figure 1 and the examples in Table 5, consonants can be subscript or superscript and final consonants can be subscript to vowels. Moreover, independent vowel graphemes can act as the base symbol. Non-linear ordering of elements in the orthographic syllable around a base symbol is a much clearer defining characteristic, and for this reason it is preferable to class Khuen as a syllabet. A sample Khuen text is given in the Appendix which gives an impression of the overall appearance of the orthography in use to complement the following presentation of individual graphemes.

6. Khuen orthography

The following four subsections relate the graphemes in the Khuen orthography to the phonemes. Subsection 6.5 describes numerals used in Khuen as well as punctuation.

6.1 Khuen consonant graphemes

Table 6 shows the 42 single consonant graphemes of Khuen in dictionary order with their corresponding phonemic values. The dictionary order here (read left to right across rows) follows Bhikku (2008), although the grapheme 𑄛 is often listed following 𑄚. All of the graphemes can serve as onsets apart from the two sibilants 𑄛 and 𑄜 which are used for writing codas in words derived from Sanskrit. The upper case subscript letters on each phoneme denote the consonant class of that phoneme: _{-L}; _{-M}; _{-H} denoting ‘Low’; ‘Medium’ and ‘High’ classes respectively. The role and meaning of consonant classes will be explained in Section 6.4. The glottal in the final row is enclosed in square brackets to show that it is not considered phonemic in this paper. Further insight into the alphabetical order can be gained by splitting the consonants into two groups: those used to write Pali and those that are not. In Table 6 the boxed graphemes are those which are not used for consonant phonemes in Pali. The two boxed /s_H/ phonemes are used in morphemes derived from Sanskrit.

It is of interest to look at these boxed graphemes to see why they were needed to write Khuen words. The four graphemes 𑄛 𑄜 𑄝 𑄞 representing phonemes posited to have been fricatives at the time the orthography was developed, can be seen to share the common feature that they were formed by adding a ‘left hook’ above the graphs for homorganic oral stops 𑄛 𑄜 𑄝 𑄞 respectively. That 𑄛 should be considered to represent the voiced velar fricative [ɣ] is based on the fact that Khuen words spelled with 𑄛 are reconstructed in Proto-Southwestern Tai (hereafter PSWT) with *ɣ (Bhikkhu 2008; Li 1997; Jonsson 1991).

Although Tai Tham is believed to have developed from Mon script and the Sukothai script of Ramkhamhaeng from old Khmer, Hartmann (1986) points out that the close connections between the kingdoms of Lan Na (based in northern Thailand) and Sukhothai (central Thailand) led to an awareness of the Sukhothai script in the Lan Na region. Brown (1985:10-11) describes how the Sukhothai language distinguished voiceless and voiced velar fricatives (represented by 𑄛 /x/ and 𑄜 /ɣ/), but the distinctions became redundant after phonemic mergers, and the graphemes became obsolete. When Edwin McFarland was struggling to find enough keys to accommodate all the Thai graphemes on the first Thai typewriter in

1892, these two graphemes were omitted and are generally not used for spelling in contemporary Thai dictionaries (Smalley 1997:247; Haas 1956:4).

Table 6: *Khuen initial consonant graphemes*

Place of Articulation in Pali/Sanskrit	Voiced Stop	Voiceless Stop				Nasal Stop	
		Unaspirated		Aspirated			
Velar		က	k ^h _H	ခ	k ^h _H	ခ	ၵ _L
		ဂ	k _L	ဂ	k ^h _L		
				ဆ	k ^h _L		
Palatal		စ	c ^h _H	ဆ	c ^h _H	ည	ၵ _L
		ဇ	c _L	န	s _L		
				ဆ	c ^h _L		
Retroflex	ၵ	ဆ	t ^h _H	ဗ	t ^h _H	ဇ	ၵ _L
				ဆ	t ^h _L		
Dental	ၵ	တ	t _H	ခ	t ^h _H	ဇ	ၵ _L
		ဒ	t _L	ဝ	t ^h _L		
Labial	ဝ	ပ	p _H	ဖ	p ^h _H	လ	ၵ _L
				ဖ	f _H		
		ပ	p _L	န	f _L		
Resonants	ယ	လ	r _L	လ	l _L	ဝ	w _L
				လ	p ^h _L		
Other	ဇ	ဆ	s _H	ဆ	s _H	ဇ	h _H
	ဇ	ၵ	ၵ	j _H	ၵ		

We note in passing that there are currently two distinct character styles of the Tai Tham script: Northern Thai and Khuen-Lue. Northern Thai style is distinguished by finer strokes of varying weight, ‘head’ terminals (small circles at the point where the pen first touches the page when handwritten) and more complex shapes. Khuen-Lue style on the other hand has uniform heavier strokes, no head terminals and simpler shapes. Based on the chronology of text samples in Sai Kam Mong (2004), the Khuen-Lue style appears to have developed to be distinct from the Northern Thai style during the 19th century. Many contemporary Khuen-Lue style character shapes are identical to corresponding characters in the Myanmar script. The first three consonants in Northern Thai style are က ခ ခ whereas in Khuen Lue style the same characters are က ခ ခ.

The reason why Khuen – and Lue (Jagacinski 1986:84) – used ခ but not the corresponding high class consonant ခ as seen in Lan Na is not clear. The ခ doesn’t appear in any of Sai Kam Mong’s inventories of Khuen consonants, the oldest being dated 1766 (Sai Kam Mong 2004) or Peltier’s works in which popular

Khuen stories from ancient manuscripts are republished in book form along with translations into Thai, French and English (Peltier 1999; 2005; 2006).

The two-volume Northern Thai Dictionary (Rungruangsri 1990) shows that the number of words spelled with high class ๕ (Khuen-Lue ๕) is small (3 pages) compared to the 30 pages for the low class ๖ (Khuen-Lue ๖). Rungruangsri explains that these two graphemes are mainly found in palm leaf manuscripts in the provinces of Phrae and Nan and seldom in Chiang Mai or Lamphun (1990:206). More recently, the single-volume Northern Thai Dictionary (Sun Watthanatham 1996) doesn't use ๕ at all, although interestingly the use of ๖ remains strong in Khuen and Northern Thai.

