

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation internationale de normalisation  
Международная организация по стандартизации

**Doc Type:** Working Group Document  
**Title:** Revised Proposal to add the Ahom Script in the SMP of the UCS  
**Source:** Martin Hosken, Stephen Morey  
**Action:** For consideration by JTC1/SC2/WG2  
**Date:** 2012-09-14

**1. Introduction:** This document is a revision of the proposal presented as ISO/IEC JTC1/SG2/WG2 N4290, Unicode L2/12-222 and replaces those documents.

The Ahom script is used in North East India for the Tai Ahom language [AHO] and there are also some bilingual manuscripts from the 18th century that are partially in Assamese [ASM]. The Ahom Kingdom was set up (traditional date 1228) by a prince of Mau Long (now in the Dehong Dai prefecture of Yunnan province China). There are stone inscriptions in Yunnan very similar to Tai Ahom, and it is possible that the Ahoms brought their script from Mau Long when they arrived in Assam. The oldest surviving Ahom text, however, is the 'Snake Pillar' now in the State Museum of Assam, Guwahati, inscribed for King Siuw Hum Miung (1497-1539). In addition to this stone inscription and a few others, there are coins, brass plates and a large corpus of manuscripts. Until the early 19th century, manuscripts were written either on cloth, or more usually on the bark of the Sasi tree (*Aquillaria Agallocha*). Many thousands of such bark manuscripts have survived, often multiple copies of the same texts. Very few have been translated because knowledge of the language by the Ahom community is partial at best.

The Tai Ahom language went into decline from the late 17th century, and by 1800 was probably not spoken at all as mother tongue in Assam. However, the traditional priests, custodians of the manuscripts, kept up some religious practice throughout the 19th century and a revival of Ahom culture and language has been under way since at least 1920 (see Terwiel 1996 for a critique of this revival, also Morey 2002). Even before this, the revival of Ahom may be said to have begun in the late 18th century, with the compilation of two bilingual texts, the Bar Amra, an Ahom to Assamese lexicon, and the Loti Amra (see Barua and Phukon 1964, Tabassum and Morey 2010).

The modern period of use of the Ahom script commences with the publication of the Ahom-Assamese-English Dictionary (G. Barua 1920). Many dictionaries, word lists and primers have followed, first printed with a font style that was significantly different from the 18th century manuscripts. Since the 1997 development of an Ahom computer font (by Stephen Morey) the publication of Ahom texts has proceeded much more rapidly and there are now large numbers of books in Assam printed with at least some Ahom content.

**2. Structure:** Ahom is of the Brahmic type with an inherent vowel, medial consonants clustered with the initial consonant and a visible virama killer character, which has only become obligatory in modern Ahom. Ahom has no independent vowels, instead they are represented by AHOM LETTER A (U+11711) followed by the corresponding dependent vowel sign. Dependent vowel signs are stored following the initial consonant cluster.



There are various irregular vowel sequences used in archaic Ahom, for example  $\text{𑜀𑜢𑜤𑜰}$  AHOM LETTER NA (U+11703) AHOM VOWEL SIGN O (U+11728) AHOM VOWEL SIGN AW (U+11727) AHOM LETTER BA (U+11708) HOM SIGN VIRAMA (U+1172B) AHOM VOWEL SIGN U (U+11724) `star`.

Ahom uses the repeating of the final vowel, vowel sequence or consonant plus virama, as a way of indicating that the word should be reduplicated. Vowels that may be so repeated are: AHOM VOWEL SIGN AA (U+11721), AHOM VOWEL SIGN II (U+11723), AHOM VOWEL SIGN AW (U+11727), AHOM VOWEL SIGN AI (U+11729), AHOM VOWEL SIGN AM (U+1172A), AHOM SIGN VIRAMA (U+1172B), and the sequence AHOM LETTER BA (U+11708) AHOM SIGN VIRAMA (U+1172B). AHOM VOWEL SIGN U (U+11724) can be used at the end of an Ahom syllable to indicate vowel length or vowel quality.

**2.1 Vowels:** This section gives a non-exhaustive list of vowels and vowel sequences found in Ahom. In the examples the consonant AHOM LETTER KA (U+11700) is used as the syllable initial.

