

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, revised 2012-10-23

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 AHOM SIGN KILLER (U+1172B), 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 𑜀𑜃𑜫 AHOM LETTER NA (U+11703) AHOM VOWEL SIGN O (U+11728) AHOM VOWEL SIGN AW (U+11727) AHOM LETTER BA (U+11708) AHOM 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.

As a Brahmic script, Ahom has a primary base consonant with diacritical vowels. For encoding order purposes, the non-spacing vowel diacritics may be grouped into 4 classes:

Medials	AHOM CONSONANT SIGN MEDIAL LA (U+1171D), AHOM CONSONANT SIGN MEDIAL RA (U+1171E), AHOM CONSONANT SIGN MEDIAL LIGATING RA (U+1171F)
Pre-Vowels	AHOM VOWEL SIGN E (U+11726)
Upper-Vowels	AHOM SIGN I (U+11722), AHOM SIGN II (U+11723), AHOM SIGN AM (U+1172A), AHOM SIGN AI (U+11729), AHOM SIGN AW (U+11727)
Lower-Vowels	AHOM SIGN U (U+11724), AHOM SIGN UU (U+11725), AHOM SIGN O (U+11728)
AA-Vowel	AHOM SIGN AA (U+11721)
A-Vowel	AHOM SIGN A (U+11720)

This allows us to specify the encoding order of non-spacing marks following a consonant as:

Consonant Medial? Pre-Vowel? Upper-Vowel Lower-Vowel* AA-Vowel* A-Vowel?*

(We use standard regular expression modifiers: ? - optional, * - 0 or more)

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	𑜀𑜃𑜫	Consonant, VOWEL SIGN AA (U+11721), VOWEL SIGN A (U+11720)
a	𑜀𑜃𑜫	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AA (U+11721)
i (with final C)	𑜀𑜃𑜫	Consonant, VOWEL SIGN I (U+11722)
i	𑜀𑜃𑜫	Consonant, VOWEL SIGN II (U+11723)
u (with final C)	𑜀𑜃𑜫	Consonant, VOWEL SIGN U (U+11724)
u	𑜀𑜃𑜫	Consonant, VOWEL SIGN UU (U+11725)
e	𑜀𑜃𑜫	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN AW (U+11727)
o (with final C)	𑜀𑜃𑜫	Consonant, VOWEL SIGN O (U+11728)
o	𑜀𑜃𑜫	Consonant, VOWEL SIGN E (U+11726), VOWEL SIGN AA (U+11721)
ɔ / ʊ	𑜀𑜃𑜫	Consonant, VOWEL SIGN I (U+11722), VOWEL SIGN U (U+11724), LETTER BA (U+11708), SIGN VIRAMA (U+1172B)
ai	𑜀𑜃𑜫	Consonant, VOWEL SIGN AI (U+11729)

oi	𑍑 ^o	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AI (U+11729)
ui	𑍑 ^u	Consonant, VOWEL SIGN U (U+11724), LETTER NYA (U+1170F), SIGN VIRAMA (U+1172B)
am	𑍑	Consonant, VOWEL SIGN AM (U+1172A)
um	𑍑 ^u	Consonant, VOWEL SIGN U (U+11724), VOWEL SIGN AM (U+1172A)
om	𑍑 ^o	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AM (U+1172A)
em	𑍑 ^r	Consonant, VOWEL SIGN AM (U+1172A), VOWEL SIGN AW (U+11727)
au	𑍑 ^r	Consonant, VOWEL SIGN O (U+11728), VOWEL SIGN AW (U+11727) ¹
iu / eu	𑍑 ^o	Consonant, VOWEL SIGN I (U+11722), LETTER BA (U+11708), SIGN VIRAMA (U+1172B)

Less Common Sequences

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

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

3. Digits: Ahom digits do not follow a radix 10 system. 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 number 20 does get used as a number. 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:

𑜇 (AHOM DIGIT TWO U+11732) 𑜇𑜇 (AHOM NUMBER TWENTY U+1173B) 𑜇𑜇 (AHOM NUMBER TEN U+1173A) 𑜇𑜇 (AHOM DIGIT EIGHT U+11738) meaning '58'. See Figure 3.