One possible reason that there is more evidence for ๕ in Northern Thai than Khuen or Lue is that Northern Thai was written earlier, so perhaps this grapheme was in use at that time, but was waning by the time the Khuen and Lue started using the script. Or perhaps the phonological loss of /x/ occurred sooner in Lue and Khuen. Further research is needed to shed light on this matter.

Table 7 shows the 33 consonant graphemes used for Pali together with their associated Pali phonemes and their phonemes in contemporary Khuen.

Table 7: Khuen Graphemes used to write Pali (adapted from Peltier 1996:36)

	Voiceless Unaspirated			Voiceless Aspirated			Voiced Unaspirated			Voiced Aspirated			Nasal		
Velar	က	k	k _H	ခ	k ^h	k ^h _H	ဂ	g	k _L	ဆ	g ^h	k ^h _L	ခ	ŋ	ŋ _L
Palatal	စ	c	c _H	ဆ	c ^h	c ^h _H	ဇ	ʃ	c _L	ဆ	ʃ ^h	c ^h _L	ဆ	ɲ	ɲ _L
Retroflex	စ	ʈ	t _H	ဗ	t ^h	t ^h _H	ဇ	ɖ	d _M	ဆ	ɖ ^h	t ^h _L	ဆ	ɳ	n _L
Dental	တ	t	t _H	ဆ	t ^h	t ^h _H	ဒ	d	t _L	ဆ	d ^h	t ^h _L	ဆ	n	n _L
Labial	ပ	p	b _M	ဖ	p ^h	p ^h _H	ပ	b	p _L	ဆ	b ^h	p ^h _L	ဆ	m	m _L
Other	ယ	j	j _L	ရ	r	r _L	လ	l	l _L	ဝ	v	w _L	ဆ	s	s _H
Other	ဆ	h	h _H	ဆ	ʎ	ʎ _L	း	~	-ŋ _L						

The IPA transcriptions of the Pali phonemes in Table 7 are based on the descriptions in Duroiselle (2007). Since the Khuen orthography has graphemes for all of the Pali phonemes it is no surprise to read that, ‘While the other Shan scripts are short of letters for writing Pali, all the Buddhist scriptures are written perfectly in Hkun script’ (Sai Kam Mong 2004:238-9). Of particular note is the final grapheme which is not a consonant but a vowel nasalisation symbol known as /maj⁵² kaŋ¹⁵ tɔ:m¹³/ in Khuen (Sanskrit: anusvāra; Pali: nigghahīta). In addition to nasalisation of the short vowels /a i u/ in Pali words, the symbol came to be associated with a nasal final: /-m/ in Burmese (Duroiselle 2007) and /-ŋ/ in Khuen (Peltier 1996:106). Over time, nasalisation has been lost and only the nasal final remains. Note that the Pali phonemes are not the same as the phonetic realisation when contemporary Khuen readers read a Pali text. Instead of the voicing distinctions between data columns 2 and 5 (voiceless) and 8 and 11 (voiced), tone is used to differentiate the plosives in columns 3 and 6 from their counterparts in columns 9 and 12. The palatal nasal is considered as /j_La⁶/ in Pali words.

The ‘karan’ symbol ⁵ is written over a consonant (normally in final position) when that consonant is not to be pronounced. This is frequently used in loans from languages with consonant clusters in the coda such as Pali (e.g., သိန္ဓိဝိသိ /som¹bu:n²/ ‘perfect’) or English (e.g., ဖိမ် /fim²/ ‘film’).

The repeat marker ၅ [maj⁵² ja⁴mok⁵³] is used to indicate reduplication of the preceding word. Adverbs, for example, are often derived by reduplicating an adjective.

6.2 Khuen orthographic clusters and finals

Orthographic clusters in Khuen add complexity because they often use subjoined forms of the consonant graphemes. These are used in syllable final position, and also in syllable initial clusters or word-medial

clusters that span a syllable boundary. We note in passing that clusters in Brahmi were written as conjunct forms which made it clear that the inherent vowel in the first consonant was not pronounced (Salomon 1996:376). Table 8 lays out the base form of each grapheme together with their subjoined forms and the phonemic value of the subjoined form.

Seven of the graphemes whose base form is wide have both wide and narrow subjoined forms which illustrate the way the Khuen orthography adapts grapheme shapes to the space available. Such adaptations help to maximise visual distinctiveness of different graphemes in handwritten texts while also minimising use of space and thus improving economy. Economy of space was important when the cost of writing materials was relatively high. A similar space saving motivation can explain the use of superscript final consonants as illustrated in Figure 1.

Table 8: Subjoined Forms of Consonant Graphemes

Base Form	Subjoined Form	Phoneme	Base Form	Subjoined Form	Phoneme	Base Form	Subjoined Form	Phoneme	Base Form	Subjoined Form	Phoneme
က	က̄	-k	ခ	ခ̄	-k	ဂ	ဂ̄	-k	ဃ		
ဆ	ဆ̄	-k	ဓ	ဓ̄	-ŋ	စ	စ̄	-t	ထ	ထ̄	-c ^h -
ဒ	ဒ̄	-t	ဇ			ဗ	ဗ̄	-c ^h -	ည	ည̄	-n
ဗ	ဗ̄	-t	ဗ	ဗ̄	-t	ည	ည̄	-t	ဃ	ဃ̄	-t
ဇ	ဇ̄	-n	ဇ	ဇ̄	-t	ဇ	ဇ̄	-t	ဇ	ဇ̄	-t
ဇ	ဇ̄	-t	ဇ	ဇ̄	-n	ဇ	ဇ̄	-p	ဇ	ဇ̄	
ဇ	ဇ̄	-p ^h -	ဇ	ဇ̄		ဇ	ဇ̄	-p ^h -	ဇ	ဇ̄	
ဇ	ဇ̄	-p	ဇ	ဇ̄	-m	ဇ	ဇ̄	-j	ဇ	ဇ̄	-n
ဇ	ဇ̄	-n	ဇ	ဇ̄	-w	ဇ	ဇ̄	-t	ဇ	ဇ̄	-t
ဇ	ဇ̄	-t	ဇ	ဇ̄	-h-	ဇ	ဇ̄	-n-	ဇ	ဇ̄	
ဇ			ဇ								

Inspection of the subjoined forms shows that most are simply downsized versions of their base forms. The subjoined forms of the graphemes for the three nasals /m n ŋ/ share a common feature of simplification from their base form. This makes them easier to distinguish the smaller size of the subjoined character. In handwriting and some older fonts, a simpler allograph (same shape as the subjoined form but same size as the base form) of the grapheme for /ŋ/ is even used when the grapheme for final /-ŋ/ is written in postscript position (for example Peltier 1996). Seven of the narrow subjoined forms have elongated vertical stokes rising from their right side – a consistent feature that adds to the visual distinctiveness of the narrow forms while not greatly increasing the amount of memorisation required by the learner. Eight of the graphemes – indicated by boxes – do not have a subjoined form. Of these, seven are characters not used to write Pali and the eighth is the vowel support. Also in boxes are the phonemes of those graphemes whose subjoined forms are never used in the syllable coda. These forms occur in polysyllabic morphemes written with medial orthographic clusters such that the subjoined form is the onset of a non-initial syllable.

