

Proposal to encode additional signs for the Tai Ahom script 29th September 2020

Prepared by Stephen Morey
La Trobe University
s.morey@latrobe.edu.au

This document proposes to add 7 consonants to the Ahom script, specifically to be able to write the characters created for the Pāli Tipiṭaka and published as part of the *World Tipiṭaka in Tai Scripts: 15-Script Set*. This book is so far the only known publication using these additional letters, and no widely available font includes them; however at the date of writing this application the Tai Ahom Community, as represented by the Society for Tai Ahom Resurgence (STAR) has expressed the strong view that these letters are needed.

A recent Thai publication, the Tai Ahom version of the *World Tipiṭaka in Tai Scripts: 15-Script Set* has created these 7 new consonant symbols in order to write the Pali scriptures in Tai Ahom script making an almost full set of consonants required to render Pāli (Pali) language with all phonemic distinctions marked. The one missing distinction in this document is a distinction between YA and JA (rendered in Thai script for Pali as ย and จ respectively), which are for reasons unknown are not distinguished in that publication, but rather both employ AHOM LETTER JA. We expect then, that to render Pali fully, an AHOM LETTER ALTERNATE JA or AHOM LETTER YA will be needed in the future.


The full set of Symbols used for consonants in the *World Tipiṭaka in Tai Scripts: 15-Script Set* are laid out in Figure 1:

ลักษณะ การเปล่ง เสียง	วักคา ปัลลปะปัลลจโส มันตา ¹ เสียงปาฬิ-พยัญชนะวักค มี 5 วักค วักคละ 5 ตัว					เสสา อวักคา พยัญชนะนา นาม ¹ เสียงปาฬิ-ชือพยัญชนะอวักค		
	อโยส ¹ ไม่ก้อง ²		โยส ¹ ก้อง ²		โยสโยส วิมุตติ ¹	โยส ¹ ก้อง ²		อโยส ¹ ไม่ก้อง ²
	ลิลิต ¹	ธนิต ¹	ลิลิต ¹	ธนิต ¹	นาสิกัญจานช ¹	ลิลิต ¹	อุสุม ¹	
	ไม่พ่นลม ²	พ่นลม ²	ไม่พ่นลม ²	พ่นลม ²	ช่องจุมุก	ไม่พ่นลม ²	เสียดแทรก ²	
กัณธช ¹ ที่เส้นเสียง ²							h ห หี่	
กัณธช ¹ ที่เพดานอ่อน ²	k ก	kh ข	g ค	gh ฌ	ṅ ง			
तालु ¹ ที่เพดานแข็ง ²	c จ	ch ฉ	j ช	jh ฌ	ñ ญ		y ย	
มूर्ध्नि ¹ ที่หลังปุ่มเหงือก ²	ṭ ฏ	ṭh ฐ	ḍ ฐ	ḍh ฑ	ṇ ณ	r ร	l ล	
तन्नि ¹ ที่ฟันบน ²	t ต	th ถ	d ด	dh ธ	n น			s ส
तन्नि ¹ ที่ลิ้นไก่ ²							v ว	
ओष्ठ ¹ ที่ริมฝีปาก ²	p ป	ph ฝ	b พ	bh ภ	m ม			
नासिकगुण ¹ ที่ช่องจมูก					ṁ ฌ	พระไตรปิฎกชาติพันธุ์ไท Tai Tipiṭaka ไตอาหม-ปาฬิ / Tai Āhom Script-Pāḷi		

Figure 1: Full set of Consonants used in Tai Ahom version of the *World Tipiṭaka in Tai Scripts: 15-Script Set* p. xxvii

The newly created letters are as follows:

AHOM LETTER CA


c
 based on AHOM LETTER SA but with an addition downward hook on the right hand side of the letter (See Figure 2)

AHOM LETTERS TTA, TTHA, DDA, DDHA

ṭ	ṭh	ḍ	ḍh
ṭ	ṭ	ḍ	ḍ

The retroflex stops; created from the Dental stops (AHOM LETTERS TA, THA, DA and DHA) with a dot in the left hand circle of each letter. This involves adding an additional circle for the retroflex ḍ (See Figures 2 and 3)

