ISO/IEC JTC1/SC2/WG2 N3207R
L2/07-007R
2007-03-21

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 for encoding the Lanna script in the BMP of the UCS**
Source: **UC Berkeley Script Encoding Initiative (Universal Scripts Project)**
Authors: **Michael Everson, Martin Hosken, and Peter Constable**
Status: **Liaison Contribution**
Replaces: **N3121, N2042, N1013**
Action: **For consideration by JTC1/SC2/WG2 and UTC**
Date: **2007-03-21**

0. Preface. N3121 was accepted by WG2 and Lanna is at present under ballot. This document revises N3121 to include additional clarifications, to correct a few errors, and to add one additional character, KHUN HIGH CHA (described on page 2). This letter and the additional symbol CAANG are coloured yellow in the code charts.

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 ᨦ LOW NGA is a sequence, ᨲ HA + ◌᩠ SAKOT + ᨦ LOW NGA = ᨶ *high nga*.

k^H	◌ᩉ			kh^H	◌᩠ᨡ	k^L	◌ᨠ	kh^L	◌ᨡ	ng^L	◌ᨦ
c^H	◌ᩈ			ch^H	◌᩠ᨧ	c^L	◌ᨩ	ch^L	◌ᨪ	ny^L	◌ᨬ
t	◌ᨲ			th^H	◌᩠ᨳ	d	◌ᨴ	th^L	◌ᨵ	n	◌ᨶ
t^H	◌ᨳ			th^H	◌᩠ᨵ	t^L	◌ᨨ	th^L	◌ᨯ	n	◌ᨱ
b	◌ᨻ	p^H	◌ᨽ	ph^H	◌᩠ᨿ	p^L	◌ᨷ	ph^L	◌ᨸ	m	◌ᨾ
y^L	◌ᨿ			r	◌ᩁ	l	◌ᩃ	w	◌ᩅ	ss	◌ᩈ᩠ᩈ
sh^H	◌ᩈ	ss^H	◌ᩈ᩠ᩅ	s^H	◌ᩈ᩠	h^H	◌ᩉ	ll	◌ᩃ᩠ᩃ	lae	◌ᩋᩮᩣ

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

kx^H	ᨢ ◌᩠ᨢ	kx^L	ᨤ ◌᩠ᨤ	s^L	ᨪ ◌᩠ᨪ	s^L	ᨪ ◌᩠ᨪ		
f^H	ᨺ ◌᩠ᨺ	f^L	ᨼ ◌᩠ᨼ	y^H	ᩀ ◌᩠ᩀ	h^L	ᩌ ◌᩠ᩌ	rue ᩂ	lue ᩄ

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

In Khün, the letter ᨨ ◌᩠ᨨ HIGH CHA is not used, being replaced by the language-specific form ᨨ ◌᩠ᨨ KHUN HIGH CHA. In Khün the script was reanalyzed and the character is formed in the same way as ᨫ ◌᩠ᨫ LOW CHA is; there is no loop and the first part of the glyph is related to that of ᩈ ◌᩠ᩈ HIGH SA. In a Khün font the shapes are: ᨨ KHUN HIGH CHA, ᨫ LOW CHA, ᩈ HIGH SA.

Due to their Brahmic etymology, some consonants are represented using sequences of characters. In particular, in Northern Thai the consonant *low cha* may be represented by the sequence ᨩ NORTHERN THAI LOW CA + ◌᩠ SAKOT + ᨿ LOW YA (yielding ᨩ᩠ᨿ), although 1A2E ᨫ LOW CHA is also often used as the spelling for this consonant. In Tai Lue, the consonant *high cha* is represented by ᨧ HIGH CA + ◌᩠ SAKOT + ᩈ HIGH SA (yielding ᨧ᩠ᩈ), and the consonant *low cha* is represented by ᨩ LOW CA + ◌᩠ SAKOT + ᨫ LOW CHA. (yielding ᨩ᩠ᨫ).

Consonants may also be subjoined to digits, as in ᪓᩠ᨴ 'thrice', which is ᪓ THAM DIGIT THREE + ◌᩠ SAKOT + ᨴ LOW TA.

A note should be made regarding ᩔ LETTER GREAT SA. Many occurrences of ᩈ SA followed by ᩈ SA are rendered using a ligature ᩔ GREAT SA (as in ᩈᩅᩔᨯᩦ *sawassaḍī* 'hello'), but there are some words in which the ligature does not occur and a normal stack of ᩈ SA and subjoined ◌᩠ᩈ SA (that is, ᩈ᩠ᩈ) may be required. Therefore the ᩔ 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 ◌᩠ᩔ would occur.

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

a ᩋ	a^{nt} ᩋ	i ᩍ	ī ᩎ
u ᩏ	ū ᩐ	ē ᩑ	ō ᩒ

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: ᨠᩕ *kra*, ᨠᩖ *kla*. ◌ᩘ CONSONANT SIGN MAI KANG LAI is used as a final *-ng* in Northern Thai and Tai Lue: ᨠᩘ *kang*. ◌ᩘ CONSONANT SIGN KHUN MAI KANG LAI is used as a final *-ng* in Khün: ᨠᩘ *kang*. ◌ᩙ CONSONANT SIGN FINAL NGA is also used as a final *-ng* in Northern Thai: ᨠᩙ *kang*. ◌ᩚ CONSONANT SIGN LOW PA is used unusually in a Tai Lue word, ᨠᩚᩣ *kappha* 'pregnant'; the normal rendering of ᨻ LOW PA + ◌᩠ SAKOT + ᨽ LOW PHA would be ᨠᨻ᩠ᨽ. The last of these signs has two readings: ◌ᩛ CONSONANT SIGN HIGH RATHA OR LOW PA represents ᨮ HIGH RATHA in ᩈᨱᩛᩣ᩠ᨶ *santhān* 'shape' and in ᩁᨭᩛᨻᩣ᩠ᩃ *ratthabāl* 'government' (note the alternate spelling of this as ᩁᨭ᩠ᨮᨻᩣ᩠ᩃ), and it represents ᨻ LOW PA in ᩈᩛ *sappa* 'omniscience' and in ᩋᩜᩛ *ampa* 'mango'. After ◌᩠ SAKOT the two base letters have their normal

subjoined forms ◌᩠ᨮ and ◌᩠ᨯ; the former is rare, but attested as in ᩈᨳᩣᨷᩢ᩠ᨶᩁᩣᨩᨽᩢ᩠ᨮᨩ᩠ᨿᨦᩉ᩠ᨾᩲ᩵ *sathāban rājabhaṭh chyaṅ hmai* 'Chiang Mai College of Further Education'; the latter is also somewhat rare, as in ᨻᨱ᩠ᨯᩰᨷᨷᩩᩁᩩᩔ *banop burus* 'disciple'. Special note should be taken that the letter U+1A30 ᨮ HIGH RATHA—whose glyph looks like U+1A35 ᨵ and U+1A5D ◌ᩛ brought together—is not the same thing as ᨭᩛ, which is U+1A2F ᨭ RATA + U+1A5D ◌ᩛ CONSONANT SIGN HIGH RATHA OR LOW PA. This is because U+1A30 ᨮ (Brahmic *ṭha*) is not linguistically decomposable whereas U+1A2F ᨭ + U+1A5D ◌ᩛ is (ᨭᩛ Brahmic *ṭṭha*). Users need to be aware of the distinction. Khün does not use U+1A30 ᨮ, but uses the sequence U+1A2F + U+1A5D ᨭᩛ (see Figure 12). In Lanna, however, ᨭᩛ can be used as alternate *spelling* for U+1A30 ᨮ.

