

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: Individual Contribution

Replaces: N2042, N1013

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

Date: 2006-08-06

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.

k ^H က		kh ^H ဣ	g ^L ဂ	kh ^L ဃ	ng ^L ဣ
c ^H ဝ		s ^H သ	ch ^L ဃ	s ^L ဝ	ny ^H ဣ
t ဝ		th ^H ဣ	d ဝ	th ^L ဝ	n ဝ
t ^H ဝ		th ^H ဣ	t ^L ဝ	th ^L ဝ	n ဝ
b ဝ	p ^H ဣ	ph ^H ဣ	p ^L ဝ	ph ^L ဝ	m ဣ
ny ^L ဝ		r ဝ	l ဝ	w ဝ	
s ^H ဝ	ss ^H ဣ	ss ^H ဣ	h ^H ဝ		le ဝ

In Northern Thai, the letters ဃ LOW CHA and ဣ LOW SA are not used, being replaced by language-specific forms ဣ NORTHERN THAI LOW CHA and ဣ NORTHERN THAI LOW SA. A number of Lanna characters do not take subjoined forms and sequences of them following sakot do not occur:

x ^H ဣ	kh ^L ဣ	s ^L ဣ	s ^L ဣ	f ^H ဣ	f ^L ဣ
y ဣ		ru ဣ	lu ဣ	laa ဣ	h ^L ဣ

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

a သ	a ^m အ	i ဣ	ī ဣ
u ဠ	ū ဠ	ē ဣ	ō ဣ

4. Dependent consonant signs. Seven dependent consonant signs are used. Two of these are used as medials: ◌ CONSONANT SIGN MEDIAL RA and ◌ CONSONANT SIGN MEDIAL LA form clusters and immediately follow a consonant: ကြာ *kra*, ကြာ ◌ *kla*. ◌ CONSONANT SIGN MAI KANG LAI is used as a final *-ng* in Northern Thai and Tai Lue: ကံ ◌ *kang*. ◌ CONSONANT SIGN KHUN MAI KANG LAI is used as a final *-ng* in Khün: ကံ ◌ *kang*. ◌ CONSONANT SIGN FINAL NGA is also used as a final *-ng* in Northern Thai: ကံ ◌ *kang*. ◌ CONSONANT SIGN LOW PA is used unusually in a Tai Lue word, ကံ ◌ *kappa* ‘pregnant’; the normal rendering of ◌ LOW PA + ◌ SAKOT + ◌ LOW PHA would be ကံ ◌. The last of these signs has two readings: ◌ CONSONANT SIGN HIGH LATHA OR LOW PA represents ◌ HIGH LATHA in ◌ *santhān* ‘shape’ and ◌ LOW PA in ◌ *sappa* ‘omniscience’. After ◌ SAKOT the two base letters have their normal subjoined forms ◌ and ◌.

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. The vowels shown here are used in Northern Thai (the examples are taken from Udom Rungruangsri’s 2004 dictionary; there are other orthographic conventions and other combinations doubtless occur):

ကံ ko	= ကံ ka ^H + ◌ ong
ကံ kaṅ	= ကံ ka ^H + ◌ mai kang
ကံ kan	= ကံ ka ^H + ◌ mai kang lai
ကံ ka	= ကံ ka ^H + ◌ mai sat
ကံ kua?	= ကံ ka ^H + ◌ sakot + ◌ wa + ◌ ong + ◌ a
ကံ kua	= ကံ ka ^H + ◌ sakot + ◌ wa + ◌ ong
ကံ kua	= ကံ ka ^H + ◌ sakot + ◌ wa
ကံ kō	= ကံ ka ^H + ◌ oh + ◌ mai kang
ကံ kō	= ကံ ka ^H + ◌ oh
ကံ ka?	= ကံ ka ^H + ◌ a
ကံ kā	= ကံ ka ^H + ◌ aa
ကံ kā	= ကံ ka ^H + ◌ tall aa
ကံ kam	= ကံ ka ^H + ◌ am
ကံ kam	= ကံ ka ^H + ◌ tall am
ကံ ki	= ကံ ka ^H + ◌ i
ကံ kī	= ကံ ka ^H + ◌ ii
ကံ kue	= ကံ ka ^H + ◌ ue
ကံ kuue	= ကံ ka ^H + ◌ uue
ကံ ku	= ကံ ka ^H + ◌ u
ကံ kū	= ကံ ka ^H + ◌ uu
ကံ ke?	= ကံ ka ^H + ◌ e + ◌ a
ကံ kē	= ကံ ka ^H + ◌ e
ကံ ke?	= ကံ ka ^H + ◌ ee + ◌ a
ကံ kē	= ကံ ka ^H + ◌ ee
ကံ kə?	= ကံ ka ^H + ◌ e + ◌ oh + ◌ i + ◌ a
ကံ kē	= ကံ ka ^H + ◌ e + ◌ oh + ◌ i (pronounced <i>kia</i> when followed by a final consonant)