### *Common Vowels*

a	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN A (U+11720), VOWEL SIGN AA (U+11721)
a	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AA (U+11721)
i (with final C)	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN I (U+11722)
i	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN II (U+11723)
u (with final C)	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN U (U+11724)
u	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN UU (U+11725)
e	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN AW (U+11727)
o (with final C)	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN O (U+11728)
o	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN AA (U+11721)
ɔ / ʊ	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN I (U+11722), VOWEL SIGN U (U+11724), LETTER BA (U+11708), SIGN VIRAMA (U+1172B)
ai	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN AI (U+11729)
oi	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AI (U+11729)
ui	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN U (U+11724), LETTER NYA (U+1170F), SIGN VIRAMA (U+1172B)
am	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN AM (U+1172A)
um	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN U (U+11724), VOWEL SIGN AM (U+1172A)
om	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AM (U+1172A)
em	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN AM (U+1172A), VOWEL SIGN AW (U+11727)
au	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AW (U+11727) <sup>1</sup>
iu / eu	$\text{𑜀𑜢𑜤𑜰}$	Consonant, VOWEL SIGN I (U+11722), LETTER BA (U+11708), SIGN VIRAMA (U+1172B)

<sup>1</sup> Spellings with SIGN VIRAMA (U+1172B) occurring instead of VOWEL SIGN AW (U+11727) are also found.



### Less Common Sequences

𑜀𑜂𑜆𑜇	Consonant, VOWEL SIGN UU (U+11725), VOWEL SIGN A (U+11721)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN O (U+11728), VOWEL SIGN AA (U+11721)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AW (U+11727)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AW (U+11727), LETTER BA (U+11707), SIGN VIRAMA (U+1172B)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN O (U+11728), LETTER BA (U+11707), SIGN VIRAMA (U+1172B)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AW (U+11727), LETTER BA (U+11708), SIGN VIRAMA (U+1172B)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN AW (U+11727)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN E (U+11726)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN II (U+11723), VOWEL SIGN AW (U+11727)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AI (U+11729), VOWEL SIGN U (U+11724)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AW (U+11727), VOWEL SIGN U (U+11724)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN I (U+11722), LETTER NGA (U+11702), SIGN VIRAMA (U+1172B), VOWEL SIGN U (U+11724)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, LETTER NGA (U+11702), SIGN VIRAMA (U+1172B), VOWEL SIGN U (U+11724)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AM (U+1172A), VOWEL SIGN U (U+11724)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN O (U+11724), VOWEL SIGN A (U+11720), LETTER NGA (U+11702), SIGN VIRAMA (U+1172B)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AA (U+11721), VOWEL SIGN AA (U+11721)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN II (U+11723), VOWEL SIGN II (U+11723)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AM (U+1172A), VOWEL SIGN AM (U+1172A)
𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇	Consonant, VOWEL SIGN AI (U+11729), VOWEL SIGN AI (U+11729)

**3. Digits:** Knowledge of Ahom digits is incomplete with Ahom specific shapes only being known for 1, 7 8 and 10. Some other digit shapes are borrowed and then localised, from Burmese: 6 and 9 and the remaining digits: 3, 4, 5 and 20 are merely the words for those numbers in Ahom spelled out. 2 is closely derived from the letter kha. Lack of knowledge of digits is exacerbated by the common mixing of digits between systems (particularly with Burmese digits) in a number. A specific digit block has been included because some modern manuscripts do use specifically Ahom numbers.

In manuscript usage of Ahom, the digit 20 does get used as a digit. Numbers above 100 are typically fully spelled out as words since they occur within text. In manuscript usage numbers above 10 tend only to be used for page numbers. The following is an example of a page number:

𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 2 U+11732) 𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 20 U+1173B) 𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 10 U+1173A) 𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 8 U+11738) meaning '58'.

𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 6 (U+11736) 𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 20 (U+1173B) 𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 10 (U+1173A) 𑜀𑜂𑜆𑜇𑜂𑜂𑜆𑜇𑜂𑜂𑜆𑜇 (AHOM DIGIT 6 (U+11736) meaning '136'.

In addition, it is not uncommon to mix digits and word spellings to create a number. For example:



𑜋 AHOM DIGIT 20 (U+1173B) 𑜊 AHOM DIGIT 10 (U+1173A) 𑜍 AHOM LETTER A (U+11712) 𑜇 AHOM VOWEL I (U+11722) 𑜎 AHOM LETTER TA (U+11704) 𑜇 AHOM SIGN VIRAMA (U+1172B) meaning '31'.

Digits can also be used as letters in a mechanism akin to that used in texting English spelling. (E.g. l8r.)

Some of the Ahom digits are visually identical to the corresponding letter spellings for the digits. As a result, digits should be excluded from International Domain Names:

𑜋	AHOM DIGIT 3 (U+11733)	AHOM LETTER SA (U+1170F)	AHOM VOWEL SIGN AM (U+1172A)
𑜋	AHOM DIGIT 4 (U+11734)	AHOM LETTER SA (U+1170F)	AHOM VOWEL SIGN II (U+11723)
𑜊	AHOM DIGIT 5 (U+11735)	AHOM LETTER HA (U+11711)	AHOM VOWEL SIGN AA (U+11721)
𑜋	AHOM DIGIT 20 (U+1173A)	AHOM LETTER SA (U+1170F)	AHOM VOWEL SIGN AW (U+11727)

In addition, 𑜂 AHOM DIGIT 2 (U+11732) is visually identical to 𑜂 AHOM LETTER KHA (U+11701). There is an expectation that if modern Ahom were to start using the digits, that glyph variation would begin to occur and that the use of AHOM DIGIT 20 (U+1173B) would be dropped.