𑜇 (AHOM DIGIT SIX (U+11736) 𑜇𑜇 (AHOM NUMBER TWENTY (U+1173B) 𑜇𑜇 (AHOM NUMBER TEN (U+1173A) 𑜇 (AHOM DIGIT SIX (U+11736) meaning '136'.

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

𑜇𑜇 (AHOM NUMBER TWENTY (U+1173B) 𑜇𑜇 (AHOM NUMBER TEN (U+1173A) 𑜇 (AHOM LETTER A (U+11712) 𑜇 (AHOM VOWEL I (U+11722) 𑜇 (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:

𑜇	AHOM DIGIT THREE (U+11733)	𑜇	AHOM LETTER SA (U+1170F)	𑜇	AHOM VOWEL SIGN AM (U+1172A)
𑜇	AHOM DIGIT FOUR (U+11734)	𑜇	AHOM LETTER SA (U+1170F)	𑜇	AHOM VOWEL SIGN II (U+11723)
𑜇	AHOM DIGIT FIVE (U+11735)	𑜇	AHOM LETTER HA (U+11711)	𑜇	AHOM VOWEL SIGN AA (U+11721)
𑜇𑜇	AHOM NUMBER TWENTY (U+1173A)	𑜇	AHOM LETTER SA (U+1170F)	𑜇	AHOM VOWEL SIGN AW (U+11727)

In addition, 𑜇 (AHOM DIGIT TWO (U+11732) is visually identical to 𑜇 (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 NUMBER TWENTY (U+1173B) would be dropped.

4. Punctuation: There are three punctuation marks. The two dandas AHOM SIGN SMALL SECTION (U+1173C) and AHOM SIGN SECTION (U+1173D) 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: 𑜇 (AHOM LETTER NGA (U+11702) 𑜇 (AHOM LETTER NA (U+11703) 𑜇 (AHOM LETTER DA (U+11713) 𑜇 (AHOM LETTER NYA (U+11710)



This is a ligature of AHOM VOWEL SIGN O (U+11728) AHOM VOWEL SIGN AA (U+11721). It is believed to convey the glide-vowel combination /wa/, as 𑜀𑜃𑜫 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/.



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.



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



This is a font variant for AHOM DIGIT FOUR (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.



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



This uses the ligating ra: AHOM LETTER PHA (U+11707) AHOM MEDIAL LIGATING RA (U+1171F).



This is a font variant of AHOM LETTER GA (U+11715) but it may also occur along with AHOM LETTER GA (U+11715) in some rare manuscripts. Where contrast needs to be encoded the AHOM LETTER ALTERNATE GA (U+11716) should be used.



This is a font variant of AHOM LETTER GA (U+11714).



This is actually two characters AHOM LETTER TA (U+11704) AHOM LETTER JA (U+11709) conjoined. But the TA has been shortened. This is an example of where the alternate TA is used. Thus this sequence is stored AHOM LETTER ALTERNATE TA (U+11705) AHOM LETTER JA (U+11709).

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) (see figure 2). This can occur rarely with the sequences AHOM VOWEL SIGN I (U+11722) AHOM VOWEL SIGN AM (U+1172A).

7. Character Naming: Character names follow the phonetics of the characters. AHOM LETTER JA (U+1170A) acts like a YA but is pronounced in modern Ahom as JA. Likewise AHOM LETTER BA (U+11708) acts like a WA but is pronounced in modern Ahom as BA.

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).

