

ESSENTIALS OF SHAN PHONOLOGY AND SCRIPT

SØREN EGEROD

University of Copenhagen

Shan is a Tai language spoken by the majority of people in the Shan States of the Union of Burma where large minorities speak Burmese, Karen and other languages.¹ It is also spoken in the Shan States of Yunnan in competition with Chinese and, sporadically, in northwest Thailand, especially in the Mae Hong Son area in competition with Siamese. There are regional variations not beyond mutual intelligibility. The following notes are concerned with educated speech of the Shan States of Burma as recorded in Taunggyi, Southern Shan States, with the purpose of making available to linguists, especially to those carrying out comparative Tai studies, material contained in such works as J.N. Cushing's *A Shan and English Dictionary*, Rangoon 1881 and 1914, and H.W. Mix' *English and Shan Dictionary*, Rangoon 1920. These works remain of little value to the modern scholar without a phonological reinterpretation.

Shan phonology is of the same general type as other Tai languages such as Siamese which have been described², but is simpler, has fewer syllabic types than Siamese. The syllable consists of a tone and a syllabic (vowel or syllabic [m]) which is always preceded by an initial consonant or consonant cluster and which may (if it is a vowel) be followed by a final consonant.

There are the following tones:

1. /˥/ [ʌ] rising
2. /˨/ [ɿ] low level

-
1. Material for this article was collected during the tenure of a grant from the John Simon Guggenheim Memorial Foundation (1956-57) for which I wish to express much gratitude. I also wish to thank my informants, U San Htwe (formerly of Hsipaw), U Sawta (formerly of Mong Tung), U Pyinnya (formerly of Ke-hsi Mansam), and U Kokhit (formerly of Nam Hu), all residing at present in Taunggyi, Southern Shan States. I am further indebted to Mgr. Lucien Lacoste, Bishop of Tali, Yunnan, residing in Chiangmai, Thailand, and Princess Sukanta Na Chiangmai for their generous help as my first teachers of Shan.
 2. For Siamese see Mary R. Haas, *The Thai System of Writing*, 1956, and bibliography in second American reprint of George Bradley McFarland, *Thai-English Dictionary*, Stanford University Press, 1956. For the rest see for instance bibliography in Fang-Kuei Li, *Consonant Clusters in Tai*, *Language* 30/3 1954.

3. /zero/ [ɹ, ɹ] middle level or slightly falling
4. /ˈ/ [ɹ, ɹ] high level or slightly rising
5. /ˋ/ [ɹ] falling

Tones 2, 4 and 5 are often accompanied by glottal friction and terminate in glottal stop in isolation.

In addition there is a neutral tone /•/ [ɹ] which occurs only with open syllables. These syllables are in more precise speech pronounced in the fifth tone or in rarer cases in the third tone, for instance, /kâti/ or /ka•ti/ 'promise'; /tâtijâ/ or /ta•ti•jâ/ 'third'; /kaanǎ/ or /ka•nǎ/ 'on'.

Words terminating phrases or sentences may take the inflexions [ɹ] or [ɹ] which should be considered manifestations of sentence intonation, for instance, [hyɹ] (also /hǎ/) 'final particle for question'; [kɔɹ] (also /kɔ/) 'also, then'.

The following chart shows the syllabic vowels:

/i/ [i]	/y/ [i]	/u/ [u]
/e/ [ə, e]	/ə/ [ə, ə]	/o/ [o, o]
/ɛ/ [æ, ɛ]	/a/ [ä, ə]	/ɔ/ [ɔ, ɔ]
	/aa/ [ä:]	

The pronunciation of vowels for which two forms are given varies even with the same individual. The second pronunciation given is more common in open syllable, but the first variety is also heard. The diphthongized forms of back vowels are less generally used than those of front and central vowels, they are most common before /j/ where pronunciations range from [oj] and [oɛj] to [oe] for /oj/ and from [ɔj] and [ɔæj] to [ɔæ] for /ɔj/; some speakers have /waj/ for /ɔj/.

[ə] alternates with [ä] for /a/ in neutral tone. In closed syllable the quality of /a/ and of the latter element of the diphthongized forms of /ɛ/ and /ɔ/ are influenced by adjoining consonants (front quality before /n/ /d/ and /j/, central quality before /m/ and /b/, rounded central quality before /w/, back quality before /ŋ/ /g/ and /ɣ/; after /w/, /a/ is rounded; some speakers substitute /wɔj/ for /waj/); whereas /aa/ is not thus influenced. The distinction of quality between /a/ and /aa/ is more audible in rapid speech than the distinction of quantity which is characteristic of careful speech and of the words in isolation.