Perhaps the most striking feature of Khuen finals that should be noted from Table 8 is the discrepancy between the number of final consonant phonemes and the number of graphemes that can be used to represent those phonemes. As with the Thai orthography, Khuen preserves historic Indic spellings even though the pronunciation has been adapted to Khuen phonology. This creates many-to-one grapheme-phoneme

mappings which greatly complicate learning to read and especially write in Khuen, as spellings have to be memorised. For example, as shown in Table 8 there are 13 graphemes that can be used to write final /t/ and 6 for final /n/. Within these mappings, however, there are huge differences in the numbers of lexemes that use a particular grapheme, the most numerous being the grapheme used to spell Tai morphemes as opposed to morphemes loaned from other languages, especially Pali and Sanskrit. Among those graphemes used for final /n/, the most commonly used is ၵ (final form ၵ) although ၶ is also very common, especially because it occurs in some frequently-used Pali prefixes such as ကၵ် /ka:n/ ‘work’. Among those graphemes used for final /t/, ၷ (final form ၷ) is by far the most common.

As noted above, there are two phonemic initial consonant clusters viz. /kw/ and /k^hw/ in spoken Khuen. These are written as one would expect, using the subjoined form of ၸ the semivowel /w/ for high and low class aspirated and unaspirated velar plosives: ကၵ် ၵ ၶ ၷ. In addition to the subjoined form of the trill ၸ, Khuen has a medial form [– for use in initial clusters with unaspirated velar, dental and labial plosives: [က [ခ [တ [ပ. Of note here is the fact that the trill is not pronounced but the plosive becomes aspirated, the tone class remaining the same as for the unaspirated plosive. Both Pittayaporn (2009:121) and Li (1977:255-6) reconstruct phonemic clusters with the trill plus velar and labial plosives in PSWT; however, Li reports that in many languages, modern reflexes are aspirated and do not involve the trill. So the written form apparently represents an older form of the spoken language, but readers learn to interpret these clusters as digraphs and pronounce them according to contemporary speech. In fact, there are signs that modern spelling practice is changing for these clusters to reflect the pronunciation rather than the etymology, i.e. these clusters are being replaced by the homorganic aspirated plosive with the same tone class (Kana Pap Aphithan 2008). Li reports that phonological /tr-/ clusters in Tai occur in loans from Khmer and Sanskrit (1977:117). These are not common in Khuen, the unaspirated high class ‘dental’ က being the only initial to form such clusters, for example, ကတ [trɛːˀ] [t^hɛːˀ] ‘bugle’ (Bhikkhu 2008) which shares a Khmer cognate ប្រែ /trae/ (Headley et al. 1997).

Another common cause of initial clusters is the use of ‘leading ha’ ၸ to change the class (to high) of the eight low class consonant graphemes that do not have corresponding high class graphemes, namely the nasals and resonants of Table 6: က ၵ ၶ ၷ ၸ ၹ ၺ ၻ. Note that the apparent duplication of forms for /j^H/, namely ၼ and ၽ is needed because ၼ is only used for four lexemes: ၼ် ‘to be situated’; ၼ်ဝ ‘do not’ (command); ၼ်ဝ် ‘type, sort’; ၼ်ဝ်ဝ် ‘to want’.

Before moving on to Khuen vowels, we note two special characters used to represent double consonants in Pali morphemes. Firstly, ၾ represents medial double /s/. Secondly ၿ represents the double Pali phoneme /-ɲɲ-/ which in Khuen is realised as [-nj-].

6.3 Khuen vowel graphemes

The Khuen orthography has two sets of vowel graphemes, termed ‘independent’ and ‘dependent’ to indicate whether or not they can be used without a base consonant (Coulmas 2003; The Unicode Consortium 2009). Independent vowels are predominantly used for vowel-initial syllables in word-initial position. Although in most instances they occur in morphemes borrowed from Pali, there are also vowel-initial native Khuen morphemes. Table 9 shows the grapheme-phoneme correspondences of the independent vowel graphemes. The superscript numbers on the phonemes denote the phonemic tone inherent in the vowel graphemes.

Table 9: Independent vowel symbols in the Khuen script

	Front		Central		Back	
	Short	Long	Short	Long	Short	Long
High	ၵ i ²	ၶ i: ²			ၷ u ²	ၸ u: ²
Mid		ၹ e: ²				ၺ o: ²
Low			ၻ a ²	ၼ a: ²		

We note in passing that the inherent tones are consistent with the interpretation as mid class consonants – which will be discussed in the section on tones below.

The vowels in Table 10 below cannot stand alone – they must always be associated with a syllable-initial consonant or cluster and possibly a syllable-final consonant. The vowel graphemes in this chart are taken from ‘Buddhism Summer Curriculum Level 1 Book 1’ (Kana Anurak Phasa 2011). It should be noted however that there are still different conventions in contemporary use, particularly in short forms of the complex vowels /-e -ε -o/ in open syllables (see for example Peltier 1996:39). The ‘-’ dash symbol in the orthographic form represents a consonant onset, and, where appropriate, a consonant coda, to show the relative placement of the vowel. In the interests of simplicity, the dash representing the coda is placed after the vowel, although as shown in Figure 1 the final consonant grapheme can be placed underneath the base consonant or the postscript vowel or above the postscript vowel position. The dashes in the phonemic representation are in the linear order of pronunciation. Recall that in open position short vowels are always followed by a glottal stop [ʔ], but since this is not represented in the orthography, it is not represented by a dash in the open short vowel columns in the table.

