

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

Doc Type: Working Group Document

Title: Proposal for encoding the Lanna script in the BMP of the UCS

Source: UC Berkeley Script Encoding Initiative (Universal Scripts Project)

Authors: Michael Everson and Martin Hosken

Status: Liaison Contribution

Replaces: N2042, N1013

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2006-09-09

1. Introduction. The Lanna script is used for three living languages: Northern Thai (that is, Kam Mu'ang), Tai Lue and Khün. In addition, the Lanna script is also used for Lao Tham (or old Lao) and other dialect variants in Buddhist palm leaves and notebooks. The script is also known as Tham or Yuan script. There are 6,000,000 speakers of Northern Thai of whom few are literate in Lanna script, although there is some resurgent interest in the script among the young. There are 670,000 speakers of Tai Lue of whom those born before 1950 are literate in Lanna script. The script has also continued to be taught in the monasteries. There are 120,000 speakers of Khün for which Lanna is the only script.

2. Consonants. Consonants have an inherent *-a* vowel sound. Most consonants have a combining subjoined form, but unlike most other Brahmic scripts, the subjoining of a consonant does not mean that the vowel of the previous consonant is killed. A subjoined consonant may be the first consonant of the following syllable. The encoding model for Lanna is more similar to the Khmer *coeng* model than to the usual *virama* model: the character LANNA SIGN SAKOT is entered before a consonant which is to take the subjoined form. A subjoined consonant may be attached to a dependent vowel sign. The table below shows the base consonants and the subjoined forms they take; it is organized according to the traditional Brahmic chart. High-tone consonants are marked with superscript ^H and low-tone consonants with superscript ^L. Note that not every low consonant has a single high-consonant equivalent. For instance, the corresponding partner to ᨧ LOW NGA is a sequence, ᨦ HA + SAKOT + ᨧ LOW NGA = ᨦᨧ *high nga*.

| | | | | | |
|---|---|--|--|---|---|
| k ^H ᨧ | | kh ^H ᨧᨦ | k ^L ᨧ | kh ^L ᨧᨦ | ng ^L ᨧ |
| c ^H ᨣ | | ch ^H ᨣᨦ | c ^L ᨣ | ch ^L ᨣᨦ | ny ^L ᨣᨦ |
| t ᨣ | | th ^H ᨣᨦ | d ᨣ | th ^L ᨣᨦ | n ᨣᨦ |
| t ^H ᨣᨦ | | th ^H ᨣᨦᨦ | t ^L ᨣᨦ | th ^L ᨣᨦᨦ | n ᨣᨦᨦ |
| b ᨣᨦ | p ^H ᨣᨦᨦ | ph ^H ᨣᨦᨦᨦ | p ^L ᨣᨦᨦ | ph ^L ᨣᨦᨦᨦ | m ᨣᨦᨦᨦ |
| y ^L ᨣᨦᨦ | | r ᨣᨦᨦᨦ | l ᨣᨦᨦᨦ | w ᨣᨦᨦᨦᨦ | ss ᨣᨦᨦᨦᨦᨦ |
| sh ^H ᨣᨦᨦᨦ | ss ^H ᨣᨦᨦᨦᨦᨦ | s ^H ᨣᨦᨦᨦᨦᨦ | h ^H ᨣᨦᨦᨦᨦᨦ | ll ᨣᨦᨦᨦᨦᨦᨦ | lae ᨣᨦᨦᨦᨦᨦᨦᨦ |

A number of Lanna characters did not traditionally take subjoined forms, but modern innovations such as that in ᨧᨦᨦᨦ *kālf* 'golf', ᨧᨦᨦᨦᨦ *krāf* 'graph', ᨧᨦᨦᨦᨦᨦᨦ *tāffī* 'toffee' suggest that fonts should make provision for subjoining behaviour for all of them but the historical vocalic r and l:

| | | | | | |
|--|---|---|--|---|--|
| kx ^H ᨧᨦᨦᨦᨦ | kx ^L ᨧᨦᨦᨦᨦᨦ | s ^L ᨣᨦᨦᨦᨦᨦ | s ^L ᨣᨦᨦᨦᨦᨦᨦ | rue ᨣᨦᨦᨦᨦᨦᨦᨦ | lue ᨣᨦᨦᨦᨦᨦᨦᨦᨦ |
| f ^H ᨣᨦᨦᨦᨦᨦ | f ^L ᨣᨦᨦᨦᨦᨦᨦ | y ^H ᨣᨦᨦᨦᨦᨦᨦ | h ^L ᨣᨦᨦᨦᨦᨦᨦᨦ | | |

In Northern Thai, the letters ᨧ᩠ᩅ LOW CA and ᨧ᩠ᩅ LOW SA are not used, being replaced by language-specific forms ᨧ᩠ᩅ NORTHERN THAI LOW CA and ᨧ᩠ᩅ NORTHERN THAI LOW SA.

Consonants may also be subjoined to digits, as in ᨧ᩠ᩅ ‘thrice’, which is ᨧ᩠ᩅ THAM DIGIT THREE + ᨧ᩠ᩅ SAKOT + ᨧ᩠ᩅ LOW TA.

A note should be made regarding ᨧ᩠ᩅ LETTER GREAT SA. Many occurrences of ᨧ᩠ᩅ SA followed by ᨧ᩠ᩅ SA are rendered using a ligature ᨧ᩠ᩅ GREAT SA (as in ᨧ᩠ᩅ *sawassadī* ‘hello’), but there are some words in which the ligature does not occur and a normal stack of ᨧ᩠ᩅ SA and subjoined ᨧ᩠ᩅ SA (that is, ᨧ᩠ᩅ) may be required. Therefore the ᨧ᩠ᩅ LETTER GREAT SA is encoded separately following the same model as that for the GREAT SA in Myanmar. In the unlikely event that GREAT SA should follow a SAKOT, the subscript form ᨧ᩠ᩅ would occur.

3. Independent vowels. Independent vowels are used as in other Brahmic scripts as well. Northern Thai LETTER A (marked with superscript ⁿ) differs from that used in Tai Lue and Khün. The LETTER OO vowel is not used in Northern Thai.

| | | | |
|-------|--------------------|-------|-------|
| a ᨧ᩠ᩅ | a ⁿ ᨧ᩠ᩅ | i ᨧ᩠ᩅ | ī ᨧ᩠ᩅ |
| u ᨧ᩠ᩅ | ū ᨧ᩠ᩅ | ē ᨧ᩠ᩅ | ō ᨧ᩠ᩅ |

4. Dependent consonant signs. Seven dependent consonant signs are used. Two of these are used as medials: ᨧ᩠ᩅ CONSONANT SIGN MEDIAL RA and ᨧ᩠ᩅ CONSONANT SIGN MEDIAL LA form clusters and immediately follow a consonant: ᨧ᩠ᩅ *kra*, ᨧ᩠ᩅ *kla*. ᨧ᩠ᩅ CONSONANT SIGN MAI KANG LAI is used as a final *-ng* in Northern Thai and Tai Lue: ᨧ᩠ᩅ *kang*. ᨧ᩠ᩅ CONSONANT SIGN KHUN MAI KANG LAI is used as a final *-ng* in Khün: ᨧ᩠ᩅ *kang*. ᨧ᩠ᩅ CONSONANT SIGN FINAL NGA is also used as a final *-ng* in Northern Thai: ᨧ᩠ᩅ *kang*. ᨧ᩠ᩅ CONSONANT SIGN LOW PA is used unusually in a Tai Lue word, ᨧ᩠ᩅ *kappa* ‘pregnant’; the normal rendering of ᨧ᩠ᩅ LOW PA + ᨧ᩠ᩅ SAKOT + ᨧ᩠ᩅ LOW PHA would be ᨧ᩠ᩅ. The last of these signs has two readings: ᨧ᩠ᩅ CONSONANT SIGN HIGH RATHA OR LOW PA represents ᨧ᩠ᩅ HIGH RATHA in ᨧ᩠ᩅ *santhān* ‘shape’ and in ᨧ᩠ᩅ *ratthabāl* ‘government’ (note the alternate spelling of this as ᨧ᩠ᩅ), and it represents ᨧ᩠ᩅ LOW PA in ᨧ᩠ᩅ *sappa* ‘omniscience’ and in ᨧ᩠ᩅ *ampa* ‘mango’. After ᨧ᩠ᩅ SAKOT the two base letters have their normal subjoined forms ᨧ᩠ᩅ and ᨧ᩠ᩅ; the former is rare, but attested as in ᨧ᩠ᩅ *sathāban rājabhaṭṭh chyañ hmai* ‘Chiang Mai College of Further Education’; the latter is also somewhat rare, as in ᨧ᩠ᩅ *banop burus* ‘disciple’. Special note should be taken that the letter U+1A30 ᨧ᩠ᩅ HIGH RATHA—whose glyph looks like U+1A35 ᨧ᩠ᩅ and U+1A5D ᨧ᩠ᩅ brought together—is not the same thing as ᨧ᩠ᩅ, which is U+1A2F ᨧ᩠ᩅ RATA + U+1A5D ᨧ᩠ᩅ CONSONANT SIGN HIGH RATHA OR LOW PA. This is because U+1A30 ᨧ᩠ᩅ (Brahmic *ṭha*) is not linguistically decomposable whereas U+1A2F ᨧ᩠ᩅ + U+1A5D ᨧ᩠ᩅ is (ᨧ᩠ᩅ Brahmic *ṭṭha*). Users need to be aware of the distinction. Khün does not use U+1A30 ᨧ᩠ᩅ, but uses the sequence U+1A2F + U+1A5D ᨧ᩠ᩅ (see Figure 12). In Lanna, however, ᨧ᩠ᩅ can be used as alternate *spelling* for U+1A30 ᨧ᩠ᩅ.