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: ᨶ NA + ◌ᩣ AA = ᨶᩣ *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):

1. ᨠᩰ ko = ᨠ ka^{H} + ◌ᩰ o
2. ᨠᩴ kaŋ = ᨠ ka^{H} + ◌ᩴ mai kang
3. ᨠᩘ kaŋ = ᨠ ka^{H} + ◌ᩘ mai kang lai
4. ᨠᩢ ka = ᨠ ka^{H} + ◌ᩢ mai sat
5. ᨠ᩠ᩅᩰᩡ kuaʔ = ᨠ ka^{H} + ◌᩠ sakot + ᩅ wa + ◌ᩰ o + ◌ᩡ a
6. ᨠ᩠ᩅᩰ kua = ᨠ ka^{H} + ◌᩠ sakot + ᩅ wa + ◌ᩰ o
7. ᨠ᩠ᩅ kua = ᨠ ka^{H} + ◌᩠ sakot + ᩅ wa
8. ᨠᩬᩴ kɔː = ᨠ ka^{H} + ◌ᩬ oa below + ◌ᩴ mai kang
9. ᨠᩬ kɔː = ᨠ ka^{H} + ◌ᩬ oa below
10. ᨠᩡ kaʔ = ᨠ ka^{H} + ◌ᩡ a
11. ᨠᩣ kaː = ᨠ ka^{H} + ◌ᩣ aa
12. ᨠᩤ kaː = ᨠ ka^{H} + ◌ᩤ tall aa
13. ᨠᩴᩣ kam = ᨠ ka^{H} + ◌ᩣ aa + ◌ᩴ mai kang
14. ᨠᩴᩤ kam = ᨠ ka^{H} + ◌ᩤ tall aa + ◌ᩴ mai kang
15. ᨠᩥ ki = ᨠ ka^{H} + ◌ᩥ i
16. ᨠᩦ kiː = ᨠ ka^{H} + ◌ᩦ ii
17. ᨠᩧ kɯ = ᨠ ka^{H} + ◌ᩧ ue
18. ᨠᩨ kɯː = ᨠ ka^{H} + ◌ᩨ uue
19. ᨠᩩ ku = ᨠ ka^{H} + ◌ᩩ u
20. ᨠᩪ kuː = ᨠ ka^{H} + ◌ᩪ uu
21. ᨠᩮᩡ keʔ = ᨠ ka^{H} + ᩮ◌ e + ◌ᩡ a
22. ᨠᩮ keː = ᨠ ka^{H} + ᩮ◌ e
23. ᨠᩯᩡ kɛʔ = ᨠ ka^{H} + ᩯ◌ ae + ◌ᩡ a
24. ᨠᩯ kɛː = ᨠ ka^{H} + ᩯ◌ ae
25. ᨠᩮᩥᩬᩡ kɜʔ = ᨠ ka^{H} + ᩮ◌ e + ◌ᩬ oa below + ◌ᩥ i + ◌ᩡ a
26. ᨠᩮᩥᩬ kɜː = ᨠ ka^{H} + ᩮ◌ e + ◌ᩬ oa below + ◌ᩥ i (pronounced kɨa when followed by a final consonant)
27. ᨠᩮᩨᩬᩡ kɜʔ = ᨠ ka^{H} + ᩮ◌ e + ◌ᩬ oa below + ◌ᩨ uue+ ◌ᩡ a (used in Khün)
28. ᨠᩮᩨᩬ kɜː = ᨠ ka^{H} + ᩮ◌ e + ◌ᩬ oa below + ◌ᩨ uue (used in Khün)
29. ᨠᩮᩢᩣ ko = ᨠ ka^{H} + ᩮ◌ e + ◌ᩢ mai sat + ◌ᩣ aa
30. ᨠᩳ kɔː = ᨠ ka^{H} + ◌ᩳ oa above
31. ᨠᩮᩣ koː = ᨠ ka^{H} + ᩮ◌ e + ◌ᩣ aa
32. ᨠᩮ᩠ᨿᩡ kiaʔ = ᨠ ka^{H} + ◌᩠ sakot + ᨿ ya^{L} + ᩮ◌ e + ◌ᩡ a
33. ᨠᩮ᩠ᨿ kia = ᨠ ka^{H} + ◌᩠ sakot + ᨿ ya^{L} + ᩮ◌ e

34. ᨠ᩠ᨿ kia = ᨠ ka^{H} + ◌᩠ sakot + ᨿ ya^{L}
35. ᨠᩮᩬᩥᩋᩡ kɨaʔ = ᨠ ka^{H} + ᩮ◌ e + ◌ᩬ oa below + ◌ᩥ i + ᩋ a^{nt} + ◌ᩡ a
36. ᨠᩮᩬᩥᩋ kɨa = ᨠ ka^{H} + ᩮ◌ e + ◌ᩬ oa below + ◌ᩥ i + ᩋ a^{nt}
37. ᨠᩰᩡ koʔ = ᨠ ka^{H} + ᩰ◌ oo + ◌ᩡ a
38. ᨠᩰ koː = ᨠ ka^{H} + ᩰ◌ oo
39. ᨠᩰᩬᩡ kɔʔ = ᨠ ka^{H} + ᩰ◌ oo + ◌ᩬ oh + ◌ᩡ a
40. ᨠᩢ᩠ᨿ kaj = ᨠ ka^{H} + ◌ᩢ mai sat + ◌᩠ sakot + ᨿ ya^{L}
41. ᨠᩲ kaj = ᨠ ka^{H} + ᩲ◌ tham ai
42. ᨠᩱ kaj = ᨠ ka^{H} + ᩱ◌ ai
43. ᨠᩱ᩠ᨿ kaj = ᨠ ka^{H} + ᩱ◌ ai + ◌᩠ sakot + ᨿ ya^{L}
44. ᨠᩭ kɔj = ᨠ ka^{H} + ◌ᩭ oy (used in Khün)

5.1. TALL AA. Both ◌ᩣ AA and ◌ᩤ TALL AA are encoded because context cannot be relied on to determine which one is used. Choice of the two can be a question of spelling: TALL AA is typically used with the following consonants: ᨷᩤ BA, ᩅᩤ WA, ᨴᩤ LOW TA, ᨵᩤ LOW THA, and ᨣᩤ LOW KA. Udom Rungrueangsri notes that some textbooks say to write ᨧᩤ HIGH CA and ᨻᩤ LOW PA, and others ᨧᩤ HIGH CA, ᩁᩤ RA, and ᨻᩤ LOW PA as well (even though these may also be written ᨧᩣ, ᨻᩣ, ᩁᩣ, and ᨻᩣ).

5.2. AM. The written representation of /am/ involves two visual components: ◌ᩣ VOWEL SIGN AA (or ◌ᩤ VOWEL SIGN TALL AA) and ◌ᩴ MAI KANG, which are stored in that order (since final consonants are always stored after their vowels). In the case of /am/ the MAI KANG is often rendered as part of the preceding cluster to VOWEL SIGN AA. Northern Thai treats AM similarly to Thai AM; it places the MAI KANG glyph to the left of the -AA vowel (whether over the previous cluster or between the clusters): ◌ᩣᩴ, ◌ᩤᩴ. In Khün and Lue, the MAI KANG render the MAI KANG over the -AA vowel: ◌ᩣᩴ, ◌ᩤᩴ.

5.3. Dependent vowel signs in Khün. The Khün character ◌ᩭ OY is not used in Northern Thai. Khün vowel order is quite different from that of Northern Thai.