Table 10: Khuen Monophthong Vowel Grapheme-phoneme Correspondences

	Front		Central		Back	
	Open	Closed	Open	Closed	Open	Closed
High	ᵛ -i	ᵛ - -i-	ᵛ -u	ᵛ - -u-	ᵛ -u	ᵛ - -u-
	ᵛ̄ -i:	ᵛ̄ - -i:-	ᵛ̄ -u:	ᵛ̄ - -u:-	ᵛ̄ -u:	ᵛ̄ - -u:-
Mid	ᵛ̄-ᵛ̄ -e		ᵛ̄-ᵛ̄ -ɣ		ᵛ̄-ᵛ̄ -o	
	ᵛ̄- -e:	ᵛ̄- -e:-	ᵛ̄-ᵛ̄ -ɣ:	ᵛ̄-ᵛ̄ -ɣ:-	ᵛ̄- -o:	ᵛ̄- -o:-
Mid (*iə *uə)					ᵛ̄-ᵛ̄ -o	
		ᵛ̄- -e:-			ᵛ̄-ᵛ̄ -o:	ᵛ̄- -o:-
Low	ᵛ̄-ᵛ̄ -ε		- or -ᵛ̄ -a	ᵛ̄ - -a-	ᵛ̄-ᵛ̄ -o	ᵛ̄ - -o-
	ᵛ̄- -ε:	ᵛ̄- -ε:-	-o -a:	-o -a:-	ᵛ̄ -o:	ᵛ̄ - -o:-

Note that the grapheme ᵛ̄ /a:/ is never used with the five base consonants ᵛ ᵛ̄ ᵛ̄ ᵛ̄ ᵛ̄ – presumably to avoid confusion with graphemes such as ᵛ ᵛ̄ ᵛ̄ ᵛ̄. The allograph ᵛ̄ is used with these and only these graphemes: ᵛ̄ ᵛ̄ ᵛ̄ ᵛ̄. Each consonant has an inherent short /-a/ vowel so open syllables ending in /-a/ can be written by the consonant alone although sometimes the grapheme ᵛ̄ is used to write /-a/ explicitly.

For high vowels in Table 10, the vowel graphemes are identical for open and closed syllables. For several of the mid and low vowels, there is consistent use of the ᵛ̄ symbol to differentiate the short vowel from the long vowel in open syllables. This repeated pattern makes for a system that is easier to memorise. Various authors (see Pittayaporn 2009a:139 for a summary) have reconstructed three diphthongs ending in open central vowels for Proto-Southwestern Tai, namely /*iə, *uə, *uə/. While these have been simplified to /e:, ɣ:, o:/ in the spoken form of Khuen, in the written form there are still separate graphemes for /e:, o, o:/. The graphemes used to write words reconstructed with the proto-diphthong *uə are the same as those presented for /ɣ:/ in Table 10 above, so they are not repeated here. The extra row for mid vowels derived from diphthongs in Table 10 makes their development explicit and explains why the present overdifferentiation exists. Note that in open syllables, there is no separate form for /e:/ derived from *iə. The grapheme for medial /e:/ derived from *iə is the same as the subjoined form of /j_l/, which creates some potential for ambiguity but in most cases ambiguity is ruled out by phonotactics. The medial /o:/ grapheme is the same as the subjoined form of /w_l/, which also creates potential for ambiguity, although again in most cases ambiguity is ruled out by phonotactics.

In this paper, two consecutive vowel sounds in the rhyme are interpreted as a nuclear vowel followed by an approximant coda (limited to /j/ or /w/). While some rhymes (/iw -e:w -ε:w -a:w -uj -u:j -ɣ:j -a:j -uj/) are written according to the closed syllable patterns laid out in Table 10 above, others (/o:j -o:j -aj -aw -e:w/) are written with ligatures or complex forms as shown in Table 11.

proto-initial consonant class and proto-tone category) and Table 13 (tone distribution by initial consonant class and orthographic tone marker).

Table 13: Summary of tone rules in the Khuen orthography

		Unchecked syllable			Checked syllable		Unchecked syllable			
		–	ˊ	ˋ	–	–	ˆ	ˊ	ˋ	
Six-tone System	Initial Consonant Class	High	1	3	5	2	3			
		Mid	2	3	5	2	3	2 [45]	1	6
		Low	2	4	6	6	4			
Five-tone System	Initial Consonant Class	High	1	4	5	2	4			
		Mid	2	4	5	2	4	2 [45]	1	6
		Low	2	4	6	6	4			

As can be seen from Table 13 the tone marks ˆ; ˊ; and ˋ can only be used with mid class consonants. We note their resemblance in shape to the vowel ˆ and the Khuen numerals ၂ <2> and ၃ <3>. This relationship is confirmed by the Khuen names of these tone marks which make direct reference to the corresponding vowel/numeral. The origin of these tone markers is not clear. They are used mainly for onomatopoeic words and do not feature in either of the Khuen dictionaries published to date, or the lexica compiled by Peltier from Khuen traditional stories (1999; 2005; 2006). The fact that they are not found in Lan Na or (traditional) Tai Lue scripts also suggests that they are recent introductions to the Khuen orthography. The reason for their introduction can be guessed by looking at the way tones are taught in complete sets in Khuen primers. The example in Table 14 shows orthographic and phonemic transcriptions based on the six-tone system.

Table 14: Tone sets in Khuen Reader (based on Peltier 1996:41)

High-Low Set	ကန	ကဝ	ကံဝ	ကံ့ဝ	ကန	ကဝ	ကံဝ	ကံ့ဝ
	ka ² [ka ⁴⁵]	ka: ¹	ka: ³	ka: ⁵	ka ⁶ [ka ⁵³]	ka: ²	ka: ⁴	ka: ⁶
Mid Set	ဒန	ဒဝ	ဒံဝ	ဒံ့ဝ		ဒံ့ဝ	ဒံ့ဝ	ဒံ့ဝ
	da ² [a ⁴⁵]	da: ²	da: ³	da: ⁵		da: ² [da: ⁴⁵]	da: ¹	da: ⁶

The High-Low set shows how all 6 phonemic tones are written on unchecked (non-shaded) syllables by the combination of high and low class unaspirated velar plosives (/k_H/ and /k_L/) with three tone markings (zero, ˊ, ˋ). The mid consonants, on the other hand, have no natural corresponding phonemes, so the three extra tone marks allow other tones to be written for the mid consonants. Looking along the final row for the mid class consonant ဒ, we see that each phonemic tone apart from tone 4 is represented. Phonemic tone 2 is represented in the first and second columns for checked and unchecked syllables respectively. The 6th column shows that the third tone mark represents the high allotone of tone 2 found in some varieties of Khuen; that is, it marks phonetic tone rather than relying on the environment to guide the reader on how to pronounce the phonemic tone marking. The lack of a way to write tone 4 suggests that these tone marks were introduced after the merger of tone 3 with tone 4 in the five-tone system described in Section 3.4.