ค๓?	= ๓ ka ^H + ๓ e + ๓ oh + ๓ uee + ๓ a (used in Khün)
ค๓	= ๓ ka ^H + ๓ e + ๓ oh + ๓ uee (used in Khün)
คaw	= ๓ ka ^H + ๓ e + ๓ mai sat + ๓ aa
คo	= ๓ ka ^H + ๓ o
คo	= ๓ ka ^H + ๓ e + ๓ aa
ค๓?	= ๓ ka ^H + ๓ sakot + ๓ nya + ๓ e + ๓ a
ค๓	= ๓ ka ^H + ๓ sakot + ๓ nya + ๓ e
ค๓	= ๓ ka ^H + ๓ sakot + ๓ nya
ค๓?	= ๓ ka ^H + ๓ e + ๓ oh + ๓ i + ๓ a ^m + ๓ a
ค๓	= ๓ ka ^H + ๓ e + ๓ oh + ๓ i + ๓ a ^m
ค๓?	= ๓ ka ^H + ๓ oo + ๓ a
คo	= ๓ ka ^H + ๓ oo
ค๓?	= ๓ ka ^H + ๓ oo + ๓ oh + ๓ a
ค๓	= ๓ ka ^H + ๓ sakot + ๓ nya + ๓ mai sat
ค๓	= ๓ ka ^H + ๓ tham ai
ค๓	= ๓ ka ^H + ๓ ai
ค๓	= ๓ ka ^H + ๓ ai + ๓ sakot + ๓ nya
ค๓	= ๓ ka ^H + ๓ oy (used in Khün)

Both ๓ AA and ๓ TALL AA are encoded because context cannot be relied on to determine which one is used. The presence of ๓ AM (and ๓ 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 ๓ oy is not used in Northern Thai. Khün vowel order is quite different from that of Northern Thai.

ค๓	= ๓ ka ^H + ๓ a
คā	= ๓ ka ^H + ๓ aa
คā	= ๓ ka ^H + ๓ tall aa
คi	= ๓ ka ^H + ๓ i
คī	= ๓ ka ^H + ๓ ii
คu	= ๓ ka ^H + ๓ u
คū	= ๓ ka ^H + ๓ uu
ค๓?	= ๓ ka ^H + ๓ e + ๓ a
คē	= ๓ ka ^H + ๓ e
ค๓	= ๓ ka ^H + ๓ sakot + ๓ nya + ๓ e
ค๓?	= ๓ ka ^H + ๓ ee + ๓ a
คē	= ๓ ka ^H + ๓ ee
คue	= ๓ ka ^H + ๓ ue
คue	= ๓ ka ^H + ๓ uee
ค๓?	= ๓ ka ^H + ๓ oo + ๓ a
คo	= ๓ ka ^H + ๓ oo
คo	= ๓ ka ^H + ๓ e + ๓ aa
ค๓?	= ๓ ka ^H + ๓ sakot + ๓ wa + ๓ ong + ๓ a
ค๓	= ๓ ka ^H + ๓ sakot + ๓ wa + ๓ ong
ค๓?	= ๓ ka ^H + ๓ oo + ๓ oh + ๓ a
คo	= ๓ ka ^H + ๓ o
ค๓?	= ๓ ka ^H + ๓ e + ๓ oh + ๓ uee + ๓ a
ค๓	= ๓ ka ^H + ๓ e + ๓ oh + ๓ uee

ကေ kaj	= က ka ^H + ငေ ai
ကေ့ kaj	= က ka ^H + ငေ့ ai + န sakot + ဃ nya
ကေ့ ko	= က ka ^H + ငေ့ oo + င် mai sat
ကေဝ kaw	= က ka ^H + ငေ့ e + င် mai sat + ဝ aa
ကေံ kaṅ	= က ka ^H + င် mai kang
ကေံ kang	= က ka ^H + င် final nga
ကေံ ka	= က ka ^H + င် mai sat + န sakot + င nga
ကေိ kiṅ	= က ka ^H + ငိ i + င် mai kang
ကေိ king	= က ka ^H + ငိ i + န sakot + င nga
ကေံဝ kam	= က ka ^H + ဝ am
ကေံဝ kam	= က ka ^H + င် mai sat + န sakot + မ ma
ကေံမ kam	= က ka ^H + င် mai sat + မ ma
ကေံမ kam	= က ka ^H + င် mai sat + မ ma + န sakot + မ ma + င် mai lahaam
ကေ့ koy	= က ka ^H + ဝှ oy (used in Khün)

6. Tone marks. Tone marks are combining characters. Lanna has two tone marks, င် SIGN TONE-1 and င် 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, င် SIGN KHUN TONE-3, င် SIGN KHUN TONE-4, and င် 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 ဝ VOWEL SIGN AM and ဝှ VOWEL SIGN TALL AM, the tone is stored *before* the vowel sign, just as in the Thai script.