Initial Consonant: Several orders exist. The proposed ordering, as approved by a meeting of Ahom community leaders held in Moran, Sibsagar District, Assam, in October 2011, is based on Barua (1920):

𑜀 𑜁 𑜂 𑜃 𑜄 𑜅 𑜆 𑜇 𑜈 𑜉 𑜊 𑜋 𑜌 𑜍 𑜎 𑜏 𑜐 𑜑 𑜒 𑜓 𑜔 𑜕 𑜖 𑜗 𑜘 𑜙 𑜚 𑜛
k kh ng n t p ph b m j ch th r l s ny h (a) d dh g gh bh jh

Another order, as given in the Bar Amra and other older Ahom manuscripts (as analysed by Stephen Morey), is:

𑜀 𑜁 𑜂 𑜃 𑜄 𑜅 𑜆 𑜇 𑜈 𑜉 𑜊 𑜋 𑜌 𑜍 𑜎 𑜏 𑜐 𑜑 𑜒 𑜓 𑜔 𑜕 𑜖 𑜗 𑜘 𑜙 𑜚 𑜛
k kh ng n ch t p d m ph th s r j ny l b h (a)

This order is found in some writing practice books of the Tai Ahom from the 18th century. However an ongoing study of the 18th century practice books and other sources suggest that there was no one ordering standard.

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:

-𑜎	-𑜄	-𑜑	-𑜕	-𑜅	-𑜖	-𑜇	-𑜓
-k	-ng	-n	-p	-m	-ny	-t	-w(b)

and the most common, from Barua and Phukan is:

-𑜓	-𑜎	-𑜄	-𑜑	-𑜇	-𑜕
-b(w)	-k	-ng	-n	-t	-p

Vowels: Vowels fall into two sequences: open and closed syllables. The open vowel sequence is (shown with an initial k):

𑜎	𑜎	𑜎 [◌]	𑜎 _◌	𑜎 [◌]	𑜎 [◌]	𑜎 [◌]	𑜎 [◌]	𑜎	𑜎	𑜎
ka	ka	ki	ku	ke	kai	ko	kav	kam	kum	kom

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

𑜎𑜎	𑜎𑜎	𑜎𑜎	𑜎𑜎	𑜎𑜎
kak	kik	kuk	kok	kvk

Finally there are two extra open syllables:

𑜎 [◌]	𑜎 [◌]
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 weight group order is: Consonants < Vowels < Medials. Each group is weighted in the same order as the order in the code chart.

AHOM LETTER KA (U+11700) < AHOM LETTER KHA (U+11701) < ... < AHOM LETTER JHA (U+11719) < AHOM SIGN A (U+11720) < AHOM SIGN AA (U+11721) < ... < AHOM SIGN AM (U+1172A) < AHOM SIGN KILLER (U+1172B) < AHOM CONSONANT SIGN MEDIAL LA (U+1171D) < AHOM CONSONANT SIGN MEDIAL RA (U+1171E) < AHOM CONSONANT SIGN MEDIAL LIGATING RA (U+1171F)

The digits sort as other digits in other scripts. Details of the symbols in section 4 of this document gives information on where the symbol characters should collate.

	1170	1171	1172	1173
0	ᩉ	ᩈ	ᩇ	ᩆ
1	ᩊ	ᩋ	ᩌ	ᩍ
2	ᩎ	ᩏ	ᩐ	ᩑ
3	ᩒ	ᩓ	ᩔ	ᩕ
4	ᩗ	ᩘ	ᩙ	ᩚ
5	ᩛ	ᩜ	ᩝ	ᩞ
6	᩟	᩠	ᩡ	ᩢ
7	ᩣ	ᩤ	ᩥ	ᩦ
8	ᩧ	ᩨ	ᩩ	ᩪ
9	ᩬ	ᩭ	ᩮ	ᩯ
A	ᩰ		ᩱ	ᩲ
B	ᩳ		ᩴ	᩵
C	᩶			᩷
D	᩸	᩹		᩺
E	᩻	᩼		᩽
F	᩿	᩺		᩻

Consonants

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

Medials

1171D	ṅ	AHOM CONSONANT SIGN MEDIAL LA
1171E	ṅ	AHOM CONSONANT SIGN MEDIAL RA
1171F	ṅ	AHOM CONSONANT SIGN MEDIAL LIGATING RA