The two-stage development of tone marking in Khuen is similar to that of Standard Thai, whose early form was ‘probably the first orthography of a fully tonal language to mark tones in a practical writing system’ (Diller 1996). Diller describes the history of tone marking in Thai, from King Ramkhamhaeng’s 1283 innovative use of <◌◌◌> to mark Gedney tone categories B and C to the 18th century development of the four marker system <◌◌◌◌> in use today. Apart from any influence Thai tone marking might have had on the first stage of tone marking in Khuen, i.e., the use of ˊ and ˋ to mark tones in the B and C columns, there is also evidence of more recent influence. Certain schools of Khuen literacy currently use the Thai tone markers <◌◌◌> instead of <◌◌◌> respectively (Jutidhammo n.d.). As shown in Table 13 ˋ or ◌◌◌ mark Tone 1 (falling rising /215/) in Khuen whereas ◌◌◌ or ◌◌◌ mark Tone 6 (high falling /51/) in Khuen. The phonetic characteristics of the respective tones in Standard Thai are falling rising /215/ and high rising falling /451/

(numbers inferred from tone plot in Abramson 1962:127.) In other words the tones associated with these markers are phonetically almost identical in the two languages.

We note in passing that the second Khuen tone marker ˊ (known as /maj⁶ sat⁶/) is the same grapheme that is used for the /-a-/ vowel in closed syllables. Both uses of the grapheme are relatively high frequency. The 1:2 grapheme-phoneme correspondence creates ambiguity when first scanned, but a unique interpretation is almost always confirmed by the presence or absence of some other vowel grapheme within the syllable. One combination that remains ambiguous is the use of ˊ with an initial cluster having /w/ as the second element, because medial ɔ can represent /w/ or the vowel /o:/. For example, ၂၈ ‘smoke’ could be read as either /kwan²/ or /ko:n⁵/. Since the Unicode encoding provides separate code points for these two uses of ˊ, computer fonts can easily solve this problem by making the glyphs visually distinct, as in the font used here. When they co-occur in the same morpheme, the tone marker is positioned above the vowel, illustrated here with /k_H/ ၂၈. Note that since the combination is unambiguous, no differentiation in shape is needed.

6.5 Khuen numerals and punctuation

All three sets of numerals laid out in Table 15 are in use by Khuen, depending on the context and ease of production. The Khuen ‘Tham’ numerals are used in religious texts, while the ‘Hora’ numerals (essentially the same as Burmese) are used in secular writings. Arabic numerals are used when the audience might be multi-ethnic. All three sets follow the decimal system.

Table 15: Numeral sets used in Khuen

Arabic	1	2	3	4	5	6	7	8	9	0
Khuen Tham	၈	၇	၆	၅	၄	၃	၂	၁	၀	၀
Khuen Hora	၁	၂	၃	၄	၅	၆	၇	၈	၉	၀

Khuen text is not spaced between words but between clauses, similar to the Thai convention. New paragraphs are started on a new line and the first clause is indented. Space is also used to separate items in a list. Apart from these conventions, punctuation practices vary. Most texts use little additional punctuation; some texts use Burmese-style markers corresponding to English comma and full stop; some texts use (obsolete) Thai markers in a similar way.

7. Assessing the user-friendliness of the Khuen orthography

The large grapheme inventory including base and subjoined forms of consonants means that the Khuen orthography is very capable of representing not only the spoken language but also the ancient Pali religious language. Furthermore, it is flexible enough to be able to accommodate modern loanwords from Burmese, Shan and English. Typologically, Khuen is isolating and has a preponderance for monosyllabic morphemes which the orthography is structurally well suited to writing. So Khuen scores highly for its capacity to represent the languages used within the Khuen-speaking community. As far as the user-friendliness of the orthography goes, the ease of use for both reading and writing must be considered, since a particular feature of the orthography might have a different impact on reading than it has on writing.

A major feature of the Khuen orthography is the large grapheme inventory which entails complex many-to-one correspondences with the phonemes, particularly with the consonants, but also with mid vowels. These many-to-one correspondences create a lack of transparency in the orthography, and more memorisation is needed when initially learning to read. Non-transparent orthographies present more difficulty in decoding words, and as a consequence the process of learning to read is slower (Goswami 2006). Furthermore, neurocognitive research on Devanagari script shows that not only lack of phonological transparency in the orthography, but also the complexity of the spatial arrangement of graphemes influences the neural and cognitive processing of written words (Rao & Singh 2015).

Perhaps the greatest negative impact of many-to-one correspondences is felt by learners when writing because knowledge of the pronunciation of a morpheme is often not sufficient to deduce its correct spelling. Thus extra memorization is required to learn spellings of lexemes involving graphemes that do not enjoy a one-to-one correspondence with their phonemes.

As well as the costs in terms of extra memorisation due to the many-to-one grapheme-phoneme correspondences, there are some benefits. One benefit to the reader of using subjoined forms is that they often help to identify syllable boundaries. Ligatures for common rhymes are also readily identifiable and act as clear markers of the end of a syllable. Even though most consonant graphemes have both regular and subjoined forms, most of the subjoined forms are simply downsized versions of the regular form, so the learner only needs to memorise a few extra character shapes. The 2:1 correspondence of initial plosive graphemes to phonemes produces a 1:1 mapping of tone for varieties that still have six phonemic tones, whereas for 5-tone varieties there is slight overdifferentiation in tone notation. The 2:1 mapping of certain vowel graphemes to phonemes should not obscure a 1:1 mapping for vowel qualities in closed and open syllables respectively. Furthermore, the form of vowel helps to identify whether the following consonant is in the coda of the preceding syllable or the onset of the next.