AHOM LETTER NNA

ṇ
 Retroflex ṇ, created by doubling the AHOM LETTER NA and turning the left hand into a mirror image of the original letter (See Figure 2)

AHOM LETTER LLA



Retroflex l, created by adding a black circle ‘dot’ to the right hand end of AHOM LETTER LA (See Appendix 1)

Note:



the letter described as va is also an addition by this publication to the set of letters in common use at the time the publication was prepared; the symbol used here is a formally similar to AHOM DIGIT ZERO, but a better choice would be AHOM LETTER ALTERNATE BA U+1171A which is already encoded.



Figure 2: Examples showing the lack of distinction between Pali /j/ and /y/ in line 1, in the Tai Ahom version of the World Tipiṭaka in Tai Scripts: 15-Script Set p. 26 (from the Vinayapitaka, Mahāvibhaṅga: ‘rules for monks’). Examples of CA, NNA, DDA are in line 8 and DDA and DDHA are in line 5

As mentioned earlier, the one phonemic distinction in Pali that is not marked here is the distinction between JA and YA, both of which employ AHOM LETTER JA. Figure 2 shows some examples of both these letters in Pali being rendered with the same symbol (YA in line 1, 4th word, YA in line 5th word).

Proposed AHOM LETTER LLA can be seen in the word Pāḷi in Appendix 1, p8 (title of the book) below.

	■-39	អ័រ	ក្រ	ប្រា	វេ រាណី ជា យា	យា ត្រា ប៊ិ រាណ តា	វិហារី ត្យា
		[at'a]	k'o	[b'aga-vā	[vē rañ jā yaŋ]	[ya-t'ā-b'i ran taŋ]	[viha ri tvā
អណ្តូង		ស៊ រេយ យា	សា កាស យា	កា ណុ កុ យា	យេ ណា	បា យា ហា បា តិ ត្រា ណា	
[anupa-gam ma]		[sō rēy yaŋ]	[saŋ kas saŋ]	[kaŋ ɲa kuj jaŋ]	[yē na]	[pa ya gapa tiṭ t'ā naŋ]	

Figure 3: Examples in the Tai Ahom version of the *World Tipiṭaka in Tai Scripts: 15-Script Set* p. 18 (from the Vinayapitaka, Mahāvibhaṅga: ‘rules for monks’) showing TTA and TTHA (2nd line)

Ahom community attitude to this proposal:

Apart from the publication of *World Tipiṭaka in Tai Scripts: 15-Script Set* there is not at tis stage additional evidence of the usage of the additional letters.

However, the addition of these 8 consonants has been broadly approved by the Tai Ahom community. On Sunday 23rd August 2020, an online meeting of members of the Society for Tai Ahom Resurgence (STAR) was held, and one of the leaders, Dr. Hemanta Gogoi Ngi Mon advised me that “Last evening we had an online meeting where most of the members advocated for the new letters”, adding that on Thursday 27th August they would have a further online meeting and “take a formal resolution”.

The formal resolution passed by the Society for Tai Ahom Resurgence is included as Appendix 2

	1170	1171	1172	1173	1174
0	ᱠ	ᱡ	ᱢ	ᱣ	ᱤ
1	ᱥ	ᱦ	ᱧ	ᱨ	ᱩ
2	ᱪ	ᱫ	ᱬ	ᱭ	ᱮ
3	ᱰ	ᱱ	ᱲ	ᱳ	ᱴ
4	ᱶ	ᱷ	ᱸ	ᱹ	ᱺ
5	ᱼ	ᱽ	᱾	᱿	᱀
6	᱁	᱂	᱃	᱄	᱅
7	᱆	᱇	᱈	᱉	
8	᱊	᱋	᱌	ᱍ	
9	ᱎ	ᱏ	᱐	᱑	
A	᱒	᱓	᱔	᱕	
B	᱖	YA	᱘	᱙	
C	ᱚ	FA		ᱛ	
D	ᱜ	ᱝ		ᱞ	
E	ᱟ	C		ᱠ	
F	ᱡ	ᱢ		ᱣ	