Vowel length is not distinctive except in the case of /a/ and /aa/. In a feature analysis it would be equally possible to count the distinction of quality,

mentioned above, as distinctive and to discount the distinction of quantity. In isolation the remaining vowels tend to be rather long in the first, second and fourth tones, somewhat shorter in the third tone, quite short in the fifth tone. Before final /b/, /d/ and /g/ they tend to be quite short in all tones (the first tone does not normally occur in such syllables); in these same surroundings the vowel /a/ mostly occurs in fourth and fifth tones, /aa/ in second and third tones. /aa/ does not occur in open syllable in the fifth and neutral tones, /a/ in open syllable only in the fifth and neutral tones.

Syllabic /m/ occurs only in the one syllable /ʔm/, an alternate form of /ʔəm/.

The above phonemic solution for vowels which will be utilized in the following allows most easily of comparison with Siamese as transcribed by Mary R. Haas (see footnote 2 above). An alternate solution is suggested in passing:

/i/ [i]	/y/ [i]	/u/ [u]
	/a/ [ä̃, ə]	
/ii/ [ẽ, e]	/yy/ [é̃, é]	/uu/ [õ, o]
/ia/ [æ̃, ɛ̃]	/aa/ [ä:]	/ua/ [ω̃, ɔ̃]

There are the following initial consonants:

p	ph	m	w
t	th	n	l
(t)s	(t)sh	h	(h)r
(c)	(ch)	(ñ)	j
k	kh	ŋ	ʔ

[p], [ph], [m], [w], [(t)s] and [(t)sh] are often palatalized before /i/, /e/, /ɛ/ (before /i/ according to the second phonemic solution suggested above for vowels) and /j/. [c], [ch] and [ñ] alternate with (often palatalized) [k], [kh] and [ŋ] respectively before /i/, /e/, /ɛ/; and [c], [ch] and [ñ] are alternate pronunciations of [kj], [khj] and [ŋj]. [ts] alternates with [s], [tsh] with [sh], [hr] with [r]. As phonemic forms we choose /ki/, /ke/, /kɛ/, /kj/, /khi/, /khe/, /khɛ/, /khj/, /ɲi/, /ɲe/, /ɲɛ/, /ɲj/, /s/, /sh/, and /r/. Some speakers substitute /l/ for /r/.

There are three medial consonants /-j-/, /-w-/, and /-r-/ which enter into the following clusters with initial consonants represented in our material:

	-j-	-w-	-r-
p-	pj-	pw-	pr-
ph-	phj-		phr-
m-	mj-		mr-
t-		tw-	tr-
th-		thw-	
l-		lw-	
s-			sr-
sh-		shw-	shr-
k-	kj-	kw-	kr-
kh-	khj-	khw-	kh-
ŋ-	ŋj-		

In dissyllabic words in which the first syllable ends in /a/ and the second syllable begins with /l/, the /a/ is sometimes dropped so that clusters containing a medial /l/ result: for instance, /pa·làŋ/ or /plàŋ/ 'throne'. Conversely, /a/ is sometimes inserted between the initial consonant and /-r-/ resulting in a dissyllabic word: for instance, /tráa/ or /ta·ráa/ 'scripture'. The cluster /jw/ occurs in Burmese loan words: for instance, /kijwâ/ 'to walk along, proceed'.

There are the following final consonants:

m	b	w
n	d	j
ŋ	g	ɣ [i]

/w/ occurs only after front vowels, /a/, and /aa/; /j/ only after back vowels, /a/, and /aa/ (and in the final particle /ʔəjv/; /ɣ/ only after /a/.

It would be natural to write initial [p], [t], [k] as /b/, /d/, /g/ and consider these as phonemically identical with final [b], [d], [g]; and to write initial [ph], [th], [kh] as /p/, /t/, /k/. To facilitate the comparison with Siamese, we refrain from this solution writing /p/, /t/, /k/ and /ph/, /th/, /kh/ initially, /b/, /d/, /g/ finally. Because of its very limited distribution

we suggest in practice to write /ɿ/ [i̯] as /y/ with the same symbol as the vowel [i].

The [ʔ] which occurs after syllables in the second, third and fifth tones is not considered phonemic.

The traditional Shan system of writing is highly imperfect in that it represents many fewer syllabic types than are actually present in the spoken language, fewer even than could be expressed by means of the symbols available. The initial and final consonants are unambiguously given, but tones are with few exceptions never written and in many cases several vowels are written alike. Furthermore there are many variations depending on individual authors and scribes as well as on the period from which the manuscripts and books date. The following data are mainly concerned with the standards used in the dictionaries of Cushing and Mix as well as modern educational publications.