**4. Punctuation:** There are three punctuation marks. The two dandas are local to this script and not shared from any other script block. The AHOM SIGN RULAI (U+1173E) is used as a paragraph mark. AHOM SYMBOL VI (U+1173F) corresponds to MYANMAR SYMBOL AITON EXCLAMATION (U+AA77).

**5. Word spacing:** Modern Ahom and some manuscripts have word spaces. Other manuscripts have no word spaces.

**6. Variant Forms:** Ahom has a number of variant and ligature glyphs that are worthy of attention.

𑜇 This is a contextual ligature of AHOM VOWEL SIGN I (U+11722) AHOM VOWEL SIGN U (U+11724). It is only used if there is no ambiguity that closing the right hand side of the initial consonant will make it look like another consonant. For example, one would not render AHOM LETTER NGA (U+11702) AHOM VOWEL SIGN I (U+11722) AHOM VOWEL SIGN U (U+11724) using this ligature (𑜇) because it would look too much like AHOM LETTER MA (U+11709) AHOM VOWEL SIGN I (U+11722) AHOM VOWEL SIGN U (U+11724) (𑜇) which can safely use this ligature.

These consonants may not take the ligature: 𑜂 AHOM LETTER NGA (U+11702) 𑜃 AHOM LETTER NA (U+11703) 𑜄 AHOM LETTER DA (U+11713) 𑜅 AHOM LETTER NYA (U+11710)

𑜇 This is a ligature of AHOM VOWEL SIGN O (U+11728) AHOM VOWEL SIGN A (U+11721). It is believed to convey the glide-vowel combination /wa/, as 𑜊 AHOM LETTER KA (U+11700) AHOM VOWEL SIGN O (U+11728) AHOM VOWEL SIGN AA (U+11721) AHOM LETTER NGA (U+11702) AHOM SIGN VIRAMA (U+1172B), pronounced /kwaang/.

𑜈 This is a font variant of AHOM LETTER JHA (U+11719) in the form found in older manuscripts. The main form AHOM LETTER JHA (U+11719) is that adopted for use at the beginning of the Ahom revival in the 1920s.

𑜊 This is a font variant for AHOM DIGIT 3 (U+11733) and is equivalent to AHOM LETTER HA (U+11711) AHOM VOWEL SIGN U (U+11724)

𑜈 This is a font variant for AHOM DIGIT 4 (U+11734) and is equivalent to AHOM LETTER THA (U+1170C)

In manuscript Ahom there are a number of variations, not found in modern Ahom, that are being analysed. The following is a discussion of such variations. While this variation may be most easily handled in texts for analysis, using a Variation Selection character, separate codepoints are proposed due to a general aversion to the use of Variation Selection characters in new encodings.

𑜂 This uses the ligating ra: AHOM LETTER KHA (U+11701) AHOM MEDIAL LIGATING RA (U+1171F).



- The sequence AHOM VOWEL SIGN AW (U+11727) AHOM VOWEL SIGN AM (U+1172A) ligates such that the AHOM VOWEL SIGN AM (U+1172A) renders before the AHOM VOWEL SIGN AW (U+11727). This can occur rarely with the sequences AHOM VOWEL SIGN I (U+11722) AHOM VOWEL SIGN AM (U+1172A).

**8. Sort order:** A standard sort order for Ahom has not been agreed. There are various in existence. Sorting Ahom gives higher priority to the final consonant than to the vowel. In fact, early sorting gave higher priority to the final consonant than to the initial consonant! But nobody is recommending this for a modern sorting. For DUCET the ordering is not expected to give precedence to the final consonant, although it would be expected for language specific tailoring(s).

m n t p ph b m j ch th r l s ny h (a) d dh g gh bh jh

m n ŋ k kh ng p d t m ph th s r j ny l b h (a)

*Final Consonants:* In modern usage, and for default collation, final consonants follow the initial consonant order, but there are historic orders for these that differ from the orders for initial consonants. The most authoritative order, from Bar Amra, is:

-v	-m	-ʃ	-ɲ	-ɒ	-v
-b(w)	-k	-ng	-n	-t	-p

ṁ	ṁ	ṁ <sup>o</sup>	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ	ṁ
ka	ka	ki	ku	ke	kai	ko	kav	kam	kum	kom



Then follows the closed syllables, here shown with initial and final k:

ကက	ကိက	ကုက	က့က	ကလက
kak	kik	kuk	kok	kvk

Finally there are two extra open syllables:

ကလိ	ကိ
kv	koi

*Medials*: Medial characters take a primary sort relation to the other characters in the order. Medials sort after both consonants and vowels. This has the effect of producing an order of the form: ka ka ki ku ... kra kra kri kru ... kla kla kli klu ... kha kha khi khu ...