5. Dependent vowel signs. Dependent vowel signs are used in a manner similar to that employed by other Brahmic scripts, though Lanna makes use of a great many of them in combination. In one instance, a ligature is formed with a consonant: ᨧ᩠ᩅ NA + ᨧ᩠ᩅ AA = ᨧ᩠ᩅ *nā*. The vowels shown here are used in Northern Thai (the examples are taken from Udom Rungrueangsri’s 2004 dictionary; there are other orthographic conventions and other combinations doubtless occur):

| | |
|---------|--|
| ᨧ᩠ᩅ ko | = ᨧ᩠ᩅ ka ^H + ᨧ᩠ᩅ o |
| ᨧ᩠ᩅ kaṅ | = ᨧ᩠ᩅ ka ^H + ᨧ᩠ᩅ mai kang |
| ᨧ᩠ᩅ kan | = ᨧ᩠ᩅ ka ^H + ᨧ᩠ᩅ mai kang lai |

| | |
|----------|--|
| ကံ ka | = က ka ^H + နံ mai sat |
| ကွဲ kua? | = က ka ^H + နံ sakot + ဝ wa + ဝိ o + ဝံ a |
| ကွဲ kua | = က ka ^H + နံ sakot + ဝ wa + ဝိ o |
| ကွဲ kua | = က ka ^H + နံ sakot + ဝ wa |
| ကွဲ kō | = က ka ^H + ဝှော below + ဝိ mai kang |
| ကွဲ kō | = က ka ^H + ဝှော below |
| ကံ ka? | = က ka ^H + ဝံ a |
| ကံ kā | = က ka ^H + ဝံ aa |
| ကံ kī | = က ka ^H + ဝိ tall aa |
| ကံ kam | = က ka ^H + ဝံ am |
| ကံ kām | = က ka ^H + ဝိ tall am |
| ကံ ki | = က ka ^H + ဝိ i |
| ကံ kī | = က ka ^H + ဝိ ii |
| ကံ kue | = က ka ^H + ဝိ ue |
| ကံ kuue | = က ka ^H + ဝိ uue |
| ကံ ku | = က ka ^H + ဝှု u |
| ကံ kū | = က ka ^H + ဝှု uu |
| ကံ ke? | = က ka ^H + ဝဲ e + ဝံ a |
| ကံ kē | = က ka ^H + ဝဲ e |
| ကံ ke? | = က ka ^H + ဝဲ ae + ဝံ a |
| ကံ kē | = က ka ^H + ဝဲ ae |
| ကံ kə? | = က ka ^H + ဝဲ e + ဝှော below + ဝိ i + ဝံ a |
| ကံ kē | = က ka ^H + ဝဲ e + ဝှော below + ဝိ i (pronounced kia when followed by a final consonant) |
| ကံ kə? | = က ka ^H + ဝဲ e + ဝှော below + ဝိ uue + ဝံ a (used in Khün) |
| ကံ kē | = က ka ^H + ဝဲ e + ဝှော below + ဝိ uue (used in Khün) |
| ကံ ko | = က ka ^H + ဝဲ e + ဝံ mai sat + ဝံ aa |
| ကံ kō | = က ka ^H + ဝဲ oa above |
| ကံ kō | = က ka ^H + ဝဲ e + ဝံ aa |
| ကံ kia? | = က ka ^H + နံ sakot + ဃ ya ^L + ဝဲ e + ဝံ a |
| ကံ kia | = က ka ^H + နံ sakot + ဃ ya ^L + ဝဲ e |
| ကံ kia | = က ka ^H + နံ sakot + ဃ ya ^L |
| ကံ kia? | = က ka ^H + ဝဲ e + ဝှော below + ဝိ i + ဝဲ a ^m + ဝံ a |
| ကံ kia | = က ka ^H + ဝဲ e + ဝှော below + ဝိ i + ဝဲ a ^m |
| ကံ ko? | = က ka ^H + ဝဲ oo + ဝံ a |
| ကံ kō | = က ka ^H + ဝဲ oo |
| ကံ ko? | = က ka ^H + ဝဲ oo + ဝှော below + ဝံ a |
| ကံ kaj | = က ka ^H + နံ sakot + ဃ ya ^L + နံ mai sat |
| ကံ kaj | = က ka ^H + ဝဲ tham ai |
| ကံ kaj | = က ka ^H + ဝဲ ai |
| ကံ kaj | = က ka ^H + ဝဲ ai + နံ sakot + ဃ ya ^L |
| ကံ koy | = က ka ^H + ဝဲ oy (used in Khün) |

Both ဝံ AA and ဝိ TALL AA are encoded because context cannot be relied on to determine which one is used. The presence of ဝံ AM (and ဝိ TALL AM) follows the Thai convention of ensuring that a final consonant is not stored before the vowel it follows. This is the only situation in which it could occur and so AM is encoded to alleviate the problem. The Khün character ဝဲ oy is not used in Northern Thai. Khün vowel order is quite different from that of Northern Thai.

| | |
|----------|---|
| ကန ka? | = က ka ^H + န a |
| ကာ kā | = က ka ^H + ှ aa |
| ကျိ kā | = က ka ^H + ဿ tall aa |
| ကိ ki | = က ka ^H + ိ i |
| ကိ kī | = က ka ^H + ိ ii |
| ကု ku | = က ka ^H + ္ u |
| ကု kū | = က ka ^H + ္ uu |
| ကေ ke? | = က ka ^H + ြ e + န a |
| ကေ kē | = က ka ^H + ြ e |
| ကျ ကျိ | = က ka ^H + ျ sakot + ဃ ya ^L + ြ e |
| ကေ ke? | = က ka ^H + ြ ae + န a |
| ကေ kē | = က ka ^H + ြ ae |
| ကွေ kue | = က ka ^H + ိ ue |
| ကွေ kuue | = က ka ^H + ိ uue |
| ကေ ko? | = က ka ^H + ြ oo + န a |
| ကေ kō | = က ka ^H + ြ oo |
| ကာ kō | = က ka ^H + ြ e + ှ aa |
| ကွေ kua? | = က ka ^H + ျ sakot + ဝ wa + ိ o + န a |
| ကွေ kua | = က ka ^H + ျ sakot + ဝ wa + ိ o |
| ကေ ko? | = က ka ^H + ြ oo + ္ oa below + န a |
| ကေ ko | = က ka ^H + ံ oa above |
| ကေ ko? | = က ka ^H + ြ e + ္ oa below + ိ uue + န a |
| ကေ ko? | = က ka ^H + ြ e + ္ oa below + ိ uue |
| ကေ kaj | = က ka ^H + ြ ai |
| ကျ ကျိ | = က ka ^H + ျ ai + ျ sakot + ဃ ya ^L |
| ကေ ko | = က ka ^H + ြ oo + ံ mai sat |
| ကာ kaw | = က ka ^H + ြ e + ံ mai sat + ှ aa |
| ကေ kaṅ | = က ka ^H + ံ mai kang |
| ကေ kang | = က ka ^H + ံ final nga |
| ကေ ka | = က ka ^H + ံ mai sat + ျ sakot + ဎ nga |
| ကေ kiṅ | = က ka ^H + ိ i + ံ mai kang |
| ကေ king | = က ka ^H + ိ i + ျ sakot + ဎ nga |
| ကာ kam | = က ka ^H + ံ am |
| ကေ kam | = က ka ^H + ံ mai sat + ျ sakot + ဃ ma |
| ကေ kam | = က ka ^H + ံ mai sat + ဃ ma |
| ကေ kam | = က ka ^H + ံ mai sat + ဃ ma + ျ sakot + ဃ ma + ံ mai ra haam |
| ကျ ကို | = က ka ^H + ျ oy (used in Khün) |

6. Tone marks. Tone marks are combining characters. Lanna has two tone marks, ^{◌̇} SIGN TONE-1 and ^{◌̈} SIGN TONE-2, which are used in Lue and in Northern Thai, which are positioned over (and follow) the vowel over the base consonant. Three additional tone marks are used in Khün, ^{◌̆} SIGN KHUN TONE-3, ^{◌̇} SIGN KHUN TONE-4, and ^{◌̈} SIGN KHUN TONE-5, which are rendered above and to the right of the vowel over the base consonant. They are stored following the vowel over the base consonant or consonant stack. If there is no vowel over a base consonant, then the tone is rendered over the consonant—this is the same way tones are treated in the Thai script. In the case of ^{◌̈} VOWEL SIGN AM and ^{◌̆} VOWEL SIGN TALL AM, the tone is stored *before* the vowel sign, just as in the Thai script.