7. Other combining marks. င် SIGN LAHAAM 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, LAHAAM is used as a final *-n*. The mark င် 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: တေံ tang ‘be different’, တေံတေံ 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 LAHAAM 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: တေံန thanon ‘path’ (without the LAHAAM, တေံန 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 တေံန khawkhong ‘belongings’ (without MAI SAM it is written တေံန).

The င် 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 7 for examples.

8. Digits. Two sets of digits are in common use, a “secular” set ဝ၁၂၃၄၅၆၇၈၉ and an “ecclesiastical” set ဝ၀၁၂၃၄၅၆၇၈၉. European digits are also found in books.

9. Punctuation. The four signs ၊ GAAN, ။ GAANGU, ၎ SATGAAN, and ၏ SATGAANGUU, 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

๓ THAI CHARACTER ANKHANKHU. The four punctuation characters are part of a patterned set in Lanna and SATGAANGUU 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, ๓ GAANGUU and ๓ HANG may be combined with ๓ REVERSED ROTATED LANA in a number of ways: ๓๓, ๓๓, ๓๓๓, and ๓๓๓ may all occur. The symbols ๓ WIANGWAAK, ๓ WIANG, and ๓ SAWAN are logographs for ‘village’, ‘city’, and ‘heaven’ respectively. European punctuation like QUESTION MARK, EXCLAMATION MARK, PARENTHESES, and QUOTATION MARKS is also used. The three signs ๓ KEOW ‘courtyard’, ๓ HOY ‘oyster’, and ๓ DOKMAI ‘flower’ are used as dingbats and as section starters; ๓ and ๓ have been observed. The mark ๓ MAI YAMOK is used in the same way as its Thai counterpart, U+0E46 ๓ 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.

๓ *mai kang* U+1A76 < ๓ *a* U+1A61 < ๓ *ka*^h U+1A20 < ๓ *kha*^h U+1A21 < ๓ *xa*^h U+1A22 < ๓ *ga*^h U+1A23 < ๓ *kha*^l U+1A24 < ๓ *khaa*^l U+1A25 < ๓ *nga*^l U+1A26 << ๓ *mai kang lai* U+1A59 << ๓ *khun mai kang lai* U+1A5A << ๓ *final nga* U+1A5B < ๓ *ca*^h U+1A27 < ๓ *sa*^h U+1A28 < ๓ *cha*^l U+1A29 <<< ๓ *N.T. cha*^l U+1A2A < ๓ *sa*^l U+1A2B <<< ๓ *N.T. sa*^l U+1A2C < ๓ *saa*^l U+1A2D < ๓ *nya*^h U+1A2E < ๓ *lata* U+1A2F < ๓ *latha*^h U+1A30 < ๓ *lada* U+1A31 < ๓ *latha*^l U+1A32 < ๓ *lana* U+1A33 < ๓ *ta*^h U+1A34 < ๓ *tha*^h U+1A35 < ๓ *ta*^l U+1A36 < ๓ *tha*^l U+1A37 < ๓ *na*^l U+1A38 < ๓ *ba* U+1A39 < ๓ *pa*^h U+1A3A < ๓ *pha*^h U+1A3B < ๓ *fa*^h U+1A3C < ๓ *pa*^l U+1A3D << ๓ *pa*^l U+1A5C <<< ๓ *latha*^h or *pa*^l U+1A5D < ๓ *fa*^l U+1A3E < ๓ *pha*^l U+1A3F < ๓ *ma* U+1A40 < ๓ *nya*^l U+1A41 < ๓ *ya* U+1A42 < ๓ *ra* U+1A43 < ๓ *ru* U+1A44 < ๓ *la* U+1A45 < ๓ *lu* U+1A46 < ๓ *wa* U+1A47 < ๓ *saa*^h U+1A48 < ๓ *ssa*^h U+1A49 < ๓ *ssaa*^h U+1A4A < ๓ *ha*^h U+1A4B < ๓ *laa* U+1A4C < ๓ *a* U+1A4D <<< ๓ *N.T. a* U+1A4E < ๓ *ha*^l U+1A55 < ๓ *le* U+1A56 < ๓ *ong* U+1A6D < ๓ *medial ra* U+1A57 < ๓ *medial la* U+1A58 < [๓ *sakot* U+1A60 ๓ *wa* U+1A47] < ๓ *oh* U+1A6E < ๓ *mai sat* U+1A62 < ๓ *aa* U+1A63 <<< ๓ *tall aa* U+1A64 < ๓ *am* U+1A65 <<< ๓ *tall am* U+1A66 < ๓ *i* U+1A4F <<< ๓ *i* U+1A67 < ๓ *ii* U+1A50 <<< ๓ *ii* U+1A68 < ๓ *ue* U+1A69 < ๓ *uue* U+1A6A < ๓ *u* U+1A51 <<< ๓ *u* U+1A6B < ๓ *uu* U+1A52 <<< ๓ *uu* U+1A6C < ๓ *ee* U+1A53 <<< ๓ *e* U+1A70 < ๓ *oo* U+1A54 <<< ๓ *ee* U+1A71 < ๓ *o* U+1A75 < ๓ *oo* U+1A72 < ๓ *ai* U+1A73 < ๓ *tham ai* U+1A74