For the purposes of default collation, vowels are ordered according to their codepoint value, likewise for the medials. The relative block order is: Consonants, Vowels, Medials.



U+11700

Ahom

U+1173F

	1170	1171	1172	1173
0	ᩈ	ᩈ	ᩈᩈ	ᩈ
1	ᩈ	ᩈ	ᩈᩈ	ᩈᩈ
2	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈ
3	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈᩈ
4	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈᩈᩈ
5	ᩈ	ᩈ	ᩈᩈᩈ	ᩈᩈ
6	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈᩈᩈ
7	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈᩈ
8	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈᩈᩈ
9	ᩈ	ᩈᩈ	ᩈᩈᩈ	ᩈᩈᩈ
A	ᩈ		ᩈᩈᩈ	ᩈᩈᩈ
B	ᩈ		ᩈᩈᩈ	ᩈᩈᩈ
C	ᩈ			ᩈ
D	ᩈ	ᩈᩈᩈ		ᩈᩈ
E	ᩈ	ᩈᩈᩈ		ᩈᩈᩈ
F	ᩈ	ᩈ		ᩈᩈᩈ



11700	ṡ	AHOM LETTER KA
11701	ṛ	AHOM LETTER KHA
11702	ṣ	AHOM LETTER NGA
11703	ṭ	AHOM LETTER NA
11704	ṭṛ	AHOM LETTER TA
11705	ṭṛ	AHOM LETTER ALTERNATE TA
11706	ṭṛ	AHOM LETTER PA
11707	ṭṛ	AHOM LETTER PHA
11708	ṭṛ	AHOM LETTER BA
11709	ṭṛ	AHOM LETTER MA
1170A	ṭṛ	AHOM LETTER JA
1170B	ṭṛ	AHOM LETTER CHA
1170C	ṭṛ	AHOM LETTER THA
1170D	ṭṛ	AHOM LETTER RA
1170E	ṭṛ	AHOM LETTER LA
1170F	ṭṛ	AHOM LETTER SA
11710	ṭṛ	AHOM LETTER NYA
11711	ṭṛ	AHOM LETTER HA
11712	ṭṛ	AHOM LETTER A
11713	ṭṛ	AHOM LETTER DA
11714	ṭṛ	AHOM LETTER DHA
11715	ṭṛ	AHOM LETTER GA
11716	ṭṛ	AHOM LETTER ALTERNATE GA
11717	ṭṛ	AHOM LETTER GHA
11718	ṭṛ	AHOM LETTER BHA
11719	ṭṛ	AHOM LETTER JHA

1171D	◌ʰ	AHOM CONSONANT SIGN MEDIAL LA
1171E	◌ᵛ	AHOM CONSONANT SIGN MEDIAL RA
1171F	◌ᵝ	AHOM CONSONANT SIGN MEDIAL LIGATING RA

11720		AHOM VOWEL SIGN A
11721		AHOM VOWEL SIGN AA
11722		AHOM VOWEL SIGN I
11723		AHOM VOWEL SIGN II
11724		AHOM VOWEL SIGN U
11725		AHOM VOWEL SIGN UU
11726		AHOM VOWEL SIGN E
11727		AHOM VOWEL SIGN AW
11728		AHOM VOWEL SIGN O
11729		AHOM VOWEL SIGN AI
1172A		AHOM VOWEL SIGN AM
1172B		AHOM SIGN VIRAMA

11730	୦	AHOM DIGIT 0
11731	୧	AHOM DIGIT 1
11732	୨	AHOM DIGIT 2
11733	୩	AHOM DIGIT 3
11734	୪	AHOM DIGIT 4
11735	୫	AHOM DIGIT 5
11736	୬	AHOM DIGIT 6
11737	୭	AHOM DIGIT 7
11738	୮	AHOM DIGIT 8
11739	୯	AHOM DIGIT 9
1173A	୧୦	AHOM DIGIT 10
1173B	୧୧	AHOM DIGIT 20