7. Other combining marks. ◌̣ SIGN RA HAAM is used in Northern Thai to indicate that the character or characters it follows are not sounded. The precise range of characters not to be sounded is not defined (that is, it's a reading rule), although it does not extend beyond one cluster. In Tai Lue, RA HAAM is used as a final *-n*. The mark ◌̣ SIGN MAI SAM has a range of uses in Northern Thai:

- It is used as a repetition mark, stored as the last character in the word to be repeated: တွဲ *tang* ‘be different’, တွဲတွဲ *tangtang* ‘be different in my view’.
- It is used to disambiguate the use of a subjoined letters. A subjoined letter may be a medial or final, or it may be the start of a new syllable. When the RA HAAM is used to indicate that a consonant begins a new syllable, it is stored following the subjoined form to indicate the consonant being at the start of a new syllable: ထွဲ *thanon* ‘path’ (without the RA HAAM, ထွဲ would be *thonra*).
- It is used to mark “double-acting” consonants. It is stored where the consonant would be stored if there were a separate consonant used, so ထွဲ *khawkhong* ‘belongings’ (without MAI SAM it is written ထွဲ).

The ◌̣ COMBINING CRYPTOGRAMMIC DOT is used singly or multiply beneath letters to give each letter a different value according to some hidden agreement between reader and writer. See Figure 6 for examples.

8. Digits. Two sets of digits are in common use, a “secular” set တွဲ၃၄၅၆၇၈၉ and an “ecclesiastical” set တွဲ၃၄၅၆၇၈၉. European digits are also found in books.

9. Punctuation. The four signs ၊ KAAN, ။ KAANKUU, ၌ SATKAAN, and ၍ SATKAANKUU, are used in a variety of ways, with progressive values of finality. It can be observed that the last of these is similar to U+0E5A ၎ THAI CHARACTER ANKHANKHU. The four punctuation characters are part of a patterned set in Lanna and SATKAANKUU will not be identical to ANKHANKHU in fonts, so it is best to encode the whole set for Lanna. At the end of a section, ။ KAANKUU and ၌ HANG may be combined with ၍ REVERSED ROTATED RANA in a number of ways: ၍၍, ၍၍၍, ၍၍၍၍, and ၍၍၍၍ may all occur. The symbols ၎ WIANG, ၏ WIANGWAAK, and ၐ SAWAN are logographs for ‘city’, ‘village’ (literally ‘open city’), and ‘heaven’ respectively. European punctuation like QUESTION MARK, EXCLAMATION MARK, PARENTHESES, and QUOTATION MARKS is also used. The three signs ၑ KEOW ‘courtyard’, ၒ HOY ‘oyster’, and ၓ DOKMAI ‘flower’ are used as dingbats and as section starters; ၑၑ and ၓၓ have been observed. The mark ၔ MAI YAMOK is used in the same way as its Thai counterpart, U+0E46 ၕ THAI CHARACTER MAIYAMOK.

10. Collating order. There is no firmly established sorting order for Lanna script. Each dictionary seems to have its own basic order that it aims to follow, though often inconsistently. There are various sort orders evidenced by the two lists given in the Dependent Vowels section. Even when an order can be established, it is not often one that is amenable to being expressed using the UCA. Therefore the order proposed here is merely as a consistent order that goes some way towards a possible sort for Lanna. This order is based on Northern Thai and Thai.

◌̣ *mai kang* U+1A76 < ◌̣ *a* U+1A61 < ◌̣ *ka^h* U+1A20 < ◌̣ *kha^h* U+1A21 < ◌̣ *kxa^h* U+1A22 < ◌̣ *ka^l* U+1A23 < ◌̣ *kxa^l* U+1A24 < ◌̣ *kha^l* U+1A25 < ◌̣ *nga^l* U+1A26 << ◌̣ *mai kang lai* U+1A59 << ◌̣ *khun mai kang lai* U+1A5A << ◌̣ *final nga* U+1A5B < ◌̣ *ca^h* U+1A27 < ◌̣ *cha^h* U+1A28 < ◌̣ *ca^l* U+1A29 <<< ◌̣ *N.T. ca^l* U+1A2A < ◌̣ *sa^l* U+1A2B <<< ◌̣ *N.T. sa^l* U+1A2C < ◌̣ *cha^l* U+1A2D < ◌̣ *nya^l* U+1A2E < ◌̣ *rata* U+1A2F < ◌̣ *ratha^h* U+1A30 < ◌̣ *da* U+1A31 < ◌̣ *ratha^l* U+1A32 < ◌̣ *rana* U+1A33 < ◌̣ *ta^h* U+1A34 < ◌̣ *tha^h* U+1A35 < ◌̣ *ta^l* U+1A36 < ◌̣ *tha^l* U+1A37 < ◌̣ *na^l* U+1A38 < ◌̣ *ba* U+1A39 < ◌̣ *pa^h* U+1A3A < ◌̣ *pha^h* U+1A3B < ◌̣ *fa^h* U+1A3C < ◌̣ *pa^l* U+1A3D << ◌̣ *pa^l* U+1A5C <<< ◌̣ *ratha^h* or *pa^l*

U+1A5D < ฝ *fa^l* U+1A3E < ๗ *pha^l* U+1A3F < ๗ *ma* U+1A40 < ๗ *ya^l* U+1A41 < ฝ *ya^h* U+1A42 < ๗ *ra*
 U+1A43 < ๗ *rue* U+1A44 < ๗ *la* U+1A45 < ๗ *lue* U+1A46 < ๗ *wa* U+1A47 < ๗ *sha^h* U+1A48 < ๗ *ssa^h*
 U+1A49 < ๗ *sa^h* U+1A4A < ๗ *ha^h* U+1A4B < ๗ *lla* U+1A4C < ๗ *a* U+1A4D <<< ๗ *N.T. a* U+1A4E <
 ๗ *ha^l* U+1A55 < ๗ *lae* U+1A56 < ๗ *o* U+1A6D < ฝ *medial ra* U+1A57 < ๗ *medial la* U+1A58 < [ฝ *sakot*
 U+1A60 ๗ *wa* U+1A47] < ๗ *oa below* U+1A6E < ๗ *mai sat* U+1A62 < ๗ *aa* U+1A63 <<< ๗ *tall aa*
 U+1A64 < ๗ *am* U+1A65 <<< ๗ *tall am* U+1A66 < ๗ *i* U+1A4F <<< ๗ *i* U+1A67 < ๗ *ii* U+1A50 <<<
 ๗ *ii* U+1A68 < ๗ *ue* U+1A69 < ๗ *uue* U+1A6A < ๗ *u* U+1A51 <<< ๗ *u* U+1A6B < ๗ *uu* U+1A52 <<< ๗ *uu*
 U+1A6C < ๗ *ee* U+1A53 <<< ๗ *e* U+1A70 < ๗ *oo* U+1A54 <<< ๗ *ae* U+1A71 < ๗ *oa above* U+1A75 <
 ๗ *oo* U+1A72 < ๗ *ai* U+1A73 < ๗ *tham ai* U+1A74

๗ *tone-1* U+1A77 << ๗ *tone-2* U+1A78 << ๗ *khun tone-3* U+1A79 << ๗ *khun tone-4* U+1A7A << ๗ *khun*
tone-5 U+1A7B << ๗ *ra haam* U+1A7C << ๗ *mai sam* U+1A7D << ๗ *cryptographic dot* U+1A7F

[ฝ *sakot* U+1A60 is ignored for sorting purposes.

11. Linebreaking. Opportunities for linebreaking are lexical, but a linebreak may not be inserted between a base letter and a combining diacritic. There is no line-breaking hyphenation.

12. Extensions. Extensions for the archaic version of Lanna used in Laos, called Lao Tham, may prove necessary in future. The user community for Lao Tham is extremely small, and this proposal meets the immediate requirements of the living communities which use Lanna script.