Another feature of the orthography is the 1:2 mapping caused by using the same grapheme for the second tone marker and the /-a-/ vowel in closed syllables. This does create ambiguities for readers when initially sighted, but when the remainder of the syllable is considered, a unique interpretation almost always results. Notwithstanding, this temporary ambiguity probably slows the reading process while the interpretation is carried out. This problem is eased by computer fonts that make the two uses visually distinct, in effect creating two separate graphemes, as provided for by the Unicode encoding.

Various reading studies in recent years have explored the effect of spacing on how readers process text. Like many neighbouring orthographies (Burmese, Thai, Lao), Khuen does not use inter-word spacing. One benefit of this is that it avoids having to identify word boundaries. Studies of Thai reading (see for example Kasisopa 2011) showed no significant difference in reading speed for adults between normal unspaced Thai text and text with added word spaces. Children, on the other hand, do read word-spaced text significantly faster, as would be expected given that word-spaced text is the norm in reading instruction until the end of first grade. Studies of eye movement in Thai reading show that readers' first fixation (the place where the eyes first focus when looking at a string of text) tends to be near the centre of words, regardless of whether the text has word spacing or not. Kasisopa et al. (2013) showed that key determinants in this ability to identify word boundaries even without spaces was the presence of boundary characters, that is characters with higher frequency of occurring at the start or end of words. With the prolific use of subjoined forms, Khuen has more clues to the position in the syllable of an individual graphic element than there are in Thai text, so it might be hypothesised that proficient Khuen readers would be able to read unspaced text with roughly the same speed as spaced text. Furthermore, boundary characters can be expected to play an important role in assisting Khuen readers to identify words in unspaced text.

8. Conclusions

The Khuen orthography has a rich inventory of graphemes that provide good coverage of the sounds in Khuen as well as allowing for unambiguous representation of Pali religious texts. Accommodation of loans with new clusters illustrates the capacity of the Khuen orthography to create new combinations of graphemes for representing clusters not native to Khuen. This suggests that the Khuen grapheme inventory is rich enough to represent other unrelated languages, such as Ruching Palaung, which currently uses an orthography based on that of Khuen. An assessment of the use of Khuen orthography for writing Ruching Palaung is a topic for further research.

Sound changes over the time since the Khuen orthography was adopted have decreased the phonological transparency of the present orthography. Furthermore, the use of spelling conventions that preserve the etymology of ancient Indic loans creates many-to-one grapheme-phoneme correspondences, which greatly increase the memorisation required (and consequently the number of spelling errors) on the part of learners. After memorising the distinct spellings for non-homographic homophones, however, the learner has access to unambiguous correspondences between morphemes and graphic forms, which may help comprehension. The development of a standard dictionary (preferably available online using a Unicode font) would greatly facilitate standardisation of spellings in the language as well as orthographic conventions such as graphemes of short vowels and tone markers.

Even though there are no inter-word spaces in Khuen text, the structure of the Khuen orthographic syllable with the use of subjoined forms plus several special ligatures for rhymes give extra visual cues for syllable boundaries. Moreover, the lack of inter-word spaces obviates the need for writers to identify word

boundaries, although readers do need to identify word boundaries in order to comprehend the text. Notwithstanding, there remains much scope for further study of how readers process Khuen text.

Perhaps the final word regarding the Khuen orthography should point to its high level of acquisition, especially amongst men but also amongst women, which is remarkable for a minority script not taught in government schools. Further study of the place of Khuen writing in the cultural and religious life of the community might well provide insights to why Khuen enjoys such widespread usage while other closely related communities that use the Tai Tham script – such as Tai Lue in Mong Yawng – currently have much lower levels of acquisition (Na Tiqwah 2016).

Eat.04 ခေဝ် နွဲ သဉ္စု ရီၣ် တီၣ်ညးန ဝဲၣ် ဣၣ်ကတဲာ် ဣၣ် တဲာ်ခေဝ် ရီၣ် ခေဝ်
kʰaw¹ tan² sa:m¹ kʰɔn² pi:⁴nɔ:ŋ⁶ daj⁵ co:n²kan¹ ca:j⁴ pɔ:⁴mɛ:⁴ kʰɔn² tʰaw⁵
 they all three CLF.person sibling did together help parents person old
 ခုၣ် ဝဲၣ် ဝဲၣ် သွၣ် လွံ ကျွဲ နွဲ ဝဲၣ် ချဲ ဝဲၣ် ကတဲာ် ဝှဲၣ်
kʰut² raj⁴ tʰaj¹ so:n¹ le:w⁶ kep² pʰak² paj² kʰa:j¹ naj² ka:t³ we:ŋ²
 dig field plough garden then collect vegetable go sell in market city
 They all three siblings together help their elderly parents by digging the field and ploughing the garden, collecting the vegetables and going to sell them in the city market.

Eat.05 လှဲၣ် ဣၣ် ခေဝ်သဉ္စု ဘွဲၣ် ကျွဲ ဣၣ်က တီၣ် ဝဲၣ် သ့ၣ် နေၣ် တဲာ်ခေဝ်
le:w⁶kɔ:⁶ su:⁶ kʰaw⁵sa:n¹ pʰik⁶ kɔ:¹ pɔ:k⁴ kʰu:n² ma:² su:³ rɔ:n² pɔ:⁴mɛ:⁴
 and.then buy uncooked.rice chili salt return return come to house parents
 ဝဲၣ်ခေဝ် ခေဝ် နွဲၣ် ဣၣ် ဣၣ်က ဝဲၣ် ဝဲၣ် ခုၣ်
hɛ:ŋ³ kʰaw¹ day³ ni:⁶ tuk⁶ wan² baw³ kʰa:t³
 place they like this every day not be.missing
 Then they buy rice, chillies and salt and bring them back to their parents' house like this every day without fail.

Eat.06 မိၣ် ဝဲၣ် တီၣ်ခေဝ် ခေဝ် သဉ္စု ရီၣ် တီၣ်ညးန ဝဲၣ် ဣၣ်ကတဲာ် ဝဲၣ်
mi:² wan² nuŋ³ kʰaw¹ sa:m¹ kʰɔn² pi:³nɔ:ŋ⁶ daj⁵ co:n²kan¹ paj²
 have day one they three person sibling did together go
 ဣၣ်က နွဲၣ် ဝဲၣ် ချဲ ဝဲၣ် ကတဲာ် လှဲၣ်
aw² pʰak² paj² kʰa:j¹ naj² ka:t³ le:w⁶
 take vegetable go sell in market already
 One day the three siblings went together to sell the vegetables in the market.