1. ᨠᩡ kaʔ = ᨠ ka^{H} + ◌ᩡ a
2. ᨠᩣ kaː = ᨠ ka^{H} + ◌ᩣ aa
3. ᨠᩤ kaː = ᨠ ka^{H} + ◌ᩤ tall aa
4. ᨠᩥ ki = ᨠ ka^{H} + ◌ᩥ i
5. ᨠᩦ kiː = ᨠ ka^{H} + ◌ᩦ ii
6. ᨠᩩ ku = ᨠ ka^{H} + ◌ᩩ u
7. ᨠᩪ kuː = ᨠ ka^{H} + ◌ᩪ uu
8. ᨠᩮᩡ keʔ = ᨠ ka^{H} + ᩮ◌ e + ◌ᩡ a
9. ᨠᩮ keː = ᨠ ka^{H} + ᩮ◌ e
10. ᨠ᩠ᨿᩮ kia = ᨠ ka^{H} + ◌᩠ sakot + ᨿ ya^{L} + ᩮ◌ e
11. ᨠᩯᩡ kɛʔ = ᨠ ka^{H} + ᩯ◌ ae + ◌ᩡ a
12. ᨠᩯ kɛː = ᨠ ka^{H} + ᩯ◌ ae
13. ᨠᩧ kɯ = ᨠ ka^{H} + ◌ᩧ ue
14. ᨠᩨ kɯː = ᨠ ka^{H} + ◌ᩨ uue
15. ᨠᩰᩡ koʔ = ᨠ ka^{H} + ᩰ◌ oo + ◌ᩡ a
16. ᨠᩰ koː = ᨠ ka^{H} + ᩰ◌ oo
17. ᨠᩮᩣ koː = ᨠ ka^{H} + ᩮ◌ e + ◌ᩣ aa
18. ᨠ᩠ᩅᩫᩡ kuaʔ = ᨠ ka^{H} + ◌᩠ sakot + ᩅ wa + ◌ᩫ o + ◌ᩡ a
19. ᨠ᩠ᩅᩫ kua = ᨠ ka^{H} + ◌᩠ sakot + ᩅ wa + ◌ᩫ o
20. ᨠᩰᩬᩡ kɔʔ = ᨠ ka^{H} + ᩰ◌ oo + ◌ᩬ oa below + ◌ᩡ a

21. ᨠᩴ kɔː = ᨠ ka^{H} + ᩴ oa above
22. ᨠᩮᩨᩢ kɜʔ = ᨠ ka^{H} + ᩮ e + ᩨ oa below + ᩦ uue+ ᩢ a
23. ᨠᩮᩨᩦ kɜː = ᨠ ka^{H} + ᩮ e + ᩨ oa below + ᩦ uue
24. ᨠᨡ kaj = ᨠ ka^{H} + ᨡ ai
25. ᨠᨡᩣ kaj = ᨠ ka^{H} + ᨡ ai + ◌᩠ sakot + ᨿ ya^{L}
26. ᨠᨣ ko = ᨠ ka^{H} + ᨠ oo + ᩫ mai sat
27. ᨠᩮᨣ kaw = ᨠ ka^{H} + ᩮ e + ᩫ mai sat + ᩣ aa
28. ᨠᩴ kaŋ = ᨠ ka^{H} + ᩴ mai kang
29. ᨠᩬ kaŋ = ᨠ ka^{H} + ᩬ final nga
30. ᨠᩫ᩠ᨣ ka = ᨠ ka^{H} + ᩫ mai sat + ◌᩠ sakot + ᨣ nga
31. ᨠᩴᩦ kiŋ = ᨠ ka^{H} + ᩦ i + ᩴ mai kang
32. ᨠᩦ᩠ᨣ kiŋ = ᨠ ka^{H} + ᩦ i + ◌᩠ sakot + ᨣ nga
33. ᨠᩴᩣ kam = ᨠ ka^{H} + ᩣ aa + ᩴ mai kang
34. ᨠᩫ᩠ᨾ kam = ᨠ ka^{H} + ᩫ mai sat + ◌᩠ sakot + ᨾ ma
35. ᨠᩫᨾ kam = ᨠ ka^{H} + ᩫ mai sat + ᨾ ma
36. ᨠ᩵ kɔj = ᨠ 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: ᨲ᩶ᨦ᩵ *tang* 'be different', ᨲ᩶ᨦ᩵ᨲᩫᨦ *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 MAI SAM 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: ᨴᩣ᩠ᨶᩫᨶ *thanon* 'path' (without the MAI SAM, ᨴᩣᨶᨶ 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 ᨡᩫᩅ᩠ᨡᨦ *khawkhong* 'belongings' (without MAI SAM it is written ᨡᩮᩅᨡᨦ).

The effect of these rules is that MAI SAM is stored in the same place regardless of the precise function it is being used for.

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 ๐๑๒๓๔๕๖๗๘๙ and an "ecclesiastical" set ๐๑๒๓๔๕๖๗๘๙. European digits are are also found in books. Since digits from either set may take subjoined consonants and vowels, the secular set was not unified with the Myanmar digits to which they are related.

9. 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+1A74 < ◌ะ *a* U+1A61 < ก *ka*^{H} U+1A20 < ข *kha*^{H} U+1A21 <
ฃ *kxa*^{H} U+1A22 < ค *ka*^{L} U+1A23 < ฅ *kxa*^{L} U+1A24 < ฆ *kha*^{L} U+1A25 <
ง *nga*^{L} U+1A26 << ◌ึ *mai kang lai* U+1A5A << ◌ั *khun mai kang lai* U+1A5B << ◌ิ *final nga* U+1A5C <
จ *ca*^{H} U+1A27 < ฉ *cha*^{H} U+1A28 <<< [จ *ca*^{H} U+1A27 ◌์ *sakot* U+1A60 ฉ *sa*^{H} U+1A4B] <<<
ฉ *Kh. cha*^{H} U+1A29 < ช *ca*^{L} U+1A2A <<< ฌ *N.T. ca*^{L} U+1A2B < ◌ิ *sa*^{L} U+1A2C <<< ◌ิ *N.T. sa*^{L} U+1A2D <
ฌ *cha*^{L} U+1A2E <<< [ฌ *N.T. ca*^{L} U+1A2B ◌์ *sakot* U+1A60 ฃ *ya*^{L} U+1A42] <<<
[ช *ca*^{L} U+1A2A ◌์ *sakot* U+1A60 ฌ *cha*^{L} U+1A2E] < ญ *nya*^{L} U+1A2F <
ฐ *rata* U+1A30 < ฑ *ratha*^{H} U+1A31 < ฒ *da* U+1A32 < ฐ *ratha*^{L} U+1A33 < ณ *rana* U+1A34 <
ด *ta*^{H} U+1A35 < ต *tha*^{H} U+1A36 < ถ *ta*^{L} U+1A37 < ท *tha*^{L} U+1A38 < น *na*^{L} U+1A39 <
บ *ba* U+1A3A < ป *pa*^{H} U+1A3B < ผ *pha*^{H} U+1A3C < ฝ *fa*^{H} U+1A3D <
ภ *pa*^{L} U+1A3E << ◌ิ *pa*^{L} U+1A5D <<< ฎ *ratha*^{H} *or pa*^{L} U+1A5E < ◌ิ *fa*^{L} U+1A3F <
พ *pha*^{L} U+1A40 < ฟ *ma* U+1A41 < ย *ya*^{L} U+1A42 < ◌ั *ya*^{H} U+1A43 < ร *ra* U+1A44 < ◌ึ *rue* U+1A45 <
ล *la* U+1A46 < ◌ึ *lue* U+1A47 < ว *wa* U+1A48 <
ศ *sha*^{H} U+1A49 < ษ *ssa*^{H} U+1A4A < ส *sa*^{H} U+1A4B < ห *ha*^{H} U+1A4C < ฬ *lla* U+1A4D <
อ *a* U+1A4E <<< ฮ *N.T. a* U+1A4F < ฮ *ha*^{L} U+1A56 < ◌ั *lae* U+1A57 <
◌ั *o* U+1A6B < ◌ิ *medial ra* U+1A58 < ◌ิ *medial la* U+1A59 < [◌์ *sakot* U+1A60 ว *wa* U+1A48] <
◌ุ *oa below* U+1A6C < ◌ั *mai sat* U+1A62 < ◌ะ *aa* U+1A63 <<< ◌า *tall aa* U+1A64 <
[◌ำ *am* U+1A63 + U+1A74] <<< [◌า *tall am* U+1A64 U+1A74] < ◌ิ *i* U+1A65 < ◌ี *ii* U+1A66 <
◌ึ *ue* U+1A67 < ◌ื *uue* U+1A68 < ◌ุ *u* U+1A69 < ◌ู *uu* U+1A6A <
◌เ *e* U+1A6E < ◌แ *ae* U+1A6F < ◌็ *oa above* U+1A73 <
◌ิ *oo* U+1A70 < ◌ไ *ai* U+1A71 < ◌ิ *tham ai* U+1A72 < ◌ๅ *oy* U+1A6D <
[อ *a* U+1A4E ◌ิ *i* U+1A65] <<< [ฮ *N.T. a* U+1A4F ◌ิ *i* U+1A65] < ฦ *i* U+1A50 <
[อ *a* U+1A4E ◌ี *ii* U+1A66] <<< [ฮ *N.T. a* U+1A4F ◌ี *ii* U+1A66] < ฦ *ii* U+1A51 <
[อ *a* U+1A4E ◌ุ *u* U+1A69] <<< [ฮ *N.T. a* U+1A4F ◌ุ *u* U+1A67] < ฯ *u* U+1A52 <
[อ *a* U+1A4E ◌ู *uu* U+1A6A] <<< [ฮ *N.T. a* U+1A4F ◌ู *uu* U+1A6A] < ฯ *uu* U+1A53 <
[อ *a* U+1A4E ◌เ *e* U+1A6E] <<< [ฮ *N.T. a* U+1A4F ◌เ *e* U+1A6E] < ฯ *ee* U+1A54 <
[อ *a* U+1A4E ◌แ *ae* U+1A6F] <<< [ฮ *N.T. a* U+1A4F ◌แ *ae* U+1A6F] < ◌ิ *oo* U+1A55 <
◌่ *tone-1* U+1A75 << ◌้ *tone-2* U+1A76 << ◌๊ *khun tone-3* U+1A77 << ◌๋ *khun tone-4* U+1A78 <<
◌๎ *khun tone-5* U+1A79 << ◌ั *ra haam* U+1A7A << ◌ั *mai sam* U+1A7B <<
◌ฺ *cryptogrammic dot* U+1A7F