13. Unicode Character Properties. Combining classes for Lanna diacritics are all 0. Using other combining classes might work, but we do not favour this for two reasons. First, it will not do away with visual ambiguity, since such ambiguity is inherent in the script. Second, there are bound to be other interesting sequences which require odd combinations that this canonical ordering will probably break. Note in particular that SAKOT is not 9, because unlike other VIRAMA/COENG-type characters, SAKOT can follow vowels in Lanna, which VIRAMA/COENG doesn't do in any other script we know of.

```
1A20;LANNA LETTER HIGH KA;Lo;0;L;;;;N;;;;;
1A21;LANNA LETTER HIGH KHA;Lo;0;L;;;;N;;;;;
1A22;LANNA LETTER HIGH KXA;Lo;0;L;;;;N;;;;;
1A23;LANNA LETTER LOW KA;Lo;0;L;;;;N;;;;;
1A24;LANNA LETTER LOW KXA;Lo;0;L;;;;N;;;;;
1A25;LANNA LETTER LOW KHA;Lo;0;L;;;;N;;;;;
1A26;LANNA LETTER LOW NGA;Lo;0;L;;;;N;;;;;
1A27;LANNA LETTER HIGH CA;Lo;0;L;;;;N;;;;;
1A28;LANNA LETTER HIGH CHA;Lo;0;L;;;;N;;;;;
1A29;LANNA LETTER LOW CA;Lo;0;L;;;;N;;;;;
1A2A;LANNA LETTER NORTHERN THAI LOW CA;Lo;0;L;;;;N;;;;;
1A2B;LANNA LETTER LOW SA;Lo;0;L;;;;N;;;;;
1A2C;LANNA LETTER NORTHERN THAI LOW SA;Lo;0;L;;;;N;;;;;
1A2D;LANNA LETTER LOW SA;Lo;0;L;;;;N;;;;;
1A2E;LANNA LETTER LOW NYA;Lo;0;L;;;;N;;;;;
1A3F;LANNA LETTER RATA;Lo;0;L;;;;N;;;;;
1A30;LANNA LETTER HIGH RATHA;Lo;0;L;;;;N;;;;;
1A31;LANNA LETTER DA;Lo;0;L;;;;N;;;;;
1A32;LANNA LETTER LOW RATHA;Lo;0;L;;;;N;;;;;
1A33;LANNA LETTER RANA;Lo;0;L;;;;N;;;;;
1A34;LANNA LETTER HIGH TA;Lo;0;L;;;;N;;;;;
1A35;LANNA LETTER HIGH THA;Lo;0;L;;;;N;;;;;
1A36;LANNA LETTER LOW TA;Lo;0;L;;;;N;;;;;
1A37;LANNA LETTER LOW THA;Lo;0;L;;;;N;;;;;
1A38;LANNA LETTER NA;Lo;0;L;;;;N;;;;;
1A39;LANNA LETTER BA;Lo;0;L;;;;N;;;;;
1A3A;LANNA LETTER HIGH PA;Lo;0;L;;;;N;;;;;
1A3B;LANNA LETTER HIGH PHA;Lo;0;L;;;;N;;;;;
```

1A3C;LANNA LETTER HIGH FA;Lo;0;L;;;;N;;;;;
1A3D;LANNA LETTER LOW PA;Lo;0;L;;;;N;;;;;
1A3E;LANNA LETTER LOW FA;Lo;0;L;;;;N;;;;;
1A3F;LANNA LETTER LOW PHA;Lo;0;L;;;;N;;;;;
1A40;LANNA LETTER MA;Lo;0;L;;;;N;;;;;
1A41;LANNA LETTER LOW YA;Lo;0;L;;;;N;;;;;
1A42;LANNA LETTER HIGH YA;Lo;0;L;;;;N;;;;;
1A43;LANNA LETTER RA;Lo;0;L;;;;N;;;;;
1A44;LANNA LETTER RUE;Lo;0;L;;;;N;;;;;
1A45;LANNA LETTER LA;Lo;0;L;;;;N;;;;;
1A46;LANNA LETTER LUE;Lo;0;L;;;;N;;;;;
1A47;LANNA LETTER WA;Lo;0;L;;;;N;;;;;
1A48;LANNA LETTER HIGH SHA;Lo;0;L;;;;N;;;;;
1A49;LANNA LETTER HIGH SSA;Lo;0;L;;;;N;;;;;
1A4A;LANNA LETTER HIGH SA;Lo;0;L;;;;N;;;;;
1A4B;LANNA LETTER HIGH HA;Lo;0;L;;;;N;;;;;
1A4C;LANNA LETTER LLA;Lo;0;L;;;;N;;;;;
1A4D;LANNA LETTER A;Lo;0;L;;;;N;;;;;
1A4E;LANNA LETTER NORTHERN THAI A;Lo;0;L;;;;N;;;;;
1A4F;LANNA LETTER I;Lo;0;L;;;;N;;;;;
1A50;LANNA LETTER II;Lo;0;L;;;;N;;;;;
1A51;LANNA LETTER U;Lo;0;L;;;;N;;;;;
1A52;LANNA LETTER UU;Lo;0;L;;;;N;;;;;
1A53;LANNA LETTER EE;Lo;0;L;;;;N;;;;;
1A54;LANNA LETTER OO;Lo;0;L;;;;N;;;;;
1A55;LANNA LETTER LOW HA;Lo;0;L;;;;N;;;;;
1A56;LANNA LETTER LAE;Lo;0;L;;;;N;;;;;
1A57;LANNA CONSONANT SIGN MEDIAL RA;Mc;0;L;;;;N;;;;;
1A58;LANNA CONSONANT SIGN MEDIAL LA;Mn;0;NSM;;;;N;;;;;
1A59;LANNA SIGN MAI KANG LAI;Mn;0;NSM;;;;N;;;;;
1A5A;LANNA SIGN KHUN MAI KANG LAI;Mn;0;NSM;;;;N;;;;;
1A5B;LANNA CONSONANT SIGN FINAL NGA;Mn;0;NSM;;;;N;;;;;
1A5C;LANNA CONSONANT SIGN LOW PA;Mn;0;NSM;;;;N;;;;;
1A5D;LANNA CONSONANT SIGN HIGH RATHA OR LOW PA;Mn;0;NSM;;;;N;;;;;
1A5E;LANNA LETTER GREAT SA;Lo;0;L;;;;N;;;;;
1A60;LANNA SIGN SAKOT;Mn;0;NSM;;;;N;;;;;
1A61;LANNA VOWEL SIGN A;Mc;0;L;;;;N;;;;;
1A62;LANNA VOWEL SIGN MAI SAT;Mn;0;NSM;;;;N;;;;;
1A63;LANNA VOWEL SIGN AA;Mc;0;L;;;;N;;;;;
1A64;LANNA VOWEL SIGN TALL AA;Mc;0;L;;;;N;;;;;
1A65;LANNA VOWEL SIGN AM;Mc;0;L;;;;N;;;;;
1A66;LANNA VOWEL SIGN TALL AM;Mc;0;L;;;;N;;;;;
1A67;LANNA VOWEL SIGN I;Mn;0;NSM;;;;N;;;;;
1A68;LANNA VOWEL SIGN II;Mn;0;NSM;;;;N;;;;;
1A69;LANNA VOWEL SIGN UE;Mn;0;NSM;;;;N;;;;;
1A6A;LANNA VOWEL SIGN UUE;Mn;0;NSM;;;;N;;;;;
1A6B;LANNA VOWEL SIGN U;Mn;0;NSM;;;;N;;;;;
1A6C;LANNA VOWEL SIGN UU;Mn;0;NSM;;;;N;;;;;
1A6D;LANNA VOWEL SIGN O;Mn;0;NSM;;;;N;;;;;
1A6E;LANNA VOWEL SIGN OA BELOW;Mn;0;NSM;;;;N;;;;;
1A6F;LANNA VOWEL SIGN OY;Mc;0;L;;;;N;;;;;
1A70;LANNA VOWEL SIGN E;Mc;0;L;;;;N;;;;;
1A71;LANNA VOWEL SIGN AE;Mc;0;L;;;;N;;;;;
1A72;LANNA VOWEL SIGN OO;Mc;0;L;;;;N;;;;;
1A73;LANNA VOWEL SIGN AI;Mc;0;L;;;;N;;;;;
1A74;LANNA VOWEL SIGN THAM AI;Mc;0;L;;;;N;;;;;
1A75;LANNA VOWEL SIGN OA ABOVE;Mn;0;NSM;;;;N;;;;;
1A76;LANNA SIGN MAI KANG;Mn;0;NSM;;;;N;;;;;
1A77;LANNA SIGN TONE-1;Mn;0;NSM;;;;N;;;;;
1A78;LANNA SIGN TONE-2;Mn;0;NSM;;;;N;;;;;
1A79;LANNA SIGN KHUN TONE-3;Mn;0;NSM;;;;N;;;;;
1A7A;LANNA SIGN KHUN TONE-4;Mn;0;NSM;;;;N;;;;;
1A7B;LANNA SIGN KHUN TONE-5;Mn;0;NSM;;;;N;;;;;
1A7C;LANNA SIGN RA HAAM;Mn;0;NSM;;;;N;;;;;
1A7D;LANNA SIGN MAI SAM;Mn;0;NSM;;;;N;;;;;
1A7F;LANNA COMBINING CRYPTOGRAMMIC DOT;Mn;0;NSM;;;;N;;;;;
1A80;LANNA DIGIT ZERO;Nd;0;L;0;0;0;N;;;;;
1A81;LANNA DIGIT ONE;Nd;0;L;1;1;1;N;;;;;