1173C	𑀓	AHOM SIGN SMALL SECTION
1173D	𑀔	AHOM SIGN SECTION
1173E	𑀕	AHOM SIGN RULAI
1173F	𑀖	AHOM SYMBOL VI



## Unicode Properties

11700;AHOM LETTER KA;Lo;0;L;;;;N;;;;;  
11701;AHOM LETTER KHA;Lo;0;L;;;;N;;;;;  
11702;AHOM LETTER NGA;Lo;0;L;;;;N;;;;;  
11703;AHOM LETTER NA;Lo;0;L;;;;N;;;;;  
11704;AHOM LETTER TA;Lo;0;L;;;;N;;;;;  
11705;AHOM LETTER ALTERNATE TA;Lo;0;L;;;;N;;;;;  
11706;AHOM LETTER PA;Lo;0;L;;;;N;;;;;  
11707;AHOM LETTER PHA;Lo;0;L;;;;N;;;;;  
11708;AHOM LETTER BA;Lo;0;L;;;;N;;;;;  
11709;AHOM LETTER MA;Lo;0;L;;;;N;;;;;  
1170A;AHOM LETTER JA;Lo;0;L;;;;N;;;;;  
1170B;AHOM LETTER CHA;Lo;0;L;;;;N;;;;;  
1170C;AHOM LETTER THA;Lo;0;L;;;;N;;;;;  
1170D;AHOM LETTER RA;Lo;0;L;;;;N;;;;;  
1170E;AHOM LETTER LA;Lo;0;L;;;;N;;;;;  
1170F;AHOM LETTER SA;Lo;0;L;;;;N;;;;;  
11710;AHOM LETTER NYA;Lo;0;L;;;;N;;;;;  
11711;AHOM LETTER HA;Lo;0;L;;;;N;;;;;  
11712;AHOM LETTER A;Lo;0;L;;;;N;;;;;  
11713;AHOM LETTER DA;Lo;0;L;;;;N;;;;;  
11714;AHOM LETTER DHA;Lo;0;L;;;;N;;;;;  
11715;AHOM LETTER GA;Lo;0;L;;;;N;;;;;  
11716;AHOM LETTER ALTERNATE GA;Lo;0;L;;;;N;;;;;  
11717;AHOM LETTER GHA;Lo;0;L;;;;N;;;;;  
11718;AHOM LETTER BHA;Lo;0;L;;;;N;;;;;  
11719;AHOM LETTER JHA;Lo;0;L;;;;N;;;;;  
  
1171D;AHOM CONSONANT SIGN MEDIAL LA;Mn;0;NSM;;;;N;;;;;  
1171E;AHOM CONSONANT SIGN MEDIAL RA;Mn;0;NSM;;;;N;;;;;  
1171F;AHOM CONSONANT SIGN MEDIAL LIGATING RA;Mn;0;NSM;;;;N;;;;;  
  
11720;AHOM VOWEL SIGN A;Mc;0;L;;;;N;;;;;  
11721;AHOM VOWEL SIGN AA;Mc;0;L;;;;N;;;;;  
11722;AHOM VOWEL SIGN I;Mn;0;NSM;;;;N;;;;;  
11723;AHOM VOWEL SIGN II;Mn;0;NSM;;;;N;;;;;  
11724;AHOM VOWEL SIGN U;Mn;0;L;;;;N;;;;;  
11725;AHOM VOWEL SIGN UU;Mn;0;NSM;;;;N;;;;;  
11726;AHOM VOWEL SIGN E;Mn;0;NSM;;;;N;;;;;  
11727;AHOM VOWEL SIGN AW;Mn;0;NSM;;;;N;;;;;  
11728;AHOM VOWEL SIGN O;Mn;0;NSM;;;;N;;;;;  
11729;AHOM VOWEL SIGN AI;Mn;0;NSM;;;;N;;;;;  
1172A;AHOM VOWEL SIGN AM;Mn;0;NSM;;;;N;;;;;  
1172B;AHOM SIGN VIRAMA;Mn;0;NSM;;;;N;;;;;  
  
11730;AHOM DIGIT 0;Nd;0;L;;0;0;0;N;;;;;  
11731;AHOM DIGIT 1;Nd;0;L;;1;1;1;N;;;;;  
11732;AHOM DIGIT 2;Nd;0;L;;2;2;2;N;;;;;  
11733;AHOM DIGIT 3;Nd;0;L;;3;3;3;N;;;;;  
11734;AHOM DIGIT 4;Nd;0;L;;4;4;4;N;;;;;  
11735;AHOM DIGIT 5;Nd;0;L;;5;5;5;N;;;;;  
11736;AHOM DIGIT 6;Nd;0;L;;6;6;6;N;;;;;  
11737;AHOM DIGIT 7;Nd;0;L;;7;7;7;N;;;;;  
11738;AHOM DIGIT 8;Nd;0;L;;8;8;8;N;;;;;  
11739;AHOM DIGIT 9;Nd;0;L;;9;9;9;N;;;;;  
1173A;AHOM DIGIT 10;Nd;0;L;;10;10;10;N;;;;;  
1173B;AHOM DIGIT 20;Nd;0;L;;20;20;20;N;;;;;  
1173C;AHOM SIGN SMALL SECTION;Po;0;L;;;;N;;;;;  
1173D;AHOM SIGN SECTION;Po;0;L;;;;N;;;;;  
1173E;AHOM SIGN RULAI;Po;0;L;;;;N;;;;;  
1173F;AHOM SYMBOL VI;Lo;0;L;;;;N;;;;;