Vowels

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

Digits

11730	ṅ	AHOM DIGIT ZERO
11731	ṅ	AHOM DIGIT ONE
11732	ṅ	AHOM DIGIT TWO
11733	ṅ	AHOM DIGIT THREE
11734	ṅ	AHOM DIGIT FOUR
11735	ṅ	AHOM DIGIT FIVE
11736	ṅ	AHOM DIGIT SIX
11737	ṅ	AHOM DIGIT SEVEN
11738	ṅ	AHOM DIGIT EIGHT
11739	ṅ	AHOM DIGIT NINE
1173A	ṅ	AHOM NUMBER TEN
1173B	ṅ	AHOM NUMBER TWENTY

Punctuation

1173C	ṅ	AHOM SIGN SMALL SECTION
1173D	ṅ	AHOM SIGN SECTION
1173E	ṅ	AHOM SIGN RULAI
1173F	ṅ	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;Mc;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;Mc;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 KILLER;Mn;0;NSM;;;;;N;;;;;

11730;AHOM DIGIT ZERO;No;0;L;;;0;0;N;;;;;
11731;AHOM DIGIT ONE;No;0;L;;;1;1;N;;;;;
11732;AHOM DIGIT TWO;No;0;L;;;2;2;N;;;;;
11733;AHOM DIGIT THREE;No;0;L;;;3;3;N;;;;;
11734;AHOM DIGIT FOUR;No;0;L;;;4;4;N;;;;;
11735;AHOM DIGIT FIVE;No;0;L;;;5;5;N;;;;;
11736;AHOM DIGIT SIX;No;0;L;;;6;6;N;;;;;
11737;AHOM DIGIT SEVEN;No;0;L;;;7;7;N;;;;;
11738;AHOM DIGIT EIGHT;No;0;L;;;8;8;N;;;;;
11739;AHOM DIGIT NINE;No;0;L;;;9;9;N;;;;;
1173A;AHOM NUMBER TEN;No;0;L;;;10;N;;;;;
1173B;AHOM NUMBER TWENTY;No;0;L;;;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

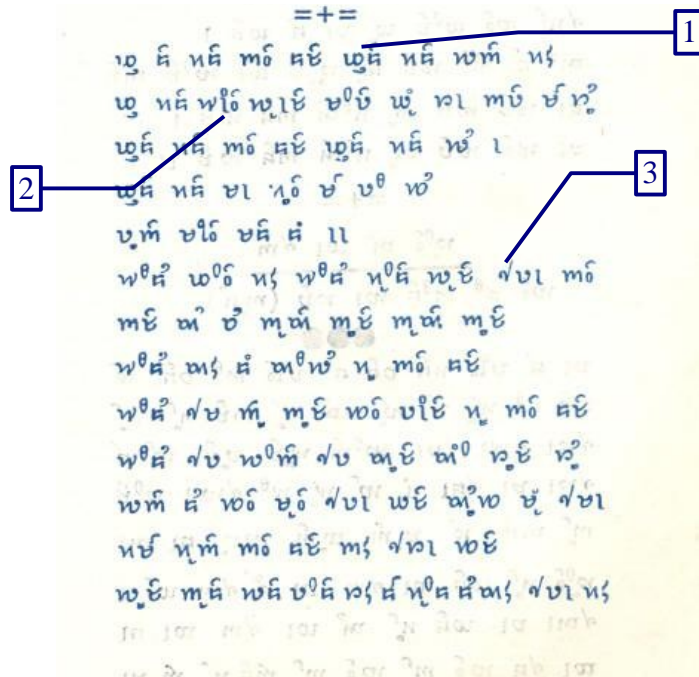


Figure 1: Lik Tai Khwam Tai page 7

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).