Below are the symbols for initial consonants in their usual alphabetic order:

/k/ ɔ, /kh/ ɔ, /ŋ/ c, /s/ ɔ, /sh/ ɔ, /ɲj/ ɔ,
 /t/ ɔ, /th/ ɔ, /n/ ɔ, /p/ ɔ, /ph/ ɔ, /m/ ɔ,
 /j/ ɔ, /r/ ɔ, /l/ ɔ, /w/ ɔ, /h/ ɔ, /ʔ/ ɔ.

In open syllables the vowels are written in the following way:

/i/ as ɔ, except in the fifth tone where it is ɔ
 /e/ and /ɛ/ as ɔ
 /y/ and /ə/ as ɔ
 /a/ as zero (fifth tone only)
 /aa/ as ɔ in the first, second and third tones, as ɔ in the fourth tone
 /u/ as ɔ, except in the fifth tone where it is ɔ
 /o/ as ɔ
 /ɔ/ as ɔ, except in the fifth tone where it is ɔ

The remaining finals (in the sinological sense) are written as follows:

/im/ /em/ and /ɛm/ as ɔ, less commonly as ɔ
 /ib/ /eb/ and /ɛb/ as ɔ
 /ym/ and /əm/ as ɔ

/yb/ and /əb/	as	ᵛᵛ
/am/ and /aam/	as	·
/ab/ and /aab/	as	ᵛ
/um/ and /om/	as	ᵛ̇, less commonly as ᵛ̇
/ub/ and /ob/	as	ᵛ̇
/om/	as	ᵛ̇, less commonly as ᵛ̇
/ob/	as	ᵛ̇
/in/ and /en/	as	ᵛ̇ᵛ̇, less commonly as ᵛ̇ᵛ̇
/en/	as	ᵛ̇ᵛ̇, less commonly as ᵛ̇ᵛ̇
/id/ /ed/ and /əd/	as	ᵛ̇ᵛ̇
/yn/ and /ən/	as	ᵛ̇ᵛ̇
/yd/ and /əd/	as	ᵛ̇ᵛ̇
/an/ and /aan/	as	ᵛ̇
/ad/ and /aad/	as	ᵛ̇
/un/ and /on/	as	ᵛ̇ᵛ̇
/ud/ and /od/	as	ᵛ̇ᵛ̇
/on/	as	ᵛ̇ᵛ̇
/od/	as	ᵛ̇ᵛ̇
/iŋ/ /eŋ/ and /əŋ/	as	ᵛ̇ᵛ̇
/ig/ /eg/ and /əg/	as	ᵛ̇ᵛ̇
/yŋ/ and /əŋ/	as	ᵛ̇ᵛ̇
/yg/ and /əg/	as	ᵛ̇ᵛ̇
/aŋ/ and /aan/	as	ᵛ̇
/ag/ and /aag/	as	ᵛ̇
/uŋ/ and /oŋ/	as	ᵛ̇
/ug/ and /og/	as	ᵛ̇
/oŋ/	as	ᵛ̇
/og/	as	ᵛ̇
/iw/ /ew/ and /əw/	as	ᵛ̇ᵛ̇ ; /ew/ is sometimes written ᵛ̇ᵛ̇, as in the word ᵛ̇ᵛ̇ /méw/ (also pronounced /mjó/)

/aw/ and /aaw/	as	၁
/əj/	as	၂, only in the syllable /ʔəjv/ 'final particle for vocative' written with the preceding graph
/aj/ and /aaj/	as	၃, compare /ɔj/ below
/uj/ and /oj/	as	၄
/ɔj/	as	၅; by some speakers pronounced /waj/; ၆ is pronounced /waj/ or /wɔj/; notice that ၇ is /waaj/
/ay/	as	၈

The symbol ၉ marks the last consonant written in each syllable where this consonant carries no other superimposed symbol; it is written with phonetically initial consonants which are graphically final, for instance, ၁ ၂, where ၁ stands for any initial consonant and ၂ for the following vowels /e/ or /ɛ/.

The medial consonants are written in the following way:

- /-j-/ as ၁; it combines with ၂ as ၁၂, with ၃ as ၁၃; /ɲj/ is written ၁၂ (see under initial consonants above)
- /-w-/ as ၄, except before ၅ with which it is usually written ၆
- /-r-/ as ၇; words written with medial ၇ are sometimes pronounced dissyllabically with /-ar-/; words written with initial consonant plus ၈ are sometimes pronounced monosyllabically with /-r-/

In Burmese loanwords ၉ is pronounced /jwa/ finally.