## Examples

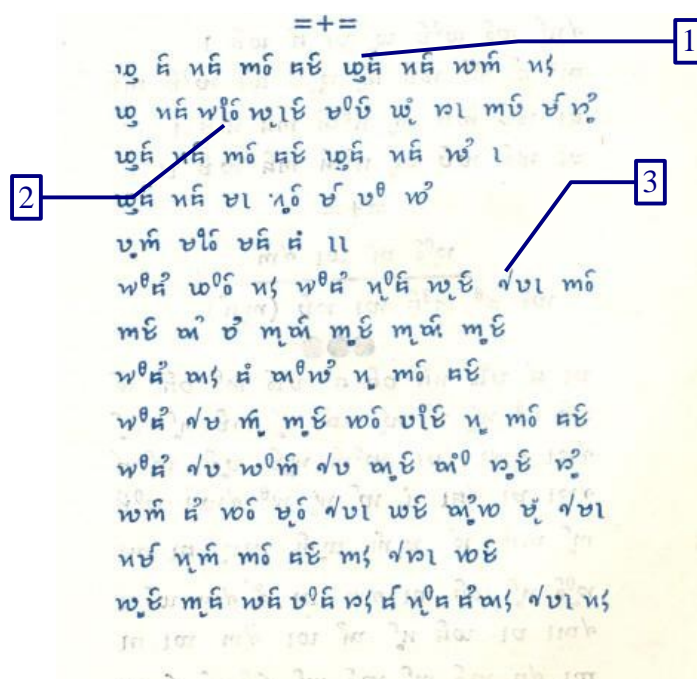


Figure 1: Lik Tai Khwam Tai page 7

This modern text was printed before the advent of computer fonts.

1. Sample of AHOM CONSONANT SIGN MEDIAL LA (U+1171D).
2. AHOM VOWEL SIGN I (U+11722) AHOM VOWEL SIGN U (U+11724) ligature.
3. Alternate, modern glyph, for AHOM VOWEL SIGN E (U+11726).

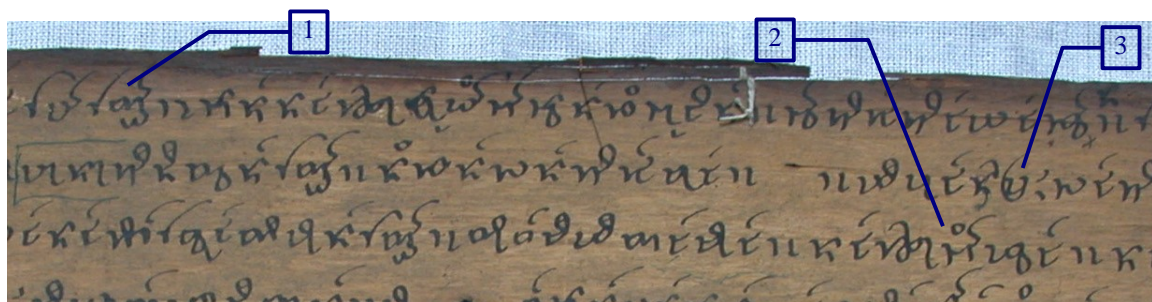


Figure 2: NemiMang p2v

1. This shows an example of a typographical insertion. The BA is to be inserted after the TA it is written below. This relation does not need to be encoded in plain text.
2. Example of AHOM VOWEL SIGN AW (U+11727) AHOM VOWEL SIGN AM (U+1172A) ligature
3. Example of AHOM CONSONANT SIGN MEDIAL RA (U+1171E).

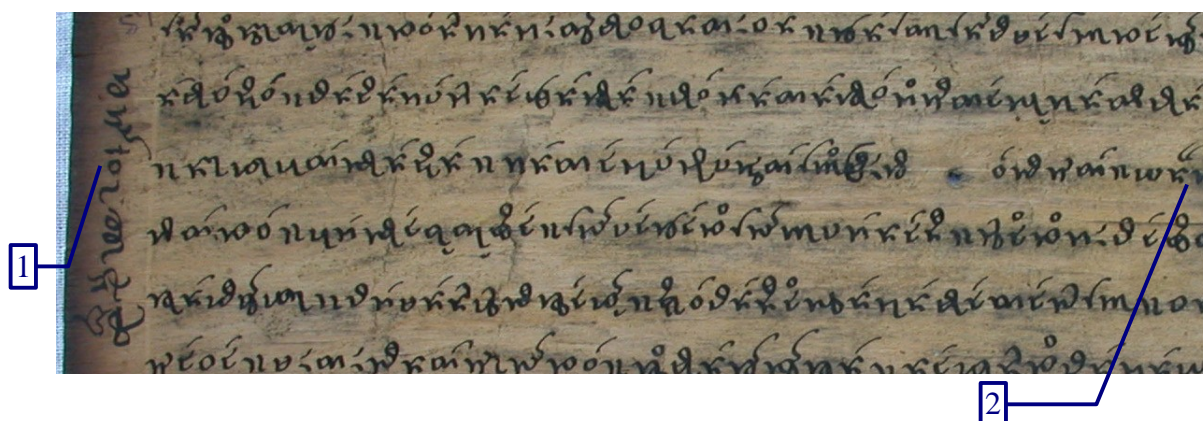




Figure 3: NemiMang p58v

1. '58' in Ahom and also in Burmese script
2. Example of reduplication through repeated AHOM SIGN VIRAMA (U+1172B).