1A82;LANNA DIGIT TWO;Nd;0;L;;2;2;2;N; ; ; ; ;
 1A83;LANNA DIGIT THREE;Nd;0;L;;3;3;3;N; ; ; ; ;
 1A84;LANNA DIGIT FOUR;Nd;0;L;;4;4;4;N; ; ; ; ;
 1A85;LANNA DIGIT FIVE;Nd;0;L;;5;5;5;N; ; ; ; ;
 1A86;LANNA DIGIT SIX;Nd;0;L;;6;6;6;N; ; ; ; ;
 1A87;LANNA DIGIT SEVEN;Nd;0;L;;7;7;7;N; ; ; ; ;
 1A88;LANNA DIGIT EIGHT;Nd;0;L;;8;8;8;N; ; ; ; ;
 1A89;LANNA DIGIT NINE;Nd;0;L;;9;9;9;N; ; ; ; ;
 1A90;LANNA THAM DIGIT ZERO;Nd;0;L;;0;0;0;N; ; ; ; ;
 1A91;LANNA THAM DIGIT ONE;Nd;0;L;;1;1;1;N; ; ; ; ;
 1A92;LANNA THAM DIGIT TWO;Nd;0;L;;2;2;2;N; ; ; ; ;
 1A93;LANNA THAM DIGIT THREE;Nd;0;L;;3;3;3;N; ; ; ; ;
 1A94;LANNA THAM DIGIT FOUR;Nd;0;L;;4;4;4;N; ; ; ; ;
 1A95;LANNA THAM DIGIT FIVE;Nd;0;L;;5;5;5;N; ; ; ; ;
 1A96;LANNA THAM DIGIT SIX;Nd;0;L;;6;6;6;N; ; ; ; ;
 1A97;LANNA THAM DIGIT SEVEN;Nd;0;L;;7;7;7;N; ; ; ; ;
 1A98;LANNA THAM DIGIT EIGHT;Nd;0;L;;8;8;8;N; ; ; ; ;
 1A99;LANNA THAM DIGIT NINE;Nd;0;L;;9;9;9;N; ; ; ; ;
 1AA0;LANNA SIGN WIANG;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA1;LANNA SIGN WIANGWAAK;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA2;LANNA SIGN SAWAN;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA3;LANNA SIGN KEOW;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA4;LANNA SIGN HOY;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA5;LANNA SIGN DOKMAI;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA6;LANNA SIGN REVERSED ROTATED RANA;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA7;LANNA SIGN MAI YAMOK;Lm;0;L;; ; ; ; ;N; ; ; ; ;
 1AA8;LANNA SIGN KAAAN;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AA9;LANNA SIGN KAANKUU;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AAA;LANNA SIGN SATKAAN;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AAB;LANNA SIGN SATKAANKUU;Po;0;L;; ; ; ; ;N; ; ; ; ;
 1AAC;LANNA SIGN HANG;Po;0;L;; ; ; ; ;N; ; ; ; ;

14. Bibliography.

Peltier, Anatole-Roger. 1996. *Lanna reader*. Chiang Mai: Wat Tha Kradas.

“ຕ້າເມືອງ” ISBN 974-9942-00-0

เอกสารคำสอน ราชวิชา ๑๕๔๒๓๐๓ อักษรไทย ๒ (อักษรล้านนา) (Lanna Alphabet II).

Udom Rungrueangsri. 2004. ລວດລຸ້ນລຽນ-ໄຊ ພັງເຂັ້ມວິບູນ (*Pacanānukrom Lānnā-Thai: Chabaph maefāhluang*). ISBN 974-685-175-9

ເບື້ງຮາວວາວລຽນ (*Baephyrar phāsā Lānnā*). ISBN 974-386-044-4

ວາວຊາເຂັ້ມລຽນ (*Phāsāmueng Lānnā*).

Acknowledgements

This project was made possible in part by a grant from the U.S. National Endowment for the Humanities, which funded the Script Encoding Initiative in respect of the Lanna encoding. Support for this project was also provided by Payap University, Chiang Mai, Thailand.

๕.๓.๑๔ คำกลุ่มพิเศษ

“คำกลุ่มพิเศษ” เป็นคำที่สร้างขึ้น มีความหมายเฉพาะ ตัวอย่างเช่น

| | | | | |
|---|-----|--------------------|--------------------|-----------|
| ว | คือ | กะใสไม้ขีด | อ่านว่า | ตัวก้อ |
| ว | คือ | ไม้กำใสก้อ | อ่านว่า | หลังตุ่น |
| อ | คือ | ตัววะใสไม้กี้ | อ่านว่า | ข้าง |
| อ | คือ | ตัวหะใสไม้กี้ | อ่านว่า | เหมี้ยง |
| อ | คือ | ตัววะซ้อนตัวระตะ | อ่านว่า | ตั้งรือ |
| ⊕ | คือ | ตัววะไขว่เกียง | อ่านว่า | เวียง |
| ⊕ | คือ | ตัววะหวากไขว่เกียง | อ่านว่า | เวียงหวาก |
| อ | | อ่านว่า | เข้าดอกดอกไม้ | |
| อ | | อ่านว่า | เข้าดอกดอกไม้เทียน | |
| ว | = | สะหวัน (สวรรณ) | | |
| ว | = | สะหรี | | |
| อ | = | ที่ | | |

Figure 5. Sample from a reader in Northern Thai, showing ⊕ SIGN WIANG, ⊕ SIGN WIANGWAAK, and ว SIGN SAWAN.

๓. ใช้พยัญชนะเป็นหลักแล้วใช้จุดวงกลมเล็กๆ เพิ่มจำนวนลงไปเพื่อ
กำหนดว่า หมายถึงอักษรใด เช่น

๐ = ๓ (ก) ๐ = (ข) ๐ = (ค) ฯลฯ

ตัวอย่าง ขขขขขขข = ๐๐๐ (วิฑูร)

[ข = ๐ (ก) ข = ๐ (ข) ข = ๐ (ค) ข = ๐ (ง) ข = ๐ (จ)]

Figure 6. Sample from a reader in Northern Thai, showing examples of ๐ COMBINING CRYPTOGRAMMIC DOT. The number of dots under a letter seem to indicate the number of places in a varga the reader should count to decode the intended letter.

| ตัวเลข | | | |
|--------------|------------|-----|--------|
| ภาษาล้านนา | | ไทย | อารบิก |
| ตัวเลขในธรรม | ตัวเลขโหรา | | |
| ๑ | 1 | ๑ | 1 |
| ๒ | 2 | ๒ | 2 |
| ๓ | 3 | ๓ | 3 |
| ๔ | 4 | ๔ | 4 |
| ๕ | 5 | ๕ | 5 |
| ๖ | 6 | ๖ | 6 |
| ๗ | 7 | ๗ | 7 |
| ๘ | 8 | ๘ | 8 |
| ๙ | 9 | ๙ | 9 |
| ๑๐ | 10 | ๑๐ | 10 |

ข้อสังเกต เลขศูนย์ของภาษาล้านนาเป็นรูปวงรีในแนวนอน หรือถ้าเป็นวงกลมจะมีขนาดเล็กกว่าตัวเลขตัวอื่น เพื่อป้องกันไม่ให้เขียนเหมือน "ตัววะ" ในขณะที่ เลขศูนย์อารบิก เป็นรูปวงรีในแนวตั้ง ส่วนเลขศูนย์ภาษาไทยกลางเป็นรูปวงกลม

Figure 7. Sample from a reader in Northern Thai, showing the Lanna digits alongside the Lanna Tham digits.

สระเดี่ยว-สระผสม

| | | |
|---------|--------|--------|
| - ๑ - ๒ | ๑ - ๒ | ๑ - ๒ |
| ๑ - ๓ | ๑ - ๓ | ๑ - ๓ |
| ๑ - ๔ | ๑ - ๔ | ๑ - ๔ |
| ๑ - ๕ | ๑ - ๕ | ๑ - ๕ |
| ๑ - ๖ | ๑ - ๖ | ๑ - ๖ |
| ๑ - ๗ | ๑ - ๗ | ๑ - ๗ |
| ๑ - ๘ | ๑ - ๘ | ๑ - ๘ |
| ๑ - ๙ | ๑ - ๙ | ๑ - ๙ |
| ๑ - ๑๐ | ๑ - ๑๐ | ๑ - ๑๐ |

Figure 8. Sample from a reader in Northern Thai, showing dependent vowels.