Eat.07 ဝဲၣ် ဣၣ် ရီၣ် ဘွဲၣ် ခေဝ် ချဲၣ် ဝဲၣ်ခေဝ် ဣၣ် ဝဲၣ် ကတဲာ်
daj⁵ han¹ kʰɔn² u:n³ kʰaw¹ kʰa:j¹ kaj³mɛ:⁴ ti:⁴ naj² ka:t³
 did see person other they sell hen at in market
 They saw another person selling hens at the market.

Eat.08 ခေဝ် သဉ္စု ရီၣ် တီၣ်ညးန ဝဲၣ် ဣၣ်ကတဲာ် ဝဲၣ် ဣၣ် ဣၣ်က ဝဲၣ် ဝဲၣ် ဝဲၣ် တီၣ်ခေဝ် တီၣ်ခေဝ်
kʰaw¹ sa:m¹ kʰɔn² pi:³nɔ:ŋ⁶ kɔ:⁶ co:n²kan¹ paj² su:⁶ aw² kaj³mɛ:⁴ to:¹ nuŋ³
 they three CLF.person sibling then together go buy take hen CLF.animal one
 လဲၣ် ဝဲၣ်ကတဲာ် တီၣ်ခေဝ် တီၣ်ခေဝ် ဝဲၣ် ဝဲၣ် ဝဲၣ် ဝဲၣ် ဝဲၣ်
le:⁶ kaj³pu:⁶ to:¹ nuŋ³ ca:k³ pʰu:³kʰa:j¹ kaj³ nan⁶ se:⁴le:w⁶
 and cockerel CLF.animal one from seller chicken that already
 ဝဲၣ် ဣၣ်ကတဲာ် ဣၣ်က ဣၣ်က ဝဲၣ် ဝဲၣ် ဝဲၣ် ဝဲၣ် ဝဲၣ်
kɔ:⁶ co:n²kan¹ hɔ:p³ aw² kaj³ nan⁶ pɔ:k⁴ ma:² su:³ rɔ:n² le:w⁶
 then together carry take chicken that return come to house already
 The three siblings then together went and bought one hen and one cockerel from that chicken seller then together carried those chickens back home.

Eat.09 ဣၤကၢ် လျၢ်နၢ် ပျဲၤ ငိၤန့ၣ် ဝဲၤ ခၢ်
co:n²kan¹ le:ŋ⁶du:² pen¹ ni⁶ran² baw³ kʰa:t³
 together raise.animal be perpetual not be.missing
 Together they took care of the chickens continually without fail.

Eat.10 ဣၤကၢ် ဝဲၤ ငိၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
tʰuŋ¹mɹ:⁴ kaj³ nan⁶ daj⁵ tɛ:k³pʰɛ:³ la:j¹ kʰuŋ⁵ ma:² lɛ:w⁶ daŋ³an⁵
 when chicken that did multiply many rise come already thus
 ခေၣ် သၣ် ဣၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
kʰaw¹ sa:m¹ kʰɔn² ko:⁶ aw² kaj³ taŋ² la:j¹ nan⁶ paj² kʰa:j¹ ti:⁴ naj² ka:t³
 they three CLF.person then take chicken all many that go sell at in market
 When those chickens had multiplied to be many, the three then took all those chickens to sell in the market.

Eat.11 ဣၤကၢ် ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
mɹ:⁴ kʰa:j¹ kaj³ daj⁵ ŋɔn² ma:²
 when sell chicken get money come
 When they sold the chickens they got money.

Eat.12 ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
ko:⁶ cu:¹sam⁶ co:n²kan¹ su:⁶ aw² mu:¹nɔ:j⁶ la:j¹la:j¹ to:¹
 then again together buy take piglet many CLF:animal
 ဝဲၤ လျၢ် ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
ma:² le:ŋ⁶ tɛ:n² kaj³ nan⁶ paj² tʰɛ:ŋ⁵
 come raise.animal instead.of chicken that go again
 They then again together bought many piglets to raise instead of the chickens.

Eat.13 ဣၤကၢ် ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
tʰu:ŋ¹mɹ:⁴ kʰaw¹ daj⁵ le:ŋ⁶ mu:¹nɔ:j⁶ taŋ²la:j¹ nan⁶ na:n² daj⁵ pi:¹ pa:j¹
 when they did raise.animal piglet all that for.a.long.time did year over
 ဝဲၤ လျၢ် ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
ma:² lɛ:w⁶ daŋ³an⁵ mu:¹nɔ:j⁶ la:j¹ cɹ:² nan⁶ ko:⁶ jaj³ ma:² lɛ:w⁶
 come already thus piglet many those then big come already
 When they had raised those piglets for over a year, the piglets grew big.

Eat.14 ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ ဝဲၤ
kʰaw¹ ciŋ³ co:n²kan¹ kʰap² aw² mu:¹ taŋ²la:j¹ nan⁶ paj² kʰa:j¹ ti:⁴ naj² ka:t³
 they so together drive take pig all that go sell at in market
 So together they drove all those pigs to sell them at the market.