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

10. Linebreaking. Opportunities for linebreaking are lexical, but a linebreak may not be inserted between a base letter and and a following combining mark. A linebreak should also not be inserted between SAKOT and a following consonant letter. There is no line-breaking hypenation.

11. 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 ᩋ 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 ᨠ KEOW 'courtyard', ᨡ HOY 'oyster', and ᨢ DOKMAI 'flower' are used as dingbats and as section starters; ᨠ᩟ and ᩟ᨢ᩟ have been observed. Another symbol used is ᩃ CAANG 'elephant', normally accompanied with the long AA: ᩃᩣ (compare ᨧᩣᨦ *cāng* 'elephant'). The mark ᩦ MAI YAMOK is used in the same way as its Thai counterpart, U+0E46 ๆ THAI CHARACTER MAIYAMOK.

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. A regular expression description. The following description may assist in implementation:

```
Ci (((S Cs)|H|MS)? M? Vm? Vp? Vl? Vu? T? Vf? A? (S Cf)? Fu?)* Fm?
```

Where:

```
Ci (Initial Consonant) =1A20 .. 1A57, 1A5F, 1A80 .. 1A89, 1A90 .. 1A99
MS (Mai Sam) = 1A7D
S  (Sakot) = 1A60
Cs (Subjoined Consonant) = 1A20 .. 1A4D, 1A57 excluding anything that cannot be subjoined
H  (Hang) = 1A5E
M  (Medial) = 1A58, 1A59, 1A60 1A48
Vm (Medial Vowel) = 1A60 [1A42, 1A48]
Vp (Pre Vowel) = 1A6E .. 1A72
Vl (Lower Vowel) = 1A69, 1A6A, 1A6C
Vu (Upper Vowel) = 1A62, 1A65 .. 1A68, 1A6B, 1A73, 1A74
T  (Tone) = 1A75 .. 1A79
Vf (Following Vowel) = 1A63 .. 1A64, 1A4E, 1A4F
A (Short a) = 1A61
Cf (Final Consonant) = 1A20 .. 1A4D
Fu (Final Upper diacritic) = 1A5A .. 1A5D, 1A6D, 1A74
Fm (Final Modifier) = 1A7A
```

This does not describe a complete linguistic syllable, because that can have a following base consonant and subsequent chaining. It is not a cluster because Vf could well take a cursor before it. It describes the orthographic syllable. The regular expression describes has a number of properties.

- Every word forming character in the Lanna block from 1A20..1A7B appears somewhere in the regular expression. So that at least means we know at least one place for each code to go.
- A number of characters appear in more than one element. This shows some of the ambiguity that exists in the Lanna script. For example, 1A74 can be both a Vu and an Fu. Notice also the number of places a SAKOT WA (1A60 1A47) can turn up.

While the regular expression given here is fairly complete when considering modern usage of the Lanna script, it is also too generous allowing ridiculous sequences that would never occur. According to this expression you could have WA SAKOT WA SAKOT WA SAKOT WA SAKOT WA. The purpose of this regular

expression, therefore, is not legal sequence constraint, but ordering. Given a string of elements, how should they be ordered? The Lanna script, historically, has been very creative in its spelling and therefore there is probably historic precedent for almost any appropriate deviation from this ordering for some historic case or other. It may, therefore, not be appropriate for this order, or any other, to be enshrined and enforced as part of the encoding standard. It is up to the implementations to support their users.

14. Consonant conjoining behaviours. While Lanna is a Brahmic script and shares many attributes common to Brahmic scripts, it is atypical in the variety of ways in which consonant conjoining occurs. The encoding model for Lanna is similar to that for Myanmar and Khmer, using a CEONG-like character plus some combining medial-consonant characters. In order to fully grasp the Lanna encoding model, however, it is necessary to understand the various ways in which consonant conjoining is used.

Typical behaviour for Brahmic scripts is to have structural units in the text often referred to as *orthographic syllables*. These units consist of an initial consonant or consonantal sequence plus satellite vowel marks. Because all consonants are combined into the initial portion of the orthographic syllable, the boundaries do not align exactly with phonological syllables; yet orthographic and phonological syllables correlated roughly one-to-one.

Lanna script has these same behaviours, but it also uses conjoining in ways not found in Brahmic scripts generally. For instance, a single orthographic syllable can encompass two full phonological syllables while maintaining the same basic formal organizational structure of the textual elements: a stack of conjoining consonant typeforms with satellite vowel marks.

The following is a break-down of the various ways in which consonant conjoining is used in Lanna script. The following notation is used:

- boundaries of phonological syllables are indicated by the FULL STOP "."
- CURLY BRACKETS are used to indicate how text elements are organized into conjoining units (base + conjoining consonant + other signs); that is, a string "{… }" represents one conjoining unit
- the consonant letter ᩉ HIGH HA is indicated as "h"
- consonant letters ᩅ LOW YA and ᩋ WA when used in the representation of diphthongs /ia/ and /ua/ will be indicated as "C_V".

Examples are provided, along with the encoded character sequences that would be used to represent them. This is done to make clear how the encoding model would apply to these various kinds of text scenario.

In describing the different ways of using conjoining, it is only necessary to refer to consonant and vowel elements. In actual text, other elements will occur, such as tone marks, and, of course, the character SAKOT.

14.1. High-tone-class sonorant consonants are written by conjoining the low-class sonorant to HIGH HA (this is common in Southeast Asian scripts):

h C … is written as {h C …}

For example, /hni:/: ᩉᩢᨶᩦ = HA^H + SAKOT + NA + II

14.2. Phonological syllable-initial consonant clusters are conjoined (this is typical across Brahmic scripts, though not used in Thai or Lao):

. C1 C2 … is written as {C1 C2 …}

For example, /khwaen/: ᨰᩫᩮᨶ = KXA^{L} + SAKOT + WA + AE + TONE-2 + RA

14.3. The consonant letters LOW YA and WA are written in conjoined form as part of the representation for certain vowels, such as diphthongs /ia/ and /ua/. (This may be derived historically from syllable-initial clusters, and so could be considered a variation of 14.2.)