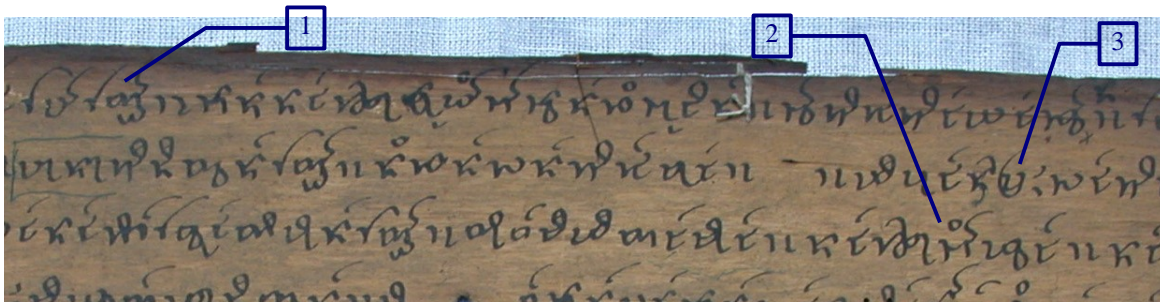


Figure 2: NemiMang p2v

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).

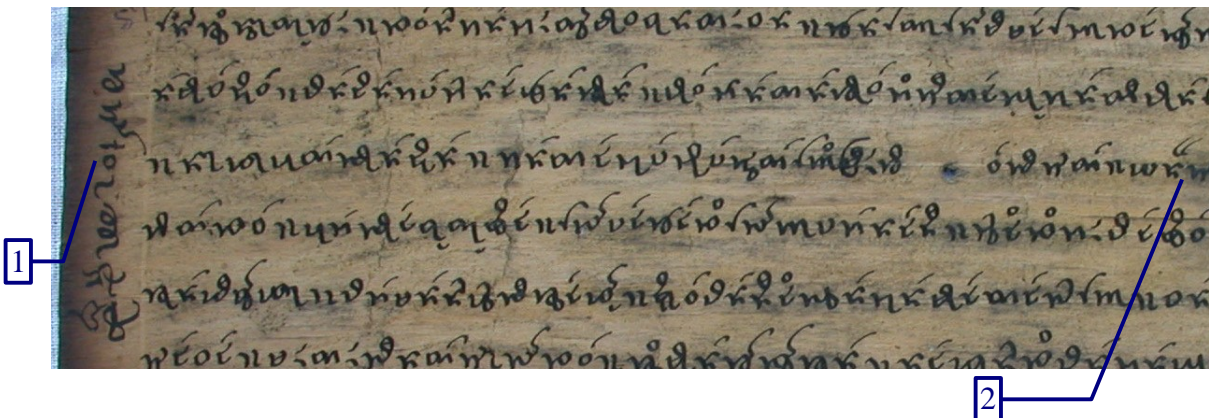


Figure 3: NemiMang p58v

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

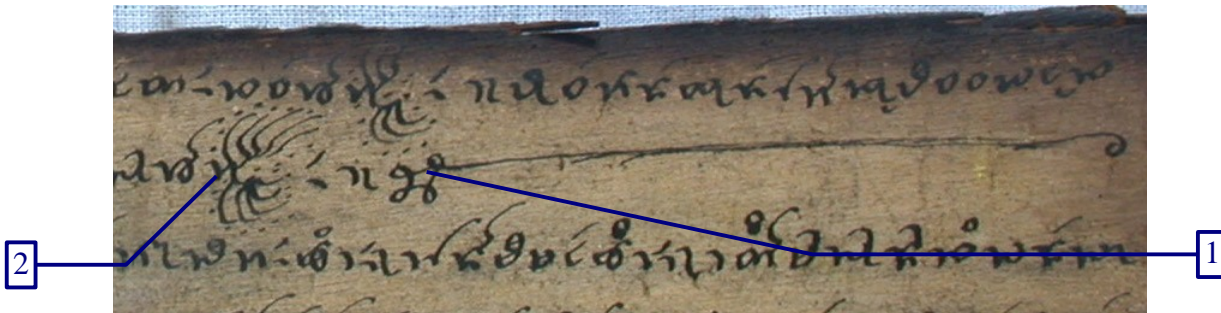


Figure 4: NemiMang p66r showing [1] text final embellishment, perhaps a character akin to TAI THAM SIGN KEOW (U+1AA3). This only occurs in the one text, so there is no intent to encode this within the Ahom block. Notice [2] the highly embellished /vi/ AHOM SYMBOL VI (U+1173F).