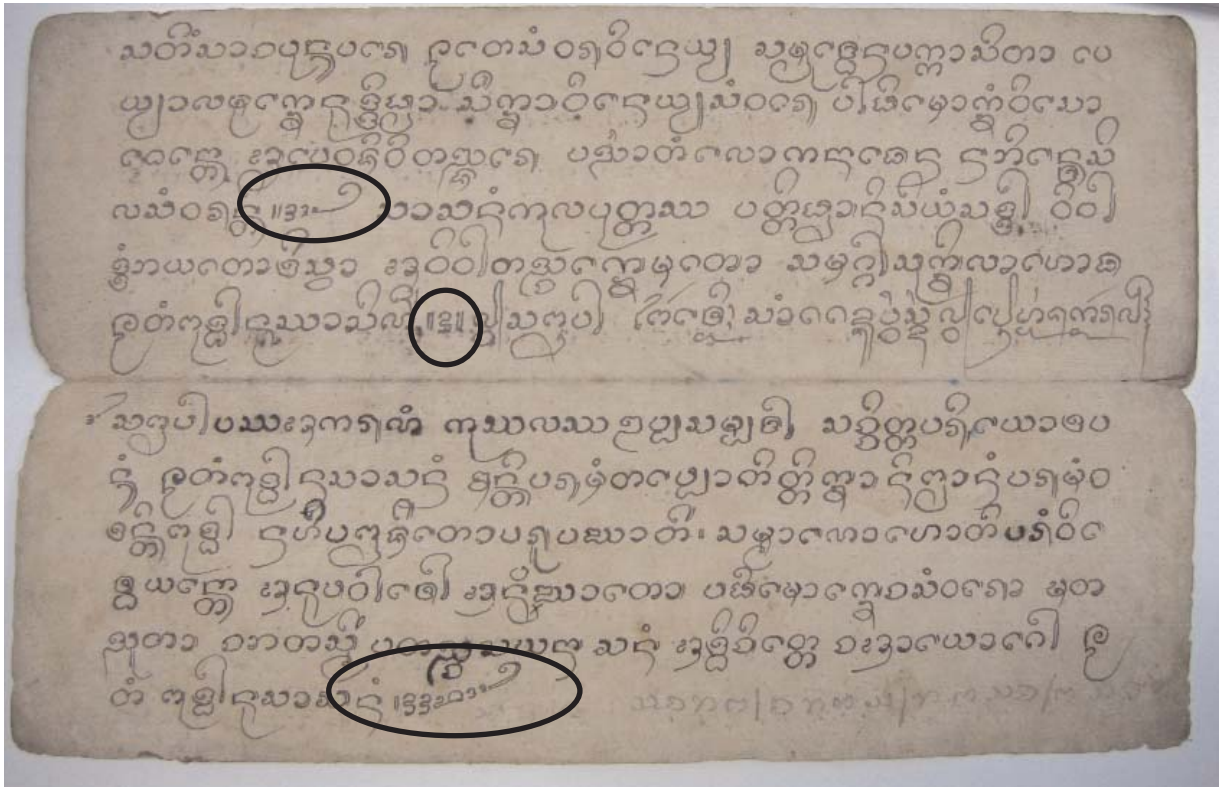
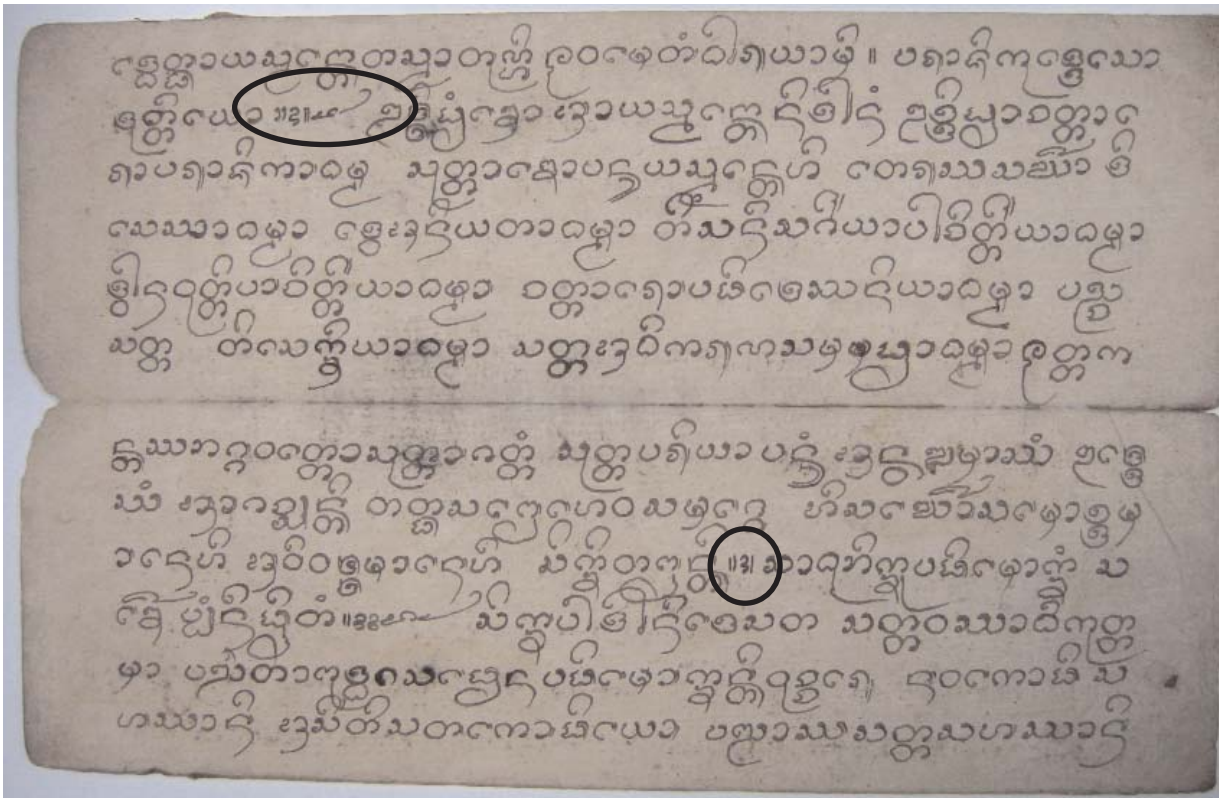


Figure 9a above and 9b below. Sample from a Northern Thai manuscript, showing a variety of marks built up out of individual punctuation marks: ||၃၁, ||၃၂, ||၃၃, ||၃၄, ||၃၅.