๓ *tone-1* U+1A77 << ๓ *tone-2* U+1A78 << ๓ *khun tone-3* U+1A79 << ๓ *khun tone-4* U+1A7A << ๓ *khun tone-5* U+1A7B << ๓ *lahaam* U+1A7C << ๓ *mai sam* U+1A7D << ๓ *cryptogrammic dot* U+1A7F

[๓ *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 XA;Lo;0;L;;;;N;;;;;
1A23;LANNA LETTER LOW GA;Lo;0;L;;;;N;;;;;
1A24;LANNA LETTER LOW KHA;Lo;0;L;;;;N;;;;;
1A25;LANNA LETTER LOW KHAA;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 SA;Lo;0;L;;;;N;;;;;
1A29;LANNA LETTER LOW CHA;Lo;0;L;;;;N;;;;;
1A2A;LANNA LETTER NORTHERN THAI LOW CHA;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 SAA;Lo;0;L;;;;N;;;;;
1A2E;LANNA LETTER HIGH NYA;Lo;0;L;;;;N;;;;;
1A3F;LANNA LETTER LATA;Lo;0;L;;;;N;;;;;
1A30;LANNA LETTER HIGH LATHA;Lo;0;L;;;;N;;;;;
1A31;LANNA LETTER LADA;Lo;0;L;;;;N;;;;;
1A32;LANNA LETTER LOW LATHA;Lo;0;L;;;;N;;;;;
1A33;LANNA LETTER LANA;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 NYA;Lo;0;L;;;;N;;;;;
1A42;LANNA LETTER YA;Lo;0;L;;;;N;;;;;
1A43;LANNA LETTER RA;Lo;0;L;;;;N;;;;;
1A44;LANNA LETTER RU;Lo;0;L;;;;N;;;;;
1A45;LANNA LETTER LA;Lo;0;L;;;;N;;;;;
1A46;LANNA LETTER LU;Lo;0;L;;;;N;;;;;
1A47;LANNA LETTER WA;Lo;0;L;;;;N;;;;;
1A48;LANNA LETTER HIGH SAA;Lo;0;L;;;;N;;;;;
1A49;LANNA LETTER HIGH SSA;Lo;0;L;;;;N;;;;;
1A4A;LANNA LETTER HIGH SSAA;Lo;0;L;;;;N;;;;;
1A4B;LANNA LETTER HIGH HA;Lo;0;L;;;;N;;;;;
1A4C;LANNA LETTER LAA;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 LE;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 LATHA OR LOW PA;Mn;0;NSM;;;;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 ONG;Mn;0;NSM; ; ; ; ;N; ; ; ; ;
1A6E;LANNA VOWEL SIGN OH;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 EE;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 O;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 LAHAAM;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 WIANGWAAK;Po;0;L; ; ; ; ;N; ; ; ; ;
1AA1;LANNA SIGN WIANG;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 LANA;Po;0;L; ; ; ; ;N; ; ; ; ;
1AA7;LANNA SIGN MAI YAMOK;Lm;0;L; ; ; ; ;N; ; ; ; ;
1AA8;LANNA SIGN GAAN;Po;0;L; ; ; ; ;N; ; ; ; ;
1AA9;LANNA SIGN GAANGUU;Po;0;L; ; ; ; ;N; ; ; ; ;
1AAA;LANNA SIGN SATGAAN;Po;0;L; ; ; ; ;N; ; ; ; ;
1AAB;LANNA SIGN SATGAANGUU;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.