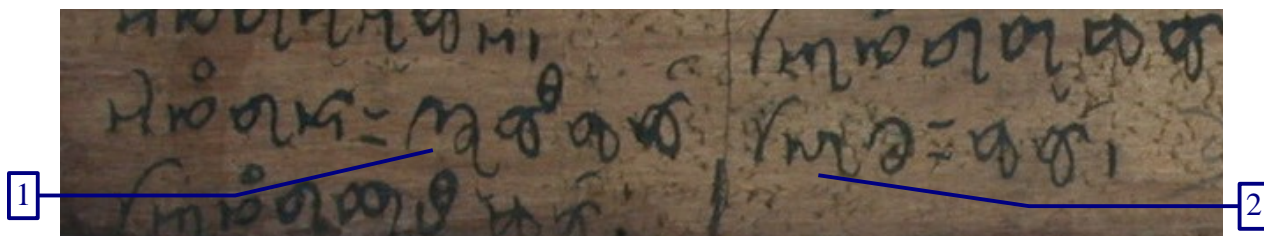


Figure 5: Phukan 1 p1v shows the two alternate forms of GA in the same document: 1) AHOM LETTER GA (U+11715) AHOM VOWEL SIGN U (U+11724), 2) AHOM LETTER ALTERNATE GA (U+11716)

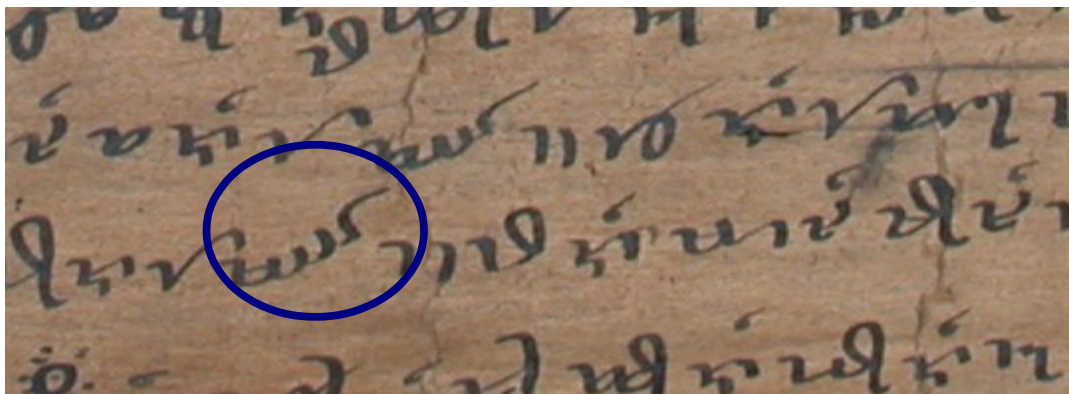


Figure 6: Mohan 9 p2r showing AHOM LETTER ALTERNATE TA (U+11705) AHOM LETTER JA (U+1170E) AHOM VOWEL SIGN AW (U+11725).

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 10646².

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:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/>
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"/>
E-Minor extinct <input type="checkbox"/>	F-Archaic Hieroglyphic or Ideographic <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:

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 <http://www.unicode.org> for such information on other scripts. Also see <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

² 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?	<i>yes</i>
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.)?	<i>yes</i>
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?	<i>yes</i>
Reference:	<i>this document</i>
4. The context of use for the proposed characters (type of use; common or rare)	<i>common</i>
Reference:	
5. Are the proposed characters in current use by the user community?	<i>yes</i>
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?	<i>no</i>
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)?	<i>yes</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>no</i>
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?	<i>no</i>
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?	<i>yes</i>
If YES, is a rationale for its inclusion provided?	<i>yes</i>
If YES, reference:	<i>this document</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	<i>yes</i>
If YES, is a rationale for such use provided?	<i>yes</i>
If YES, reference:	<i>this document</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	<i>no</i>
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	<i>no</i>
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility character(s)?	<i>no</i>
If YES, is the equivalent corresponding unified ideographic character(s) identified?	
If YES, reference:	