Consonants

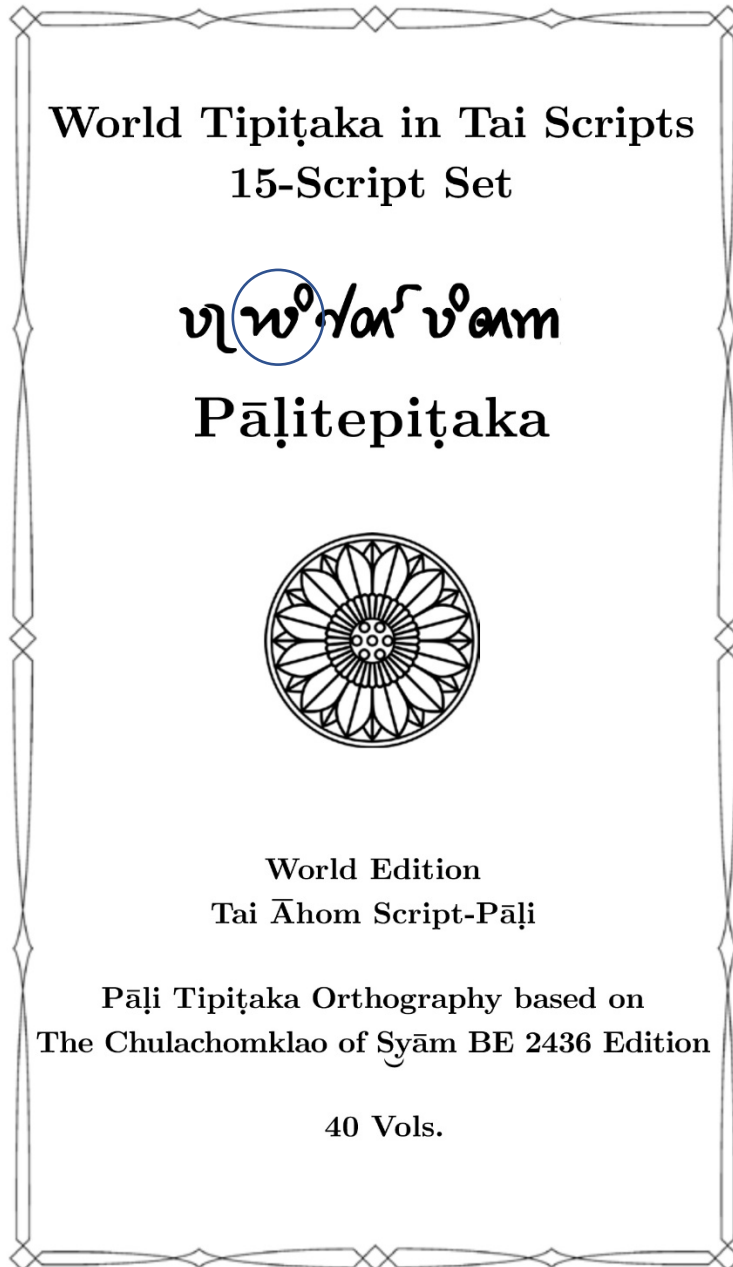
11740 AHOM LETTER CA
11741 AHOM LETTER TTA
11742 AHOM LETTER TTHA
11743 AHOM LETTER DDA
11744 AHOM LETTER DDHA
11745 AHOM LETTER NNA
11746 AHOM LETTER LLA

Unicode Properties

11740;AHOM LETTER CA ;Lo;0;L;;;;N;;;;;
11741;AHOM LETTER TTA ;Lo;0;L;;;;N;;;;;
11742;AHOM LETTER TTHA ;Lo;0;L;;;;N;;;;;
11743;AHOM LETTER DDA ;Lo;0;L;;;;N;;;;;
11744;AHOM LETTER DDHA ;Lo;0;L;;;;N;;;;;
11745;AHOM LETTER NNA ;Lo;0;L;;;;N;;;;;
11746;AHOM LETTER LLA ;Lo;0;L;;;;N;;;;;