“ต้าเมือง” ISBN 974-9942-00-0

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

Udom Rungrueangsri. 2004. ๑๐๑๑๑๑๑๑๑๑-๑๑๑๑ ๑๑๑๑๑๑๑๑๑๑ (Pacananukrom Lānnā-Thai: *Sababmēfāhluang*). ISBN 974-685-175-9

๑๑๑๑๑๑๑๑๑๑๑ (Bēbaryara bhāsā Lānnā). ISBN 974-386-044-4

๑๑๑๑๑๑๑๑๑๑๑ (Sāsāmeng 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	ตัวระซ้อนตัวระตะ	อ่านว่า	ตั้งรือ
อ	คือ	ตัววะไขว้เกียง	อ่านว่า	เวียง
อ	คือ	ตัววะหวกไขว้เกียง	อ่านว่า	เวียงหวก
อ			อ่านว่า	เข้าดอกดอกไม้
อ			อ่านว่า	เข้าดอกดอกไม้เทียน
อ	=			สะหวัน (สวรรณ)
อ	=			สะหรี
อ	=			ที่

Figure 5. Sample from a reader in Northern Thai, showing ⊕ SIGN WIANGWAAK, ⊕ SIGN WIANGWAAK, and ๐ SIGN SAWAN.

๓. ใช้พยัญชนะเป็นหลักแล้วใช้จุดวงกลมเล็กๆ เพิ่มจำนวนลงไปเพื่อ
กำหนดว่า หมายถึงอักษรใด เช่น

๐ = ๓ (ก) ๐ = (ข) ๐ = (ค) ฯลฯ

ตัวอย่าง ๐๐๐๐๐๐๐๐ = ๐๐๐๐ (วิฑูร)

[๐ = ๐ (ก) ๐ = ๐ (ข) ๐ = ๐ (ค) ๐ = ๐ (ง) ๐ = 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.