References

- Abramson, Arthur S. 1962. *The vowels and tones of standard Thai: acoustical measurements and experiments*. Bloomington, IN: Indiana University Research Center in Anthropology, Folklore, and Linguistics.
- Ananda Travel Yangan. 2009. List of the 135 tribes and ethnic group[s] in Myanmar. Accessed 07-Nov-2016 from http://www.ananda-travel.com/UK/myanmar_ethnic_group_uk.htm
- Bright, William. 1999. A matter of typology: alphasyllabaries and abugidas. *Written Language and Literacy* 2:45-56.
- Brown, J. Marvin. 1985. *From ancient Thai to modern dialects*. Bangkok, Thailand: White Lotus.
- Cook, Vivian and Benedetta Bassetti. 2005. An introduction to researching second language writing systems. *Second language writing systems*, ed. by Vivian Cook and Benedetta Bassetti, 1-67. Bristol: Multilingual Matters.
- Coulmas, Florian. 2003. *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge University Press.
- Daniels, Peter T., and William Bright (eds.) 1996. *The world's writing systems*. New York: Oxford University Press.
- Diller, Anthony V. N. 1994. Tai languages: varieties and subgroup terms. In Wijehewardene, Gehan (ed.) *Thai-Yunnan Project Newsletter No. 25*. 8-18. Canberra: Australian National University.
- Diller, Anthony V. N. 1996. Thai orthography and the history of marking tone. *Oriens Extremus* 39.2:228-254.
- Duroiselle, Charles. 2007. *A Practical Grammar of the Pāli Language*. 4th ed. Accessed on 08-Nov-2011 from <http://www.pratyeka.org/duroiselle/>
- Duanmu, San. 2010. *Syllable Structure: The Limits of Variation*. Oxford: Oxford University Press.
- Egerod, Søren. 1957. Essentials of Shan phonology and script. *Academia Sinica/Bulletin of the Institute of History and Philology* 29.121-9.
- Egerod, Søren. 1959. Essentials of Khün phonology and script. *Acta Orientalia* 24.3-4:124-146.
- Gedney, William J. 1973. A checklist for determining tone in Tai dialects. In *Studies in Linguistics in Honour of George L. Trager*. The Hague: Mouton.
- Gedney, William J. 1989. Speculations on early Tai tones. *Selected Papers on Comparative Tai Studies*, ed. by Robert J. Bickner, John Hartmann, Thomas John Hudak, and Patcharin Peyasantiwong, 207-228. No. 29. Ann Arbor, Michigan: Center for South and Southeast Asian studies, the University of Michigan. <http://purl.org/sealang/gedney1989speculations.pdf>.
- Goswami, Usha. 2006. Orthography, Phonology, and Reading Development: A Cross-Linguistic Perspective. *Handbook of orthography and literacy*, ed. by R. Malatesha Joshi & P. G. Aaron, 463-480. New Jersey: Lawrence Erlbaum Associates.
- Haas, Mary R. 1956. *The Thai System of Writing*. Ithaca, N.Y: Spoken Language Services Inc., U.S.
- Hartmann, John F. 1986. The spread of South Indic scripts in Southeast Asia, *Crossroads* 3.1:6-20. Dekalb, Ill: Northern Illinois University, Center for Southeast Asian Studies.
- Hartmann, John F. 2008. The Lue Language. *The Tai-Kadai languages* ed. by Anthony V. N. Diller, Jerold A. Edmondson and Yongxian Luo, 254-297. London: Routledge.
- Headley, Robert Kirk, Rath Chim, and Ok Soeum. 1997. *Modern Cambodian-English Dictionary*. Dunwoody Press.
- Hudak, Thomas John (ed.) 1994. *William J. Gedney's Southwestern Tai Dialects: Glossaries, Texts and Translations*. Michigan Papers on South and Southeast Asia, 42. Ann Arbor: Center for South and Southeast Asian Studies, The University of Michigan.
- Iwasaki, Shoichi and Preeya Ingkaphirom. 2005. *A reference grammar of Thai*. Cambridge University Press.
- Jagacinski, Ngampit. 1986. Tai Lue scripts: the old and the new. *Crossroads* 3.1:80-96.
- Jonsson, Nanna L. 1991. *Proto-Southwestern Tai*. Albany, NY: State University of New York at Albany. (Ph.D dissertation.)

- Rao, Chaitra and Nandini C. Singh. 2015. Visuospatial complexity modulates reading in the brain. *Brain & Language* 141:50-61.
- Rasi Petsuk. 1978. *General characteristics of the Khün language*. Bangkok: Mahidol University. (MA Thesis.)
- Rhee, Seok-Chae. 2003. Onset-Coda Asymmetries in Mishmi and Other Southeast Asian Languages. *Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society*, ed. by Karen L. Adams, Thomas John Hudak, and F. K. Lehman, 137–152. Tempe, Arizona: Arizona State University, Program for Southeast Asian Studies. <http://purl.org/sealang/rhee2003onset.pdf>.
- Sai Kam Mong. 2004. *The history and development of the Shan scripts*. Chiang Mai: Silkworm Books.
- Salomon, Richard G. 1996. Brahmi and Khartoshthi. *The world's writing systems* ed. by Peter T. Daniels and William Bright, 1-20. Oxford: Oxford University Press.
- Sarawit, Mary E. 1973. The Proto-Tai vowel system. Ann Arbor: University of Michigan. (Ph.D. dissertation.)
- Schlesinger, Joachim. 2003. *Ethnic groups of Laos*. Vol 3: Profile of Austro-Thai-Speaking Peoples. Bangkok: White Lotus.
- Smalley, William A. 1997. Early Protestant Missionaries and the Development of Thailand's Hierarchy of Multilingualism. *Southeast Asian Linguistics Studies in honor of Vichin Panupong*, ed. by Arthur S. Abramson, 237–252. Chulalongkorn University Press. <http://purl.org/sealang/smalley1997early.pdf>.
- Sun Watthanatham Cangwat Chiang Mai lae Sun Silpwatthanatham Sathaban Rajabhat Chiang Mai (ศูนย์วัฒนธรรมจังหวัดเชียงใหม่และศูนย์ศิลปวัฒนธรรมสถาบันราชภัฏเชียงใหม่) [Chiang Mai Province Culture Department and Chiang Mai Rajabhat Institute Art and Culture Department]. 1996. *Photcananukrom phasathin phak nuea* (พจนานุกรมภาษาถิ่นภาคเหนือ) [The Northern Thai dictionary]. Chiang Mai, Thailand.
- The Unicode Consortium. 2009. *The Unicode Standard, Version 5.2*. Mountain View, CA: The Unicode Consortium. <http://www.Unicode.org/versions/Unicode5.2.0/>
- Trager, Ed. 2014. Hariphunchai Tai Tham Font Project. ms. Accessed on 08-Nov-2016 from https://www.academia.edu/6119912/Hariphunchai_Tai_Tham_Font_Project.
- Udom Rungruang Sri (อุดม รุ่งเรืองศรี) (compiler). 1990. *Photcananukrom Lanna-Thai: Chabap Mae Fah Luang* (พจนานุกรมล้านนา-ไทย ฉบับแม่ฟ้าหลวง) [Lanna-Thai Dictionary: Mae Fah Luang Edition]. Thailand: Mae Fah Luang Foundation and Siam Commercial Bank.

Reviewed: Received 10 December 2016, revised text accepted 17 March 2017, published 24 March 2017

Editors: Editor-In-Chief Dr Mark Alves | Managing Eds. Dr Paul Sidwell, Dr Nathan Hill, Dr Sigrid Lew