Appendix 1:

Front cover of the Pāli Tepiṭaka in Ahom scripts



xx`

Figure 4: Title page of *World Tipiṭaka in Tai Scripts: 15-Script Set, Tai Ahom Version* showing LLA in the title (circled)



সাজযা
[saj · j^hā · ya]

তৈপিঠাক
[te · pi ṭa · ka]

The World Tipitaka Edition
Tai Ahom Script

Appendix 2, letter from the Society for Tai Ahom Resurgence (2020-08-28)



তুঁ নুঁ নুঁদি ৰ অ' তুঁ
SOCIETY FOR TAI AHOM RESURGENCE (STAR)

(Chom Hom Khuon Mau Tai Ahom)

Re. No. RS/SIVA/256/F/103

Headquarter: Balighat (Opposite Police Station), Simaluguri-785686, Assam, India.

Sivasagar, Assam, India. Email: resurgencetai@gmail.com website: taiahoms.ning.com



Resolution:

Further to the initial resolution dated 27/08/2020, the members of Society for Tai Ahom Resurgence (STAR) with other Ahom scholars resolved the following:

1. The proposed additional letters as used in the Tai Ahom Tipitaka by the World Tipitaka Foundation should be included in the Unicode standard. This refers specifically to the letters proposed by Dr. Stephen Morey in his proposal dated 21st September 2020, as

U+11740 AHOM LETTER CA;
U+11741 AHOM LETTER TTA;
U+11742 AHOM LETTER TTHA;
U+11743 AHOM LETTER DDA;
U+11744 AHOM LETTER DDHA;
U+11745 AHOM LETTER NNA;
U+11746 AHOM LETTER LLA.

i

2. STAR also resolved that we would work towards the addition of further consonants, vowels and tone marks as may be required for the revival of the Tai Ahom language.

Dr. Hemanta Kumar Gogoi
Chairman,
Society for Tai Ahom Resurgence (STAR)

Chairman
Society For Tai Ahom
Resurgence (STAR)

21.09/2020



Bibliography

- Hosken, Martin and Stephen Morey. 2012. 'Revised proposal to add the Ahom script in SMP of the UCS.' This is a technical document prepared for the Unicode Consortium. Downloaded from <http://www.unicode.org/L2/L2012/12309r-n4321r.pdf>, accessed 2015-02-18
- Morey, Stephen. 2005. *The Tai languages of Assam – a grammar and texts*. Canberra: Pacific Linguistics.

Acknowledgements

This proposal was inspired and supported by the members of the Tai Ahom community, in particular the Society for Tai Ahom Resurgence (STAR), and a number of individuals including Dr. Hemanta Ngi Mon Gogoi, Medini Madhab Mohan and Chao Medini Gogoi. I'm also very grateful to Poppy Gogoi whose careful graphemic analysis of Tai manuscripts has deepened our knowledge of the script. Much technical assistance has been given by Craig Cornelius who has helped me to understand the issues involved in the technical side and implementation of scripts, and also Zachary Scheuren who added the 7 new characters to the existing Noto Serif Ahom font at very short notice and with great skill.

Deborah Anderson has been very encouraging and helpful with comments and suggestions throughout the preparation of this document. Her strong support for the work on smaller scripts throughout the world is much valued and appreciated. Deborah's assistance has been made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Universal Scripts Project PR-253360-17 (part of the Script Encoding Initiative at UC Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment of the Humanities.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

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 – proposal to add additional characters</i>	
2. Requester's name:	<i>Stephen Morey</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>2020</i>	
5. Requester's reference (if applicable):	<i>?</i>	
6. Choose one of the following:		
This is a complete proposal:		<i>Yes</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>No</i>
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:		
Name of the existing block:	<i>Ahom</i>	
2. Number of characters in proposal:		<i>7</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	<i>?</i>	B.1-Specialized (small collection)
C-Major extinct		B.2-Specialized (large collection)
D-Attested extinct		E-Minor extinct
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols
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>an enhanced version of Noto Serif Ahom has been made by Zachary Scheuren</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>no</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.

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