Repetition of a syllable is sometimes indicated by repetition of part of the graph:

၁ ၁ as ၁၁, ၂ ၂ as ၂၂, ၃ ၃ as ၃၃, ၄ ၄ as ၄၄, ၅ ၅ as ၅၅, ၆ ၆ as ၆၆, etc.

To indicate the exact pronunciation after the graph of each word the dictionary writers, Cushing and Mix, have devised two different systems.

J. N. Cushing in his Shan and English Dictionary indicates the tone of each syllable by means of a numeral (1, 2, 3, 4, and 5 for the five tones described above) and the vowel quality by means of a letter (c for closed,

m for middle, o for open) placed after the graph. The terms closed, middle and open are not always used in a strict phonetic sense.

Vowel indication in open syllable:

/i/, /e/, /y/, /a/, /aa/, /u/, /o/, and /ɔ/ are classed as c.

/ɛ/ and /ə/ are classed as o.

Vowel indication in closed syllable:

/i/, /y/, /a/, /u/, and /ɔ/ are classed as c.

/e/ is classed as m.

/ɛ/, /ə/, /aa/ and /o/ are classed as o.

Examples of Cushing's tone and vowel indications:

၁၈	1c	/khǎw/ 'horn; they'
၁၉	2c	/khàw/ 'knee'
၂၀	3c	/khaw/ 'rice; to enter'
၂၁	4c	/kháw/ 'balustrade, railing'
၂၂	1o	/khǎaw/ 'white'
၂၃	2o	/khàaw/ 'news'
၂၄	3o	/khaaw/ 'bars or cross timbers'
၂၅	4o	/kháaw/ 'stage of a journey'
၂၆	5o	/khâaw/ 'middle-aged'
၂၇	1c	/nǐm/ 'still, quiet'
၂၈	1m	/těm/ 'full'
၂၉	1o	/lěm/ 'sharp'
၃၀	4c	/mý/ 'hand'
၃၁	3o	/mə/ 'time; when'

H. W. Mix in her English and Shan Dictionary indicates both tones and vowel quality by means of little hooks placed after the graph. A hook in a low position means the same as Cushing's c, in a middle position the same as m, in a high position the same as o. The hook points right for the first, left for the second, down for the fourth and up for the fifth tone. For the third tone the hook is "z" shaped.

The symbols of both Cushing and Mix enable the reader to interpret accurately each graph in terms of the above given phonemics.

Later than the publication of these two dictionaries a minor orthographic reform has gradually established itself under the influence of the Burmese writing system according to which two tones, the fourth and the fifth, are indicated with all syllabic types. The fourth tone is indicated by means of \therefore , which was previously restricted to the combination η . For the fifth tone a single dot under the last element of the graph is used, \cdot . Examples are ၈၆ : /wén/ 'city', ၇၆ /lûg/ 'to rise'.

In 1955 a much more sweeping orthographic reform was officially adopted by the Shan States Government. This system is too recent to have been put to widespread use. Textbooks are now being prepared incorporating this new orthography. If the reform attains general acceptance, it will alter the status of the Shan script from one of the least perfect to one of the most perfect writing systems of Southeast Asia. Under the new system all syllabic types are accounted for in a simple and exhaustive manner.

A synopsis is given below of the essential features of the latest system. In open syllables the vowels are thus distinguished:

/e/ remains ၆ , /ɛ/ becomes ၆

/y/ remains ၇၆ , /ə/ becomes ၇၆

/u/ remains ၈ (and ၇ in the fifth tone), /o/ becomes ၈

/a/ remains zero, /aa/ becomes ၇

In closed syllables the following distinctions are introduced:

/i/ remains ၇ , /e/ becomes ၆ , /ɛ/ becomes ၆

Note that /im/ becomes ၇ , /em/ becomes ၆ , /ɛm/ becomes ၆ ;

/in/ remains ၇၆ , /en/ becomes ၆၆ , /ɛn/ becomes ၆၆

/y/ remains ၇ , /ə/ becomes ၇

/a/ remains zero, /aa/ becomes ၇

Note that /am/ remains ၇ , /aam/ becomes ၇ ;

/aj/ remains ၇ , /aaj/ becomes ၇

/u/ remains ၈ , /o/ becomes ၈

The medials are unchanged, notice that /kwaa/ is written ၇၇၇ , but /kwaan/ ၇၇၇ .

The tones are written as follows:

1st tone (zero mark) 4th tone :

2nd tone , 5th tone .

3rd tone ;