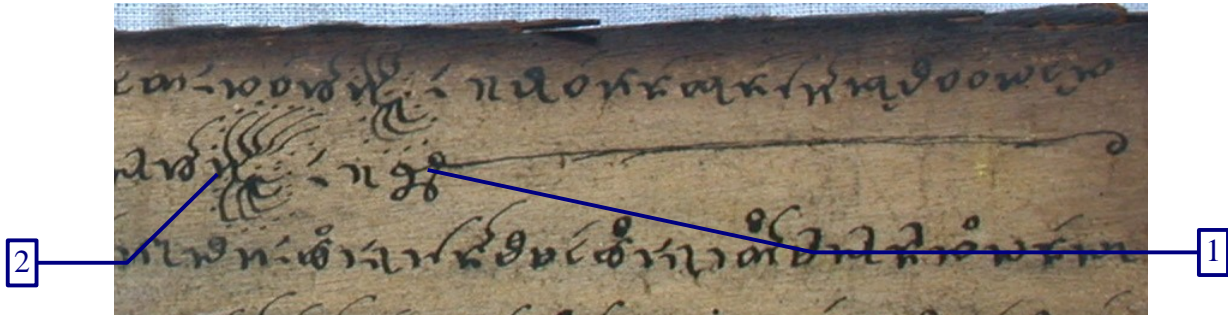


Figure 4: NemiMang p66r showing [1] text final embellishment, perhaps a character akin to TAI THAM SIGN KEOW (U+1AA3). This only occurs in the one text, so there is no intent to encode this within the Ahom block. Notice [2] the highly embellished /vi/ AHOM SYMBOL VI (U+1173F).

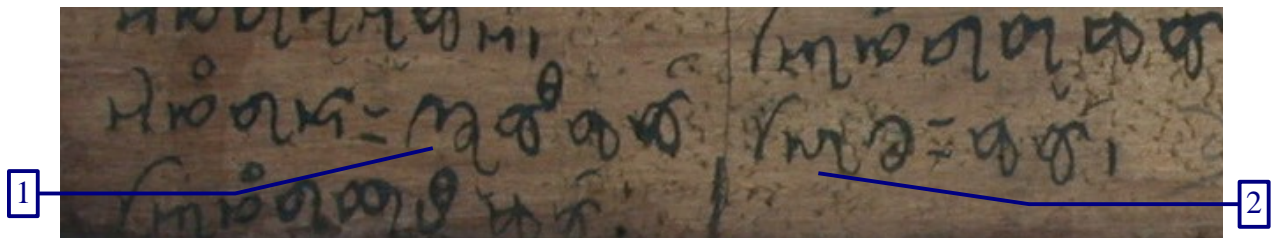


Figure 5: Phukan 1 p1v shows the two alternate forms of GA in the same document: 1) AHOM LETTER GA (U+11715) AHOM VOWEL SIGN U (U+11724), 2) AHOM LETTER ALTERNATE GA (U+11716)

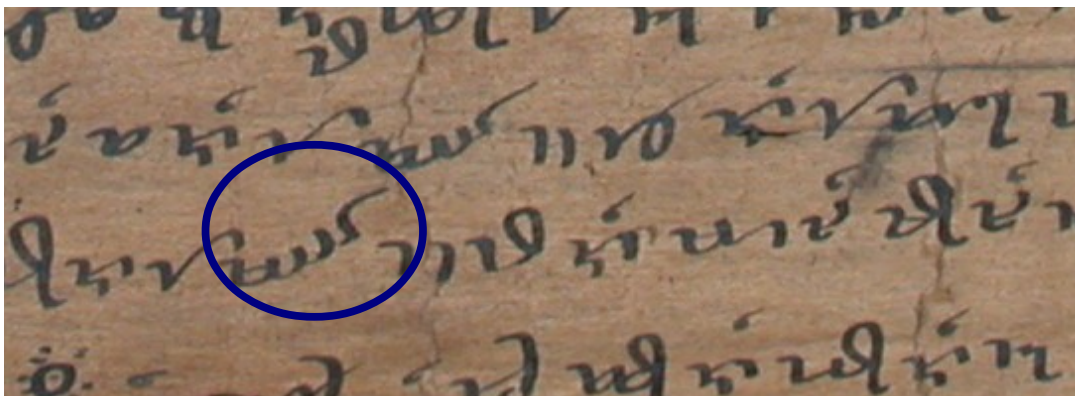


Figure 6: Mohan 9 p2r showing AHOM LETTER ALTERNATE TA (U+11705) AHOM LETTER JA (U+1170E) AHOM VOWEL SIGN AW (U+11725).