. C1 C_V … is written as {C1 C_V … }

For example, /hua/: ᩉᩪ = HA^{H} + SAKOT + WA + SIGN O

14.4. A sequence of phonological syllable-final and syllable-initial consonants are conjoined (this is typical across all Brahmic scripts, except in Southeast Asia):

… C1 . C2 … is written as {…}{C1 C2 …}

For example, /dang ni:/: ᨴᩢᩴᨦᩤ = DA + MAI SAT + TONE-1 + NGA + SAKOT + NA + II + TONE-2

14.5. The initial and final consonants of a phonological syllable are conjoined (this is uncommon among Brahmic scripts, but is very commonly-used in Lanna):

. C1 V C2 . is written as {C1 V C2}

For example, /hin/: ᩉᩤᨶ = HA^{H} + I + SAKOT + NA

14.6. The initial consonants of two consecutive phonological syllables are conjoined (this is uncommon among Brahmic scripts):

. C1 V . C2 V … is written as {C1 V C2 V}

For example, /bo.mi:/: ᨻᩮᩬᨾᩤ = BA + MAI KANG + TONE-1 + SAKOT + MA + II

Or, for example, /thanon/: ᨳᩢᩴᨶ = THA^{H} + SAKOT + NA + MAI SAM + SIGN O + RA

14.7. The final consonant of a phonological syllable is conjoined to the vowel signs AA or TALL AA—these are the vowel signs that are spacing and position to the right of the initial consonant (this is uncommon among Brahmic scripts).

. C1 V C2 . is written as {C1}{V C2}

For example, /ta:m/: ᨲᩣᨾ = TA^{H} + AA + SAKOT + MA

14.8. A further kind of conjoining is formally like 14.7 but involving two syllables where the second syllable consists of a consonant conjoined to AA or TALL AA with an above or below vowel mark (uncommon among Brahmic scripts, and rare even in Lanna).

. C1 V1 . C2 V2 is written as {C1}{V1 C2 V2}

For example, /pya:.thi/: ᨻ᩠ᨿᩣ᩠ᨴᩥ = PA^{L} + SAKOT + YA^{L} + AA + SAKOT + THA^{L} + I

14.9. These different kinds of conjoining can co-occur in a single conjoining unit. It should be noted in particular that this can result in a conjoining stack of three (or perhaps more?) consonants.

For example, /plian/, which combines 14.2 and 14.3: ᨸ᩠ᩃ᩠ᨿ᩵ᩁ = PA^{H} + SAKOT + LA + SAKOT + YA^{L} + TONE-1 + RA

Or, for example, /maen.wa:/, combining 14.5 and 14.6: ᨾᩯ᩠᩶ᨶ᩠ᩅ᩵ᩣ = MA + AE + TONE-2 + SAKOT + NA + SAKOT + WA + TONE-1 + AA

14.10. Because consonants can conjoin in different contexts, ambiguous readings can result. In other words, two different encoded sequences corresponding to two different readings may display identically. For instance, because YA and WA can occur as part of a syllable-initial consonant cluster, but can also be used to write certain vowels or can occur as syllable-final consonants, and because in all these cases they would conjoin to the first consonant, words written with these as conjoined consonants may have different readings. For example, /swe/ ᩈ᩠ᩅᩯ᩵ SA^{H} + SAKOT + WA + AE + TONE-1 /swe/ (MFL p764) displays identically to another word in the dictionary: /sew/ ᩈ᩵ᩯ᩠ᩅ SA^{H} + TONE-1 + AE + SAKOT + WA (MFL, p766). Notice that both words appear with identical renderings, but they occur in different places in the dictionary corresponding to the different readings.

15. Syllable shapes. The following description exemplifies the ways in which characters can interact to form syllables. It derives from a book called *Akśāra Lānnā*.

1. ᨿᩪ yuu = ᨿ ya^{L} + ᩪ uu
2. ᨾᩦ mii = ᨾ ma + ᩦ ii
3. ᩉ᩠ᨾᩪ hmuu = ᩉ ha^{L} + ᩠ sakot + ᨾ ma + ᩪ uu
4. ᩉ᩠ᨾᩦ hmii = ᩉ ha^{L} + ᩠ sakot + ᨾ ma + ᩦ ii
5. ᨽ᩠ᩅᩫ phua = ᨽ pha^{L} + ᩠ sakot + ᩅ wa + ᩫ o
6. ᩉ᩠ᩃᩬᩴ᩵ law = ᩉ ha^{L} + ᩠ sakot + ᩃ la + ᩬ oa below + ᩴ mai kang + ᩵ tone-1
7. ᨾᩣ ma = ᨾ ma + ᩣ aa
8. ᩱᩉ᩵ hai = ᩉ ha^{L} + ᩱ ai + ᩵ tone-1
9. ᨾ᩠ᨿ mia = ᨾ ma + ᩠ sakot + ᨿ ya^{L}
10. ᩅ᩠ᨿᨦ wiang = ᩅ wa + ᩠ sakot + ᨿ ya^{L} + ᨦ nga^{L}
11. ᩉᩣ᩠ᨾ haam = ᩉ ha^{L} + ᩣ aa + ᩠ sakot + ᨾ ma
12. ᨯᩣᩴ daam = ᨯ da^{L} + ᩣ aa + ᩴ mai kang
13. ᩮᨡ᩠ᩅ khaw = ᨡ kha^{H} + ᩮ e + ᩠ sakot + ᩅ wa
14. ᩉ᩠ᨾᩣ hmaa = ᩉ ha^{L} + ᩠ sakot + ᨾ ma + ᩣ aa
15. ᨠᩕᩣ᩠ᨸ krap = ᨠ ka^{H} + ᩕ medial ra + ᩣ aa + ᩠ sakot + ᨸ pa^{H}
16. ᨻᩕ᩵ᩣᩴ pram = ᨻ pa^{L} + ᩕ medial ra + ᩵ tone-1 + ᩣ aa + ᩴ mai kang
17. ᨠᩕᩬᨦ krong = ᨠ ka^{H} + ᩕ medial ra + ᩬ oa below + ᨦ nga^{L}
18. ᩈᩕᩫᨾ᩠ᨱ᩺ som = ᩈ sa^{H} + ᩕ medial ra + ᩫ o + ᨾ ma + ᩠ sakot + ᨱ rana + ᩺ ra haam
19. ᩈᩕ᩠ᩅᩫᨾ suam = ᩈ sa^{H} + ᩕ medial ra + ᩠ sakot + ᩅ wa + ᩫ o + ᨾ ma
20. ᩮᨯᩬᩨᩁ deuan = ᨯ da^{L} + ᩮ e + ᩬ oa below + ᩨ uue + ᩁ ra

21. ᩁᩮᩬ reua = ᩁ ra + ᩮ e + ᩬ oa below + ᩦ uue + ᩬ a
22. ᩃᩮᩬ leua = ᩉ ha^{L} + ◌᩠ sakot + ᩃ la + ᩮ e + ᩬ oa below + ᩦ uue + ᩬ a
23. ᩉ᩶ᩴ hmaam = ᩉ ha^{L} + ◌᩠ sakot + ᩋ ma + ᩶ tone-1 + ᩴ aa + ᩴ mai kang
24. ᩈᩮᩯ smer smer = ᩈ sa^{H} + ◌᩠ sakot + ᩋ ma + ᩬ oa below + ᩦ i + ᩯ mai sam
25. ᩉᩮᩤ hmeuang = ᩉ ha^{L} + ◌᩠ sakot + ᩋ ma + ᩮ e+ ᩬ oa below + ᩦ uue + ᩤ nga^{L}
26. ᩉᩮᩬᩤ hyeuang = ᩉ ha^{L} + ◌᩠ sakot + ᩌ ya^{L} + ᩮ e+ ᩬ oa below + ᩦ uue + ᩤ nga^{L}
27. ᩉ᩶ᩣ hmon = ᩉ ha^{L} + ◌᩠ sakot + ᩋ ma + ᩣ o + ᩶ tone-1 + ᩁ ra
28. ᩉᩱᩴ hlaay = ᩉ ha^{L} + ◌᩠ sakot + ᩃ la + ᩱ aa + ◌᩠ sakot + ᩌ ya^{L}
29. ᩉᩮᩬᩤ hleuang = ᩉ ha^{L} + ◌᩠ sakot + ᩃ la + ᩮ e+ ᩬ oa below + ᩦ uue + ᩤ nga^{L}