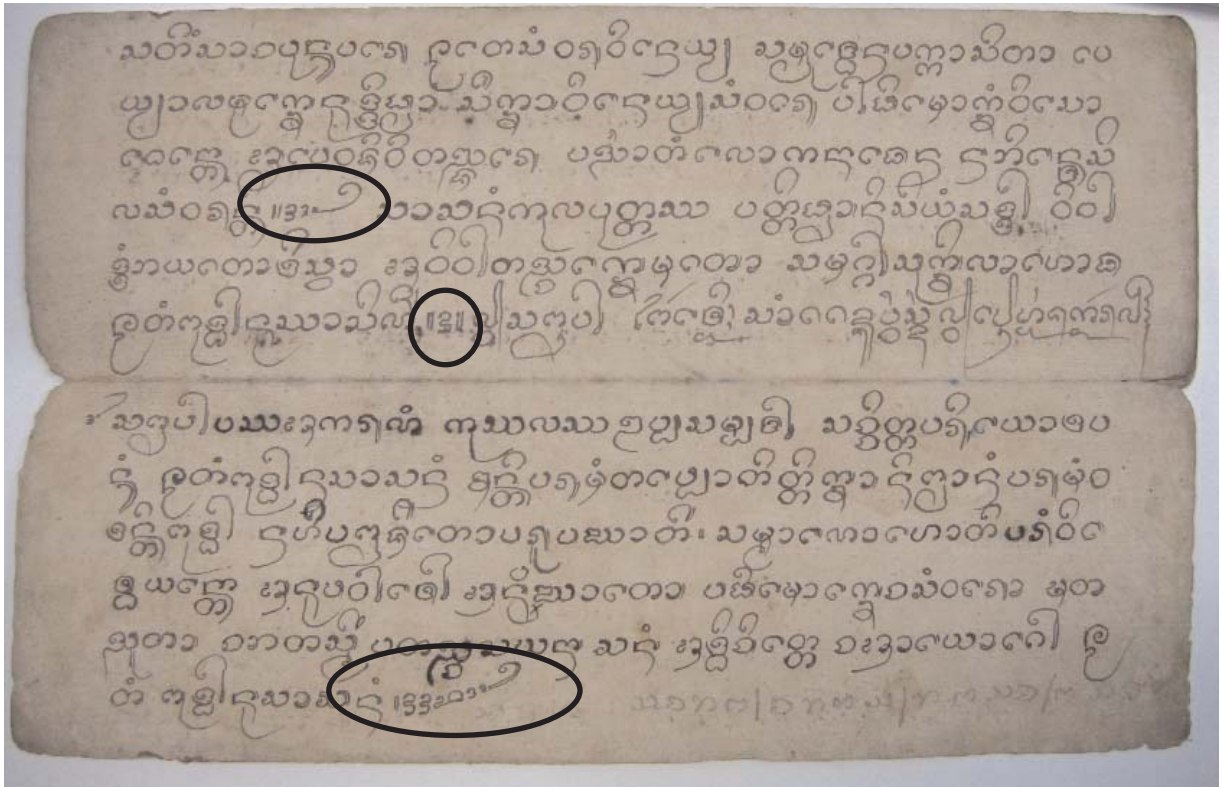
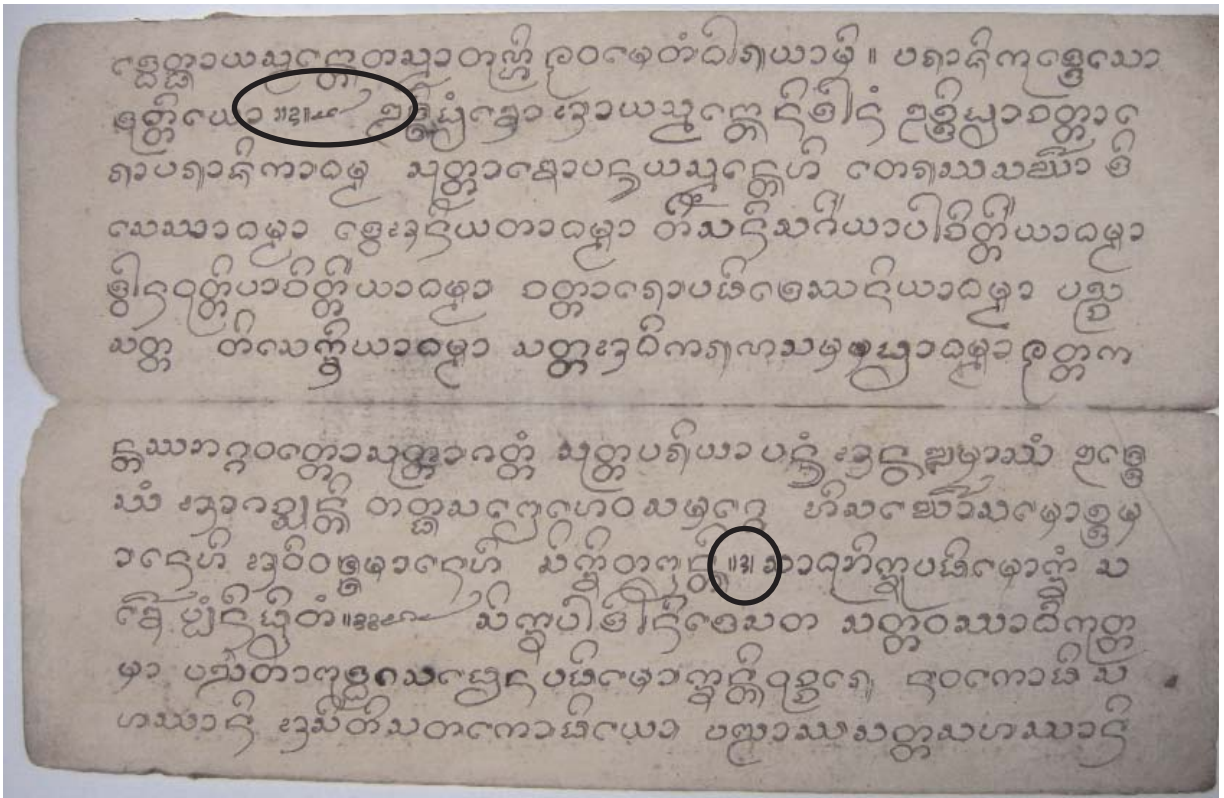


Figure 9a above and 9b below. Sample from a Northern Thai manuscript, showing a variety of marks built up out of individual punctuation marks: ||31, ||31||, ||33, ||31||, ||31.