## Bibliography

- Barua, Bimala Kanta and N.N. Deodhari Phukan. 1964. Ahom Lexicons, Based on Original Tai Manuscripts. Guwahati: Department of Historical and Antiquarian Studies
- Barua, Golap Chandra, 1920, Ahom-Assamese-English Dictionary. Calcutta: Baptist Mission Press (printed under the authority of the Assam Administration).
- Hazarika, Nagen (ed). 1990 Lik Tai Khwam Tai (Tai letters and Tai words) Souvenir of the 8th Annual conference of Ban Ok Pup Lik Mioung Tai (Eastern Tai Literary Association)
- Kar, Babul. 2005. Tai Ahom Alphabet Book. Sepon, Assam: Tai Literature Associate
- Morey, Stephen D. 2002a. Tai languages of Assam, a progress report – Does anything remain of the Tai Ahom language? in David and Maya Bradley, (eds). Language Maintenance for Endangered Languages: An Active Approach. London: Curzon Press. 98-113.
- Tabassum, Zeenat and S.D. Morey. 2010. 'Linguistic features of the Ahom Bar Amra,' , in S. Morey and M. Post (eds) North East Indian Linguistics II. Delhi: Cambridge University Press, India. 70-89
- Terwiel, B.J. 1996. 'Recreating the Past: Revivalism in Northeastern India' in *Bijdragen - Journal of the Royal Institute of Linguistics and Anthropology*, (Leiden) No. 152, p.275-292.

## Acknowledgements

Thanks go to Payap University Linguistics Institute, Chiang Mai, Thailand, under whose auspices this work is done. The work on Tai Ahom has been funded by a grant from the Volkswagen Stiftung (DoBeS program) for the project The Traditional Songs and Poetry of Upper Assam, (<http://www.mpi.nl/DoBeS>), and also the Centre for Research into Computational Linguistics, Bangkok who also maintain the on-line Ahom Dictionary (<http://sealang.net/ahom>). The translation of Ahom manuscripts has been done by Stephen Morey and Chaichuen Khamdaengyodtai (Rajabhat University, Chiang Mai), with transcriptions done by Zeenat Tabassum (Gauhati University, Assam). The traditional Ahom priests who have given great assistance include Chaw Junaram Sangbun Phukan, Chaw Tileswar Mohan and Chaw Medini Mohan. The Institute for Tai Studies and Research, especially the Director, Prof Girin Phukon, have also assisted a great deal. This proposal is the culmination of work since 1997 on the Ahom texts and script, work that commenced with the Ahom computer font made by Stephen Morey and widely used in Assam ever since.



**ISO/IEC JTC 1/SC 2/WG 2**  
**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS**  
**FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>2</sup>.**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. Title:	<i>Ahom</i>
2. Requester's name:	<i>Martin Hosken</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>02/05/12</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>X</i>
(or) More information will be provided later:	

**B. Technical – General**

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<i>X</i>
Proposed name of script:	<i>Ahom</i>
b. The proposal is for addition of character(s) to an existing block:	
Name of the existing block:	
2. Number of characters in proposal:	<i>57</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input checked="" type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
	G-Obscure or questionable usage symbols <input type="checkbox"/>
4. Is a repertoire including character names provided?	<i>yes</i>
a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?	<i>yes</i>
b. Are the character shapes attached in a legible form suitable for review?	<i>yes</i>
5. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?	<i>Stephen Morey</i>
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:	
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<i>yes</i>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<i>yes</i>
7. Special encoding issues:	
Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?	<i>yes</i>
<i>sorting</i>	
8. Additional Information:	
<p>Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <a href="http://www.unicode.org">http://www.unicode.org</a> for such information on other scripts. Also see <a href="http://www.unicode.org/Public/UNIDATA/UCD.html">http://www.unicode.org/Public/UNIDATA/UCD.html</a> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.</p>	

<sup>2</sup> Form number: N3102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03)



### C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	yes
If YES explain	<i>This finalises N3928 L2/10-359</i>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	yes
If YES, with whom?	<i>Stephen Morey</i>
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	yes
Reference:	<i>this document</i>
4. The context of use for the proposed characters (type of use; common or rare)	common
Reference:	
5. Are the proposed characters in current use by the user community?	yes
If YES, where? Reference:	
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	no
If YES, is a rationale provided?	
If YES, reference:	
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	yes
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	no
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	no
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?	yes
If YES, is a rationale for its inclusion provided?	yes
If YES, reference:	<i>this document</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	yes
If YES, is a rationale for such use provided?	yes
If YES, reference:	<i>this document</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	no
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	no
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	no
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	