16. 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 SAKOT can follow vowels in Lanna.

```
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 KHUN HIGH CHA;Lo;0;L;;;;;N;;;;;
1A2A;LANNA LETTER LOW CA;Lo;0;L;;;;;N;;;;;
1A2B;LANNA LETTER NORTHERN THAI LOW CA;Lo;0;L;;;;;N;;;;;
1A2C;LANNA LETTER LOW SA;Lo;0;L;;;;;N;;;;;
1A2D;LANNA LETTER NORTHERN THAI LOW SA;Lo;0;L;;;;;N;;;;;
1A2E;LANNA LETTER LOW CHA;Lo;0;L;;;;;N;;;;;
1A2F;LANNA LETTER LOW NYA;Lo;0;L;;;;;N;;;;;
1A30;LANNA LETTER RATA;Lo;0;L;;;;;N;;;;;
1A31;LANNA LETTER HIGH RATHA;Lo;0;L;;;;;N;;;;;
1A32;LANNA LETTER DA;Lo;0;L;;;;;N;;;;;
1A33;LANNA LETTER LOW RATHA;Lo;0;L;;;;;N;;;;;
1A34;LANNA LETTER RANA;Lo;0;L;;;;;N;;;;;
1A35;LANNA LETTER HIGH TA;Lo;0;L;;;;;N;;;;;
1A36;LANNA LETTER HIGH THA;Lo;0;L;;;;;N;;;;;
1A37;LANNA LETTER LOW TA;Lo;0;L;;;;;N;;;;;
1A38;LANNA LETTER LOW THA;Lo;0;L;;;;;N;;;;;
1A39;LANNA LETTER NA;Lo;0;L;;;;;N;;;;;
1A3A;LANNA LETTER BA;Lo;0;L;;;;;N;;;;;
1A3B;LANNA LETTER HIGH PA;Lo;0;L;;;;;N;;;;;
1A3C;LANNA LETTER HIGH PHA;Lo;0;L;;;;;N;;;;;
1A3D;LANNA LETTER HIGH FA;Lo;0;L;;;;;N;;;;;
1A3E;LANNA LETTER LOW PA;Lo;0;L;;;;;N;;;;;
1A3F;LANNA LETTER LOW FA;Lo;0;L;;;;;N;;;;;
1A40;LANNA LETTER LOW PHA;Lo;0;L;;;;;N;;;;;
1A41;LANNA LETTER MA;Lo;0;L;;;;;N;;;;;
1A42;LANNA LETTER LOW YA;Lo;0;L;;;;;N;;;;;
1A43;LANNA LETTER HIGH YA;Lo;0;L;;;;;N;;;;;
1A44;LANNA LETTER RA;Lo;0;L;;;;;N;;;;;
1A45;LANNA LETTER RUE;Lo;0;L;;;;;N;;;;;
1A46;LANNA LETTER LA;Lo;0;L;;;;;N;;;;;
1A47;LANNA LETTER LUE;Lo;0;L;;;;;N;;;;;
1A48;LANNA LETTER WA;Lo;0;L;;;;;N;;;;;
1A49;LANNA LETTER HIGH SHA;Lo;0;L;;;;;N;;;;;
1A4A;LANNA LETTER HIGH SSA;Lo;0;L;;;;;N;;;;;
1A4B;LANNA LETTER HIGH SA;Lo;0;L;;;;;N;;;;;
```

```
1A4C;LANNA LETTER HIGH HA;Lo;0;L;;;;;N;;;;;
1A4D;LANNA LETTER LLA;Lo;0;L;;;;;N;;;;;
1A4E;LANNA LETTER A;Lo;0;L;;;;;N;;;;;
1A4F;LANNA LETTER NORTHERN THAI A;Lo;0;L;;;;;N;;;;;
1A50;LANNA LETTER I;Lo;0;L;;;;;N;;;;;
1A51;LANNA LETTER II;Lo;0;L;;;;;N;;;;;
1A52;LANNA LETTER U;Lo;0;L;;;;;N;;;;;
1A53;LANNA LETTER UU;Lo;0;L;;;;;N;;;;;
1A54;LANNA LETTER EE;Lo;0;L;;;;;N;;;;;
1A55;LANNA LETTER OO;Lo;0;L;;;;;N;;;;;
1A56;LANNA LETTER LOW HA;Lo;0;L;;;;;N;;;;;
1A57;LANNA LETTER LAE;Lo;0;L;;;;;N;;;;;
1A58;LANNA CONSONANT SIGN MEDIAL RA;Mc;0;L;;;;;N;;;;;
1A59;LANNA CONSONANT SIGN MEDIAL LA;Mn;0;NSM;;;;;N;;;;;
1A5A;LANNA SIGN MAI KANG LAI;Mn;0;NSM;;;;;N;;;;;
1A5B;LANNA SIGN KHUN MAI KANG LAI;Mn;0;NSM;;;;;N;;;;;
1A5C;LANNA CONSONANT SIGN FINAL NGA;Mn;0;NSM;;;;;N;;;;;
1A5D;LANNA CONSONANT SIGN LOW PA;Mn;0;NSM;;;;;N;;;;;
1A5E;LANNA CONSONANT SIGN HIGH RATHA OR LOW PA;Mn;0;NSM;;;;;N;;;;;
1A5F;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 I;Mn;0;NSM;;;;;N;;;;;
1A66;LANNA VOWEL SIGN II;Mn;0;NSM;;;;;N;;;;;
1A67;LANNA VOWEL SIGN UE;Mn;0;NSM;;;;;N;;;;;
1A68;LANNA VOWEL SIGN UUE;Mn;0;NSM;;;;;N;;;;;
1A69;LANNA VOWEL SIGN U;Mn;0;NSM;;;;;N;;;;;
1A6A;LANNA VOWEL SIGN UU;Mn;0;NSM;;;;;N;;;;;
1A6B;LANNA VOWEL SIGN O;Mn;0;NSM;;;;;N;;;;;
1A6C;LANNA VOWEL SIGN OA BELOW;Mn;0;NSM;;;;;N;;;;;
1A6D;LANNA VOWEL SIGN OY;Mc;0;L;;;;;N;;;;;
1A6E;LANNA VOWEL SIGN E;Mc;0;L;;;;;N;;;;;
1A6F;LANNA VOWEL SIGN AE;Mc;0;L;;;;;N;;;;;
1A70;LANNA VOWEL SIGN OO;Mc;0;L;;;;;N;;;;;
1A71;LANNA VOWEL SIGN AI;Mc;0;L;;;;;N;;;;;
1A72;LANNA VOWEL SIGN THAM AI;Mc;0;L;;;;;N;;;;;
1A73;LANNA VOWEL SIGN OA ABOVE;Mn;0;NSM;;;;;N;;;;;
1A74;LANNA SIGN MAI KANG;Mn;0;NSM;;;;;N;;;;;
1A75;LANNA SIGN TONE-1;Mn;0;NSM;;;;;N;;;;;
1A76;LANNA SIGN TONE-2;Mn;0;NSM;;;;;N;;;;;
1A77;LANNA SIGN KHUN TONE-3;Mn;0;NSM;;;;;N;;;;;
1A78;LANNA SIGN KHUN TONE-4;Mn;0;NSM;;;;;N;;;;;
1A79;LANNA SIGN KHUN TONE-5;Mn;0;NSM;;;;;N;;;;;
1A7A;LANNA SIGN RA HAAM;Mn;0;NSM;;;;;N;;;;;
1A7B;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 KAAN;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;;;;;
1AAD;LANNA SIGN CAANG;Po;0;L;;;;;N;;;;;
```

17. Bibliography.

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

"ตั๋าเมือง" ISBN 974-9942-00-0

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

Udom Rungrueangsri. 2004. ᨷᨧᨶᩣ᩠ᨶᩩᨠᩕᨾ᩺ᩃ᩶ᩣ᩠ᨶᨶᩣ-ᨴᩱ ᨨ᩠ᨷᩢ᩠ᨷ ᩯᨾ᩵ᨼ᩶ᩣᩉ᩠ᩃ᩠ᩅᨦ (*Pacanānukrom Lānnā-Thai: Chabap maefāhluang*). ISBN 974-685-175-9

ᩯᨷ᩠ᨷᩁ᩠ᨿᩁ᩠ᨶᨽᩣᩈᩣᩃ᩶ᩣ᩠ᨶᨶᩣ (*Baepryarn phāsā Lānnā*). ISBN 974-386-044-4¶

ᨽᩣᩇᩣᨾᩮᩬᩥᨦᩃ᩶ᩣ᩠ᨶᨶᩣ (*Phāśāmueng Lānnā*).