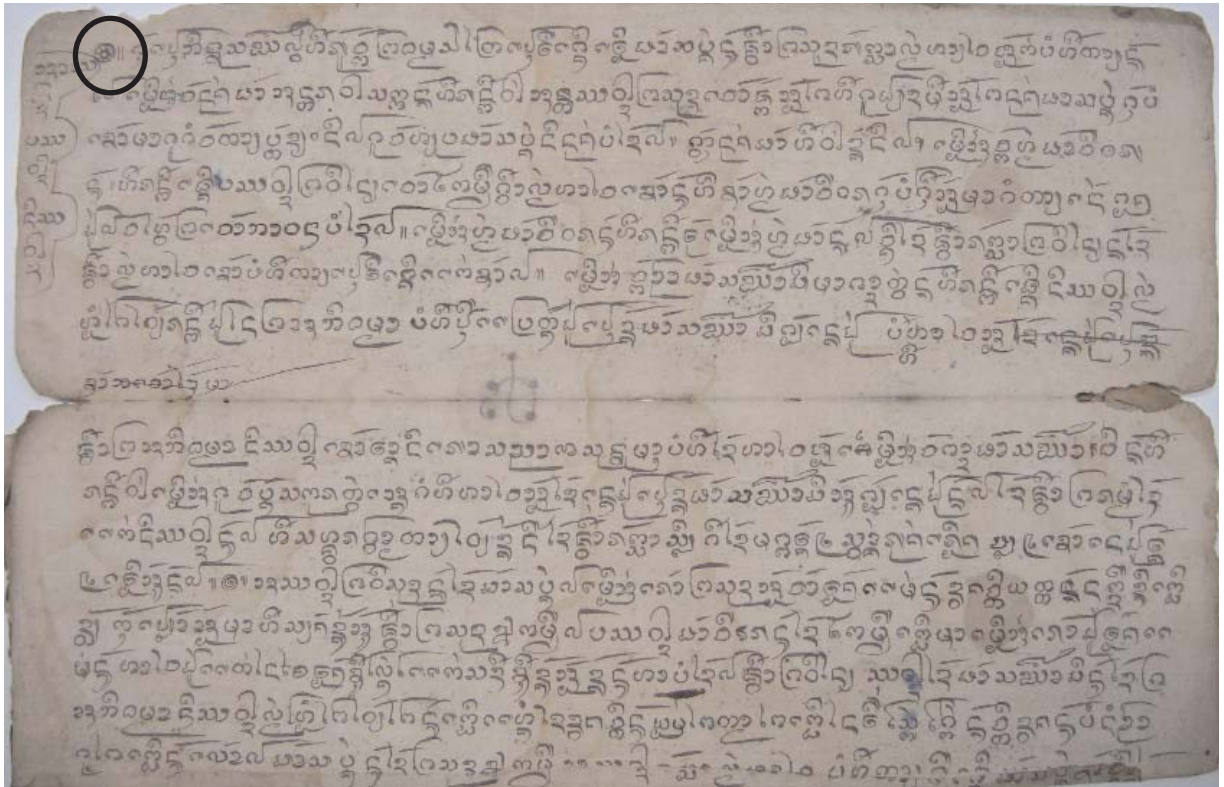


Figure 10a above and 10b below. Sample from a Northern Thai manuscript, showing a variety of marks built up out of individual punctuation marks: ๕๕, @, and ๕๕.