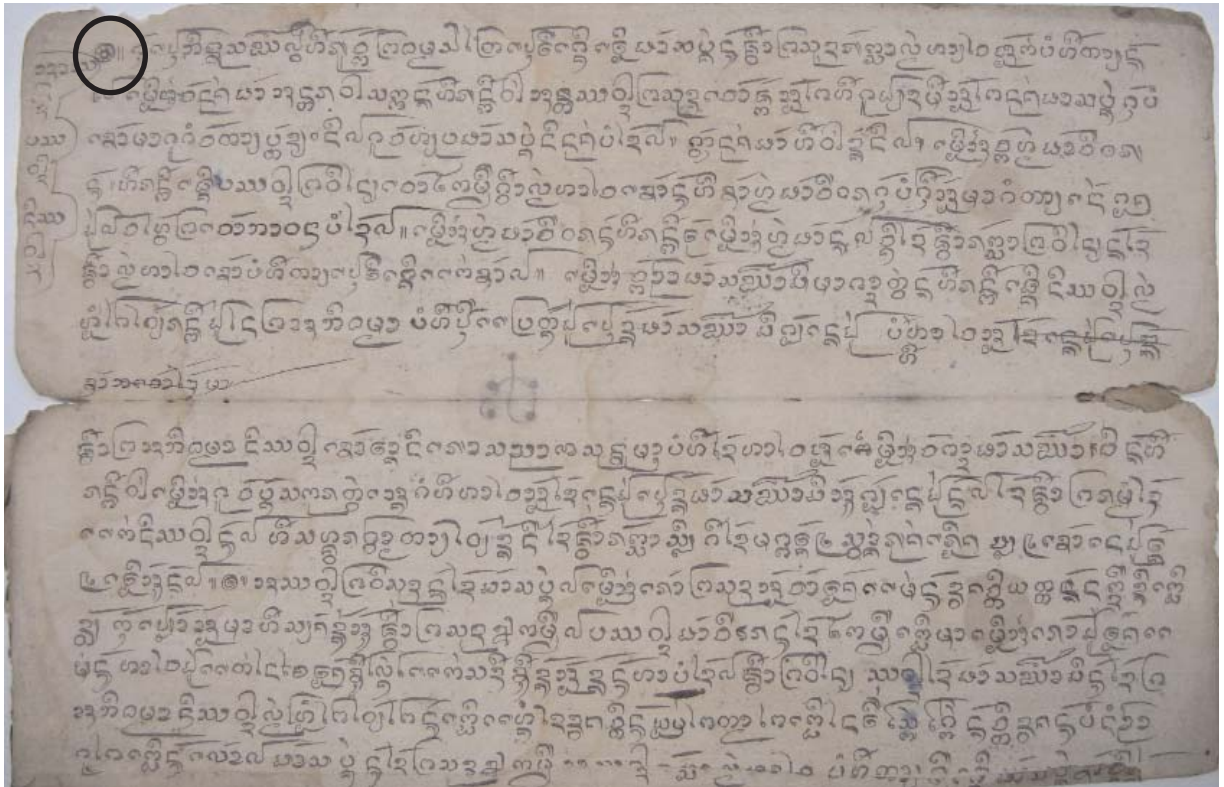
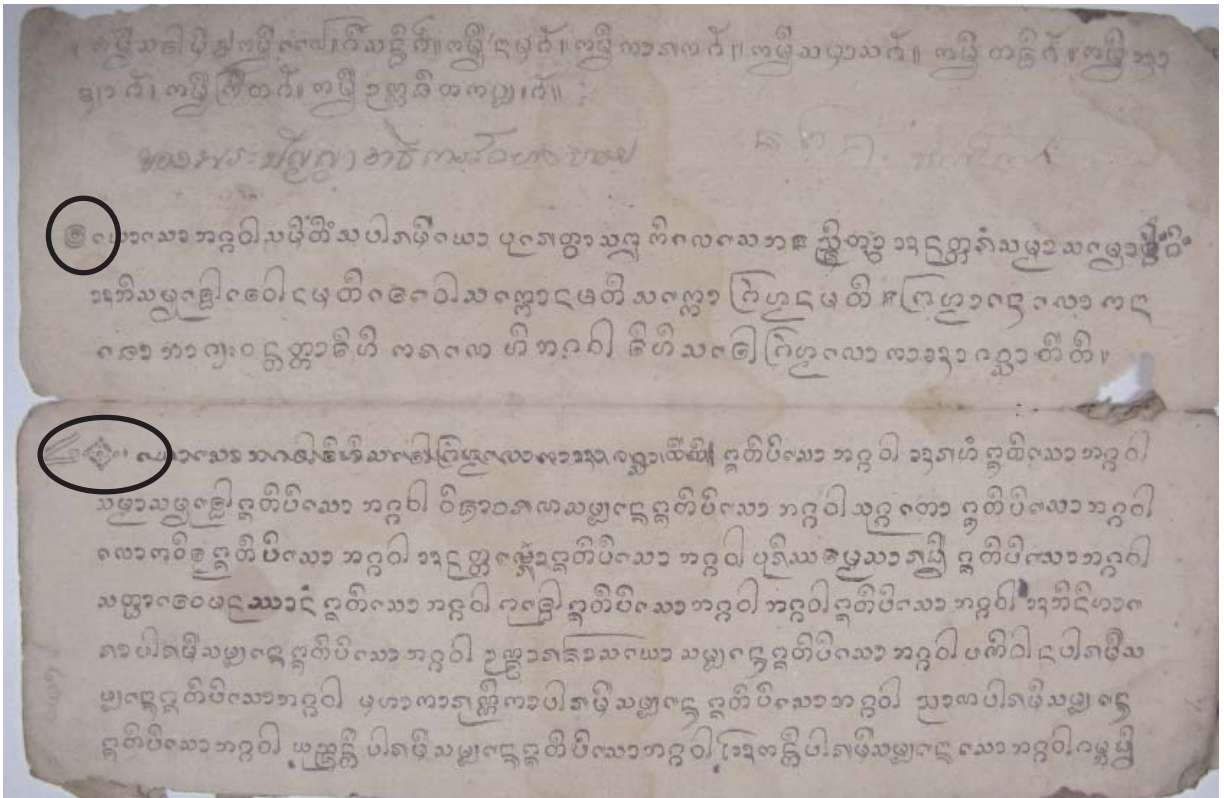


Figure 10a above and 10b below. Sample from a Northern Thai manuscript, showing a variety of marks built up out of individual punctuation marks: ๕๕, @, and ๕๕.



ဝေဝသမ္မိန္ဒရသုတ္တံ

၁. ရေသဉ္စရိ ခေမန္တ ဘဒ္ဒိယဒ္ဓိတိတံ ပေဂေါလုကတောဝဋ္ဌမ္မာ ပါလ္လေဖိဝိ ၄၀၀ သျှင်ကုဏ္ဍသံ သျှင်ကုဏ္ဍသံ ပျဗတဟူသု ယုကုဏ ခုဉ္စရိ(၂၀၀၀)လေ ခံပိဿဇပုညဇေတိဏ္ဍိတံဗုဒ္ဓိဉ္စရိခေမန္တ ဟိဇဒံရံဗုဒ္ဓိဉ္စရိဗုဒ္ဓိဉ္စရိသု ရိဗဒိဗုဒ္ဓိဉ္စရိသု သာဓု။
၂. ဝေဝဂ္ဂိဇေသုဂံဗေဟေ ဘဒ္ဒိယဒ္ဓိတိတံ ဝေဝဂ္ဂိဇေသု ပေဂေါလုကတောဝဋ္ဌမ္မာ ခုဉ္စရိ(၅၀၀)လေ ခံပိဿဇပုညဇေတိဏ္ဍိတံဗုဒ္ဓိဉ္စရိဗုဒ္ဓိဉ္စရိသု ရိဗဒိဗုဒ္ဓိဉ္စရိသု သာဓု။
၃. ရေဟူဂ္ဂကဉ္စ ဘဒ္ဒိယဒ္ဓိတိတံ ပေဂေါလုကတောဝဋ္ဌမ္မာ ပါသဉ္စရိ ဝုဗဒိဗုဒ္ဓိဉ္စရိ ခုဉ္စရိ(၅၀၀)လေ ခံပိဿဇပုညဇေတိဏ္ဍိတံဗုဒ္ဓိဉ္စရိဗုဒ္ဓိဉ္စရိသု ရိဗဒိဗုဒ္ဓိဉ္စရိသု သာဓု။

ဝိမ္ဗူကမိ ဝ ဝေဝဂ္ဂိ ၄ ဝေဟ ၁၄ ဝိ ၁၃၆၅ တံ။
(19.02.2004) ဝံသန္တာ ၃၀၀၀-လေ

Figure 12. Sample from a reader in Khün. Note that in the circled example, the sequence is 1A43 ၎ RA + 1A62 ◌ MAI SAT + 1A60 ◌ SAKOT + 1A2F ၎ RATA + 1A5D HIGH RATHA OR LOW PA ၎ = ၎; it is not 1A43 ၎ RA + 1A62 ◌ MAI SAT + 1A60 ◌ SAKOT + 1A30 ၎ HIGH RATHA (= *၎) since U+1A30 is not used in Khün.

TABLE XX - Row 1A: LANNA

| | 1A2 | 1A3 | 1A4 | 1A5 | 1A6 | 1A7 | 1A8 | 1A9 | 1AA |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 0 | ဝိ | ဖျ | ဖ | ဆွဲ | ☐ | ေ | ဝ | ဝ | ⊕ |
| 1 | ဖှ | ဖ | ဃ | ဣ | ံ | ေ | ၁ | ခ | ⊕ |
| 2 | ဖှ | ဃ | ဃ | ဣ | ံ | ေ | ၂ | ၄ | ဖှ |
| 3 | ဝ | ဝ | ၁ | ဣ | ံ | ေ | ၃ | ၂ | ⊕ |
| 4 | ဝါ | ဝ | ဝ | ဣ | ံ | ေ | ၄ | ဣ | ⊕ |
| 5 | ဃ | ဝ | ဝ | ဣ | ံ | ေ | ၅ | ဣ | ⊕ |
| 6 | ၁ | ဝ | ဝ | ဣ | ံ | ေ | ၆ | ဣ | ဝ |
| 7 | ဝ | ဝ | ဝ | ဣ | ံ | ေ | ၇ | ဃ | ဃ |
| 8 | ဃ | ဝ | ဝ | ဣ | ံ | ေ | ၈ | ဣ | ၂ |
| 9 | ၃ | ဝ | ဝ | ဣ | ံ | ေ | ၉ | ဝ | |
| A | ၁ | ဝ | ဝ | ဣ | ံ | ေ | | | ၃ |
| B | ၁ | ဝ | ဝ | ဣ | ံ | ေ | | | ၃ |
| C | ၁ | ဝ | ဝ | ဣ | ံ | ေ | | | ၁ |
| D | ဝ | ဝ | ဝ | ဣ | ံ | ေ | | | |
| E | ဣ | ဝါ | ဃ | ဃ | ံ | | | | |
| F | ဃ | ဝ | ဝ | | ံ | ေ | | | |