ลมูล จันทน์หอม. 1999. ᩋᨠ᩠ᨡᩁᩃ᩶ᩣ᩠ᨶᨶᩣ (*Aksorn Lānnā*).

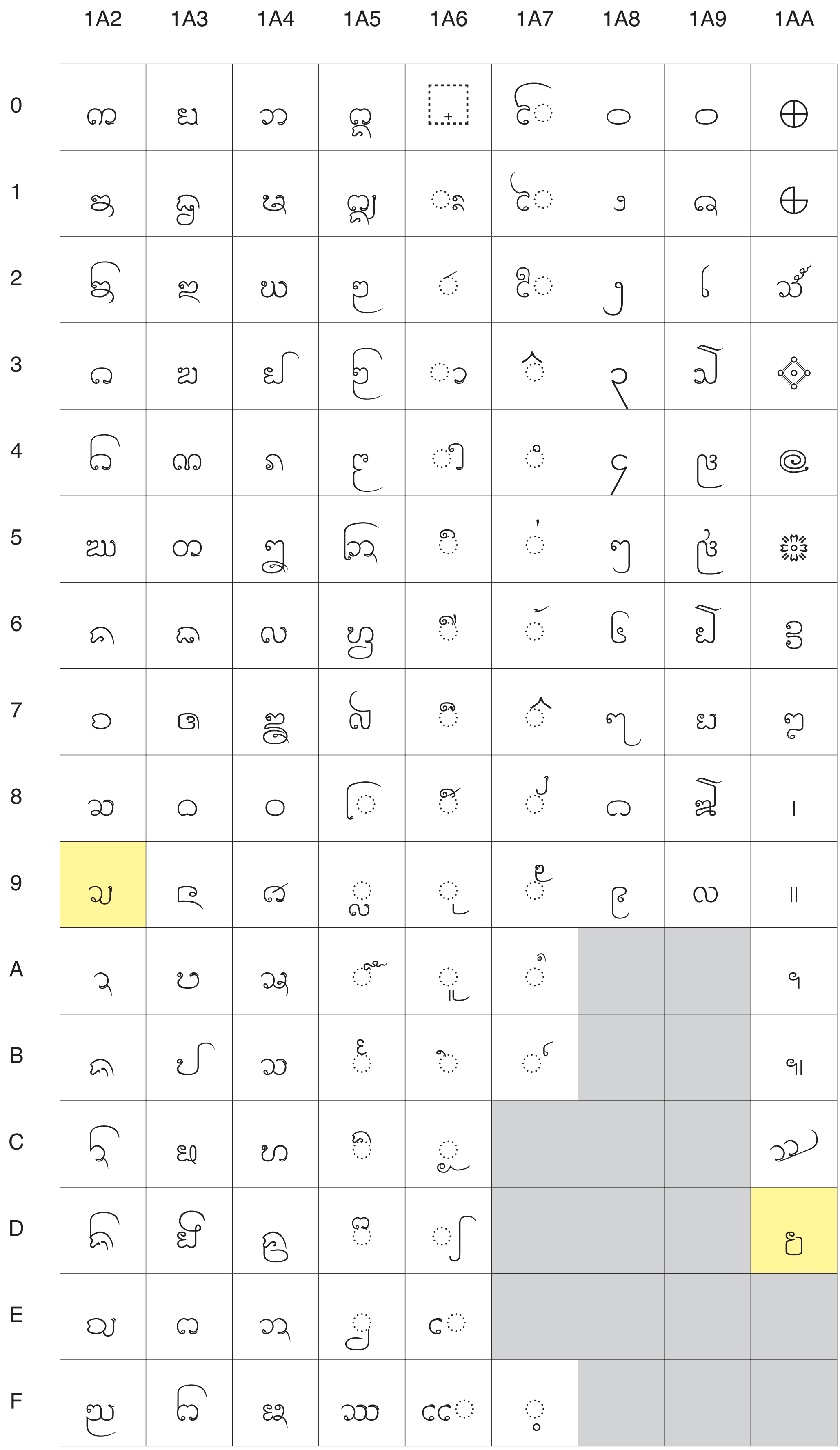
18. Code charts.

Two code charts are given below, one in a Northern Thai-style font, and one in a Khün-style font.

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.

TABLE XX - Row 1A: LANNA (Northern Thai font style)