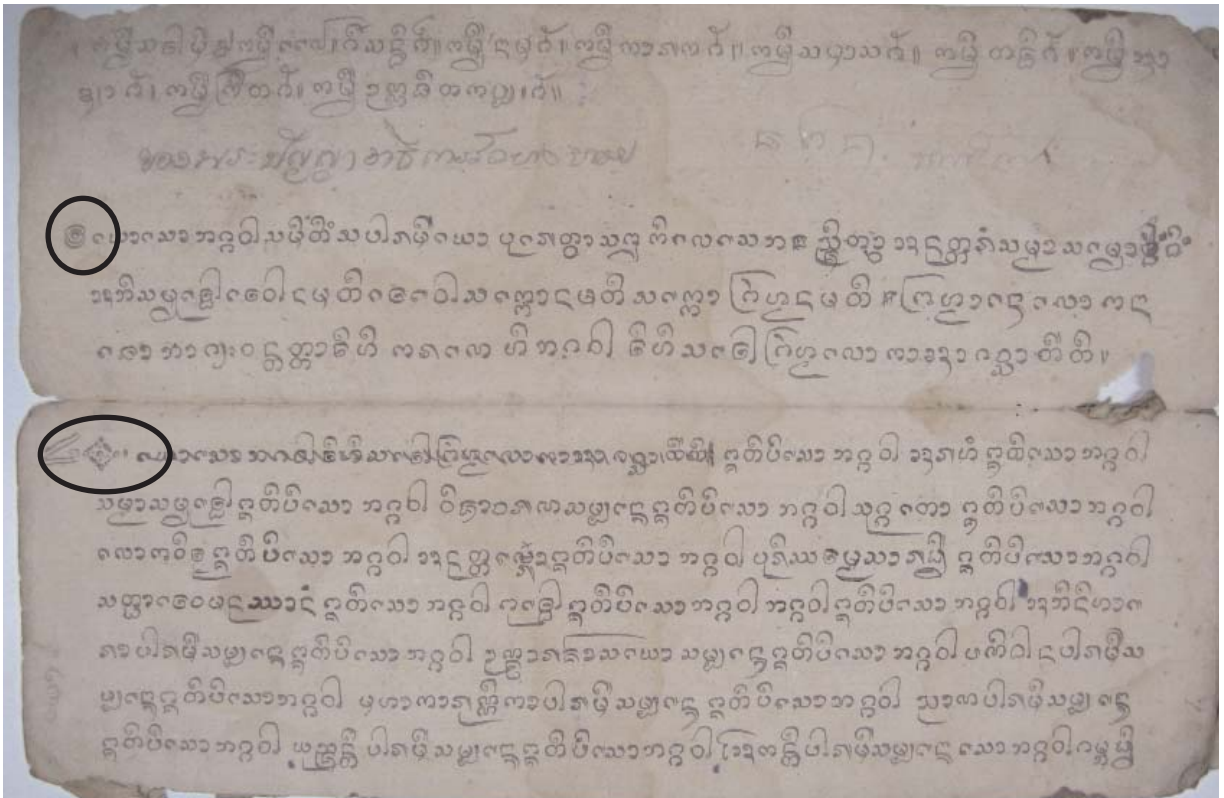


TABLE XX - Row 1A: LANNA

	1A2	1A3	1A4	1A5	1A6	1A7	1A8	1A9	1AA
0	ဝ	ဃ	ဖ	ဆ	☐	ေ	ဝ	ဝ	⊕
1	တ	ဃ	ဃ	ဃ	ံ	ေ	၁	ခ	⊕
2	တ	ဃ	ဃ	ဃ	ံ	ေ	၂	၄	မ
3	ဝ	ဝ	၁	ဃ	ံ	ေ	၃	၂	⊕
4	ဝ	ဝ	၂	ဃ	ံ	ေ	၄	၃	⊕
5	ဃ	ဝ	ဝ	ဃ	ံ	ေ	၅	၄	⊕
6	မ	ဝ	၃	ဃ	ံ	ေ	၆	၅	⊕
7	ဝ	ဝ	ဝ	ဃ	ံ	ေ	၇	၆	⊕
8	ဃ	ဝ	၄	ဃ	ံ	ေ	၈	၇	⊕
9	မ	ဝ	၅	ဃ	ံ	ေ	၉	၈	⊕
A	မ	ဝ	၆	ဃ	ံ	ေ			၉
B	မ	ဝ	၇	ဃ	ံ	ေ			၉
C	မ	ဝ	၈	ဃ	ံ	ေ			၉
D	မ	ဝ	၉	ဃ	ံ	ေ			
E	မ	ဝ		ဃ	ံ	ေ			
F	မ	ဝ		ဃ	ံ	ေ			

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 XA	7B	LANNA SIGN KHUN TONE-5
23	LANNA LETTER LOW GA	7C	LANNA SIGN LAHAAM
24	LANNA LETTER LOW KHA	7D	LANNA SIGN MAI SAM
25	LANNA LETTER LOW KHAA	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 SA	81	LANNA DIGIT ONE
29	LANNA LETTER LOW CHA	82	LANNA DIGIT TWO
2A	LANNA LETTER NORTHERN THAI LOW CHA	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 SAA	86	LANNA DIGIT SIX
2E	LANNA LETTER HIGH NYA	87	LANNA DIGIT SEVEN
2F	LANNA LETTER LATA	88	LANNA DIGIT EIGHT
30	LANNA LETTER HIGH LATHA	89	LANNA DIGIT NINE
31	LANNA LETTER LADA	8A	(This position shall not be used)
32	LANNA LETTER LOW LATHA	8B	(This position shall not be used)
33	LANNA LETTER LANA	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 NYA	9A	(This position shall not be used)
42	LANNA LETTER YA	9B	(This position shall not be used)
43	LANNA LETTER RA	9C	(This position shall not be used)
44	LANNA LETTER RU	9D	(This position shall not be used)
45	LANNA LETTER LA	9E	(This position shall not be used)
46	LANNA LETTER LU	9F	(This position shall not be used)
47	LANNA LETTER WA	A0	LANNA SIGN WIANGWAAK
48	LANNA LETTER HIGH SAA	A1	LANNA SIGN WIANG
49	LANNA LETTER HIGH SSA	A2	LANNA SIGN SAWAN
4A	LANNA LETTER HIGH SAA	A3	LANNA SIGN KEOW
4B	LANNA LETTER HIGH HA	A4	LANNA SIGN HOY
4C	LANNA LETTER LAA	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 GAAN
50	LANNA LETTER II	A9	LANNA SIGN GAANGUU
51	LANNA LETTER U	AA	LANNA SIGN SATGAAN
52	LANNA LETTER UU	AB	LANNA SIGN SATGAANGUU
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 LE	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 LATHA OR LOW PA		
5E	(This position shall not be used)		
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 ONG		
6E	LANNA VOWEL SIGN OH		
6F	LANNA VOWEL SIGN OY		
70	LANNA VOWEL SIGN E		
71	LANNA VOWEL SIGN EE		
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

Michael Everson and Martin Hosken

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

Individual contribution.

4. Submission date

2006-08-06

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

126

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?