G = 00
P = 00

TABLE XX - Row 1A: LANNA

| hex | Name | hex | Name |
|-----|---|-----|-----------------------------------|
| 20 | LANNA LETTER HIGH KA | 79 | LANNA SIGN KHUN TONE-3 |
| 21 | LANNA LETTER HIGH KHA | 7A | LANNA SIGN KHUN TONE-4 |
| 22 | LANNA LETTER HIGH KXA | 7B | LANNA SIGN KHUN TONE-5 |
| 23 | LANNA LETTER LOW KA | 7C | LANNA SIGN RA HAAM |
| 24 | LANNA LETTER LOW KXA | 7D | LANNA SIGN MAI SAM |
| 25 | LANNA LETTER LOW KHA | 7E | (This position shall not be used) |
| 26 | LANNA LETTER LOW NGA | 7F | LANNA COMBINING CRYPTOGRAMMIC DOT |
| 27 | LANNA LETTER HIGH CA | 80 | LANNA DIGIT ZERO |
| 28 | LANNA LETTER HIGH CHA | 81 | LANNA DIGIT ONE |
| 29 | LANNA LETTER LOW CA | 82 | LANNA DIGIT TWO |
| 2A | LANNA LETTER NORTHERN THAI LOW CA | 83 | LANNA DIGIT THREE |
| 2B | LANNA LETTER LOW SA | 84 | LANNA DIGIT FOUR |
| 2C | LANNA LETTER NORTHERN THAI LOW SA | 85 | LANNA DIGIT FIVE |
| 2D | LANNA LETTER LOW CHA | 86 | LANNA DIGIT SIX |
| 2E | LANNA LETTER LOW NYA | 87 | LANNA DIGIT SEVEN |
| 2F | LANNA LETTER RATA | 88 | LANNA DIGIT EIGHT |
| 30 | LANNA LETTER HIGH RATHA | 89 | LANNA DIGIT NINE |
| 31 | LANNA LETTER DA | 8A | (This position shall not be used) |
| 32 | LANNA LETTER LOW RATHA | 8B | (This position shall not be used) |
| 33 | LANNA LETTER RANA | 8C | (This position shall not be used) |
| 34 | LANNA LETTER HIGH TA | 8D | (This position shall not be used) |
| 35 | LANNA LETTER HIGH THA | 8E | (This position shall not be used) |
| 36 | LANNA LETTER LOW TA | 8F | (This position shall not be used) |
| 37 | LANNA LETTER LOW THA | 90 | LANNA THAM DIGIT ZERO |
| 38 | LANNA LETTER NA | 91 | LANNA THAM DIGIT ONE |
| 39 | LANNA LETTER BA | 92 | LANNA THAM DIGIT TWO |
| 3A | LANNA LETTER HIGH PA | 93 | LANNA THAM DIGIT THREE |
| 3B | LANNA LETTER HIGH PHA | 94 | LANNA THAM DIGIT FOUR |
| 3C | LANNA LETTER HIGH FA | 95 | LANNA THAM DIGIT FIVE |
| 3D | LANNA LETTER LOW PA | 96 | LANNA THAM DIGIT SIX |
| 3E | LANNA LETTER LOW FA | 97 | LANNA THAM DIGIT SEVEN |
| 3F | LANNA LETTER LOW PHA | 98 | LANNA THAM DIGIT EIGHT |
| 40 | LANNA LETTER MA | 99 | LANNA THAM DIGIT NINE |
| 41 | LANNA LETTER LOW YA | 9A | (This position shall not be used) |
| 42 | LANNA LETTER HIGH YA | 9B | (This position shall not be used) |
| 43 | LANNA LETTER RA | 9C | (This position shall not be used) |
| 44 | LANNA LETTER RUE | 9D | (This position shall not be used) |
| 45 | LANNA LETTER LA | 9E | (This position shall not be used) |
| 46 | LANNA LETTER LUE | 9F | (This position shall not be used) |
| 47 | LANNA LETTER WA | A0 | LANNA SIGN WIANG |
| 48 | LANNA LETTER HIGH SHA | A1 | LANNA SIGN WIANGWAAK |
| 49 | LANNA LETTER HIGH SSA | A2 | LANNA SIGN SAWAN |
| 4A | LANNA LETTER HIGH SA | A3 | LANNA SIGN KEOW |
| 4B | LANNA LETTER HIGH HA | A4 | LANNA SIGN HOY |
| 4C | LANNA LETTER LLA | A5 | LANNA SIGN DOKMAI |
| 4D | LANNA LETTER A | A6 | LANNA SIGN REVERSED ROTATED LANA |
| 4E | LANNA LETTER NORTHERN THAI A | A7 | LANNA SIGN MAI YAMOK |
| 4F | LANNA LETTER I | A8 | LANNA SIGN KAAAN |
| 50 | LANNA LETTER II | A9 | LANNA SIGN KAANKUJ |
| 51 | LANNA LETTER U | AA | LANNA SIGN SATKAAN |
| 52 | LANNA LETTER UU | AB | LANNA SIGN SATKAANKUU |
| 53 | LANNA LETTER EE | AC | LANNA SIGN HANG |
| 54 | LANNA LETTER OO | AD | (This position shall not be used) |
| 55 | LANNA LETTER LOW HA | AE | (This position shall not be used) |
| 56 | LANNA LETTER LAE | AF | (This position shall not be used) |
| 57 | LANNA CONSONANT SIGN MEDIAL RA | | |
| 58 | LANNA CONSONANT SIGN MEDIAL LA | | |
| 59 | LANNA SIGN MAI KANG LAI | | |
| 5A | LANNA SIGN KHUN MAI KANG LAI | | |
| 5B | LANNA CONSONANT SIGN FINAL NGA | | |
| 5C | LANNA CONSONANT SIGN LOW PA | | |
| 5D | LANNA CONSONANT SIGN HIGH RATHA OR LOW PA | | |
| 5E | LANNA LETTER GREAT SA | | |
| 5F | (This position shall not be used) | | |
| 60 | LANNA SIGN SAKOT | | |
| 61 | LANNA VOWEL SIGN A | | |
| 62 | LANNA VOWEL SIGN MAI SAT | | |
| 63 | LANNA VOWEL SIGN AA | | |
| 64 | LANNA VOWEL SIGN TALL AA | | |
| 65 | LANNA VOWEL SIGN AM | | |
| 66 | LANNA VOWEL SIGN TALL AM | | |
| 67 | LANNA VOWEL SIGN I | | |
| 68 | LANNA VOWEL SIGN II | | |
| 69 | LANNA VOWEL SIGN UE | | |
| 6A | LANNA VOWEL SIGN UUE | | |
| 6B | LANNA VOWEL SIGN U | | |
| 6C | LANNA VOWEL SIGN UU | | |
| 6D | LANNA VOWEL SIGN O | | |
| 6E | LANNA VOWEL SIGN OA BELOW | | |
| 6F | LANNA VOWEL SIGN OY | | |
| 70 | LANNA VOWEL SIGN E | | |
| 71 | LANNA VOWEL SIGN AE | | |
| 72 | LANNA VOWEL SIGN OO | | |
| 73 | LANNA VOWEL SIGN AI | | |
| 74 | LANNA VOWEL SIGN THAM AI | | |
| 75 | LANNA VOWEL SIGN O | | |
| 76 | LANNA SIGN MAI KANG | | |
| 77 | LANNA SIGN TONE-1 | | |
| 78 | LANNA SIGN TONE-2 | | |

A. Administrative

1. Title

Proposal for encoding the Lanna script in the UCS.

2. Requester's name

UC Berkeley Script Encoding Initiative (Universal Scripts Project); authors: Michael Everson and Martin Hosken

3. Requester type (Member body/Liaison/Individual contribution)

Liaison contribution.

4. Submission date

2006-09-09

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

Yes.

1b. Proposed name of script

Lanna.

1c. The proposal is for addition of character(s) to an existing block

No.

1d. Name of the existing block

2. Number of characters in proposal

127

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3)

Level 2

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Lanna requires Level 2 implementation as other Brahmic scripts do.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

8. 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)?

Yes.

9. 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 Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

See above.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

Yes. See N2042 and N1013.

2a. 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.

2b. If YES, with whom?

A. Boonkit Wacharasat (Chiang Mai), A. Manop Tanyo (Chiang Mai)

2c. 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?

Lanna is used in eastern Myanmar, northern Thailand, and southern China.

4a. The context of use for the proposed characters (type of use; common or rare)

Used to write the Khün, Northern Thai, and Tai Lue languages, as well as Pali and Sanskrit.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

In eastern Myanmar, northern Thailand, and southern China.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes. Positions 1A20-1AAF are proposed.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Contemporary use and accordance with the Roadmap.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Yes, but only superficially so.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

Similarities with other related scripts are to be expected but disunification is as per normal.

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

Yes.

11c. If YES, reference

Brahmic vowels.

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?