G = 00
P = 00

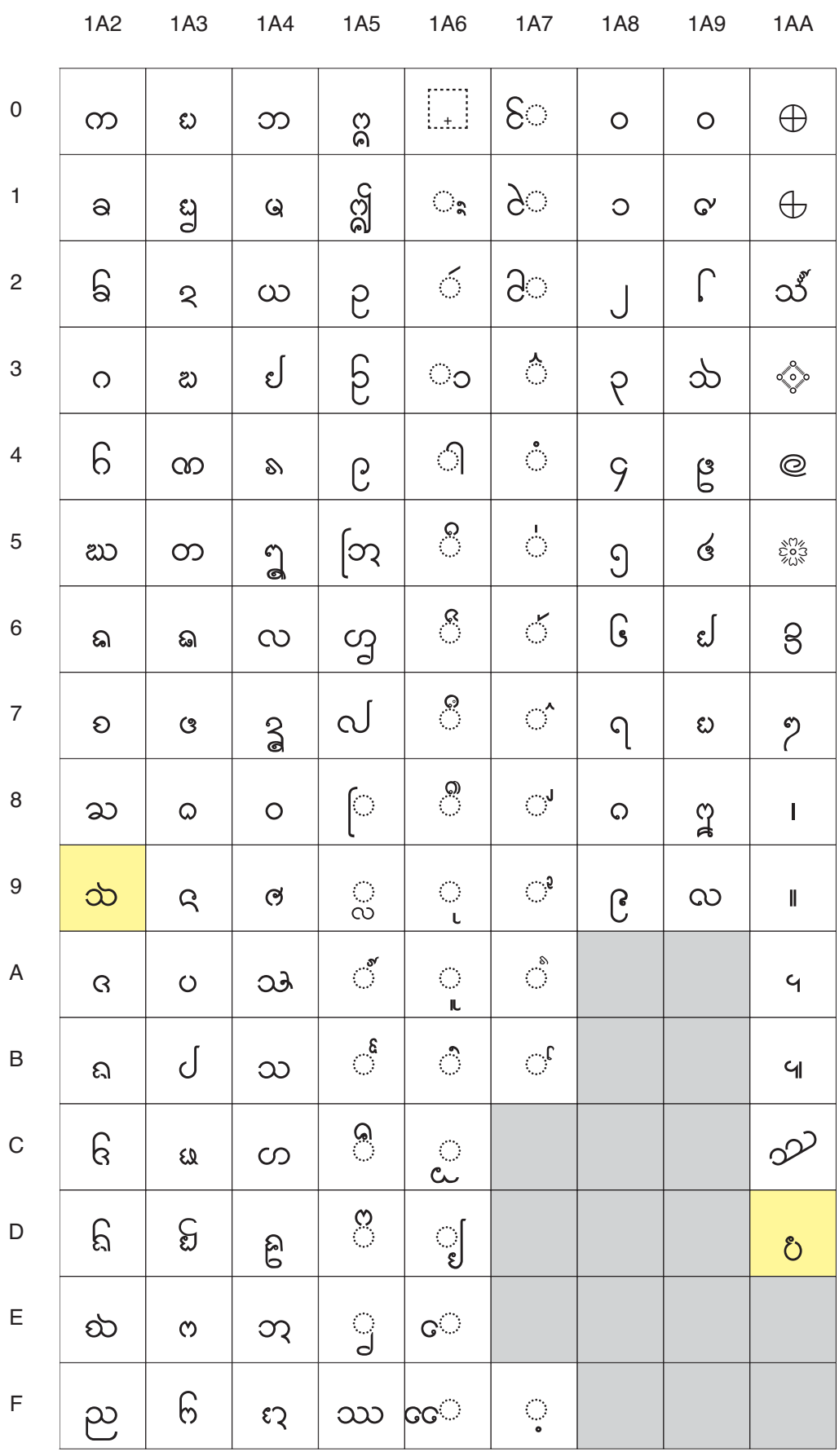
TABLE XX - Row 1A: LANNA

hex	Name
20	LANNA LETTER HIGH KA
21	LANNA LETTER HIGH KHA
22	LANNA LETTER HIGH KXA
23	LANNA LETTER LOW KA
24	LANNA LETTER LOW KXA
25	LANNA LETTER LOW KHA
26	LANNA LETTER LOW NGA
27	LANNA LETTER HIGH CA
28	LANNA LETTER HIGH CHA
29	LANNA LETTER KHUN HIGH CHA
2A	LANNA LETTER LOW CA
2B	LANNA LETTER NORTHERN THAI LOW CA
2C	LANNA LETTER LOW SA
2D	LANNA LETTER NORTHERN THAI LOW SA
2E	LANNA LETTER LOW CHA
2F	LANNA LETTER LOW NYA
30	LANNA LETTER RATA
31	LANNA LETTER HIGH RATHA
32	LANNA LETTER DA
33	LANNA LETTER LOW RATHA
34	LANNA LETTER RANA
35	LANNA LETTER HIGH TA
36	LANNA LETTER HIGH THA
37	LANNA LETTER LOW TA
38	LANNA LETTER LOW THA
39	LANNA LETTER NA
3A	LANNA LETTER BA
3B	LANNA LETTER HIGH PA
3C	LANNA LETTER HIGH PHA
3D	LANNA LETTER HIGH FA
3E	LANNA LETTER LOW PA
3F	LANNA LETTER LOW FA
40	LANNA LETTER LOW PHA
41	LANNA LETTER MA
42	LANNA LETTER LOW YA
43	LANNA LETTER HIGH YA
44	LANNA LETTER RA
45	LANNA LETTER RUE
46	LANNA LETTER LA
47	LANNA LETTER LUE
48	LANNA LETTER WA
49	LANNA LETTER HIGH SHA
4A	LANNA LETTER HIGH SSA
4B	LANNA LETTER HIGH SA
4C	LANNA LETTER HIGH HA
4D	LANNA LETTER LLA
4E	LANNA LETTER A
4F	LANNA LETTER NORTHERN THAI A
50	LANNA LETTER I
51	LANNA LETTER II
52	LANNA LETTER U
53	LANNA LETTER UU
54	LANNA LETTER EE
55	LANNA LETTER OO
56	LANNA LETTER LOW HA
57	LANNA LETTER LAE
58	LANNA CONSONANT SIGN MEDIAL RA
59	LANNA CONSONANT SIGN MEDIAL LA
5A	LANNA SIGN MAI KANG LAI
5B	LANNA SIGN KHUN MAI KANG LAI
5C	LANNA CONSONANT SIGN FINAL NGA
5D	LANNA CONSONANT SIGN LOW PA
5E	LANNA CONSONANT SIGN HIGH RATHA OR LOW PA
5F	LANNA LETTER GREAT SA
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 I
66	LANNA VOWEL SIGN II
67	LANNA VOWEL SIGN UE
68	LANNA VOWEL SIGN UUE
69	LANNA VOWEL SIGN U
6A	LANNA VOWEL SIGN UU
6B	LANNA VOWEL SIGN O
6C	LANNA VOWEL SIGN OA BELOW
6D	LANNA VOWEL SIGN OY
6E	LANNA VOWEL SIGN E
6F	LANNA VOWEL SIGN AE
70	LANNA VOWEL SIGN OO
71	LANNA VOWEL SIGN AI
72	LANNA VOWEL SIGN THAM AI
73	LANNA VOWEL SIGN OA ABOVE
74	LANNA SIGN MAI KANG
75	LANNA SIGN TONE-1
76	LANNA SIGN TONE-2
77	LANNA SIGN KHUN TONE-3
78	LANNA SIGN KHUN TONE-4

hex	Name
79	LANNA SIGN KHUN TONE-5
7A	LANNA SIGN RA HAAM
7B	LANNA SIGN MAI SAM
7C	(This position shall not be used)
7D	(This position shall not be used)
7E	(This position shall not be used)
7F	LANNA COMBINING CRYPTOGRAMMIC DOT
80	LANNA DIGIT ZERO
81	LANNA DIGIT ONE
82	LANNA DIGIT TWO
83	LANNA DIGIT THREE
84	LANNA DIGIT FOUR
85	LANNA DIGIT FIVE
86	LANNA DIGIT SIX
87	LANNA DIGIT SEVEN
88	LANNA DIGIT EIGHT
89	LANNA DIGIT NINE
8A	(This position shall not be used)
8B	(This position shall not be used)
8C	(This position shall not be used)
8D	(This position shall not be used)
8E	(This position shall not be used)
8F	(This position shall not be used)
90	LANNA THAM DIGIT ZERO
91	LANNA THAM DIGIT ONE
92	LANNA THAM DIGIT TWO
93	LANNA THAM DIGIT THREE
94	LANNA THAM DIGIT FOUR
95	LANNA THAM DIGIT FIVE
96	LANNA THAM DIGIT SIX
97	LANNA THAM DIGIT SEVEN
98	LANNA THAM DIGIT EIGHT
99	LANNA THAM DIGIT NINE
9A	(This position shall not be used)
9B	(This position shall not be used)
9C	(This position shall not be used)
9D	(This position shall not be used)
9E	(This position shall not be used)
9F	(This position shall not be used)
A0	LANNA SIGN WIANG
A1	LANNA SIGN WIANGWAAK
A2	LANNA SIGN SAWAN
A3	LANNA SIGN KEOW
A4	LANNA SIGN HOY
A5	LANNA SIGN DOKMAI
A6	LANNA SIGN REVERSED ROTATED RANA
A7	LANNA SIGN MAI YAMOK
A8	LANNA SIGN KAAN
A9	LANNA SIGN KAANKUU
AA	LANNA SIGN SATKAAN
AB	LANNA SIGN SATKAANKUU
AC	LANNA SIGN HANG
AD	LANNA SIGN CAANG
AE	(This position shall not be used)
AF	(This position shall not be used)

Group 00 **Plane 00** **Row** 1A

TABLE XX - Row 1A: LANNA (Khün font style)



G = 00
P = 00

Figures

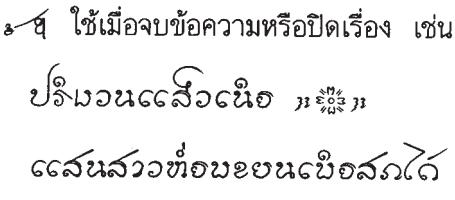


Figure 1. Sample text in Thai and a variety of Lanna script, showing the ᩞ, SIGN DOKMAI flanked by two SIGN KAANKUU characters.

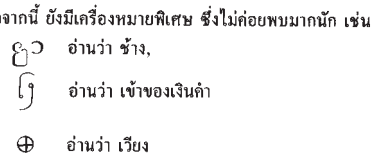


Figure 2. Sample from a reader in Northern Thai, showing CAANG with AA as a special sign for *chāng* ‘elephant’. Also shown are ᧖ THAM DIGIT TWO and ᧒ DIGIT TWO together as a special sign *ngein* ‘silver’, as well as ⊕ SIGN WIANG.

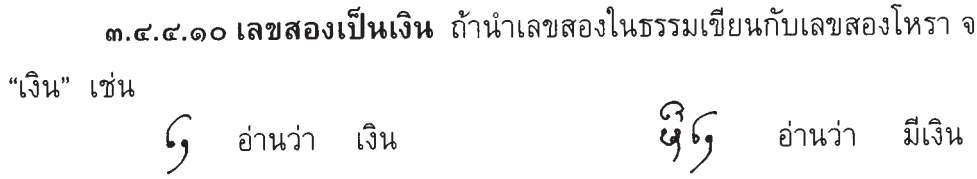


Figure 3. Sample from a reader in Northern Thai, showing ᧖ THAM DIGIT TWO and ᧒ DIGIT TWO together, alongside an example ᨾᧂᧉ *mīngein*.

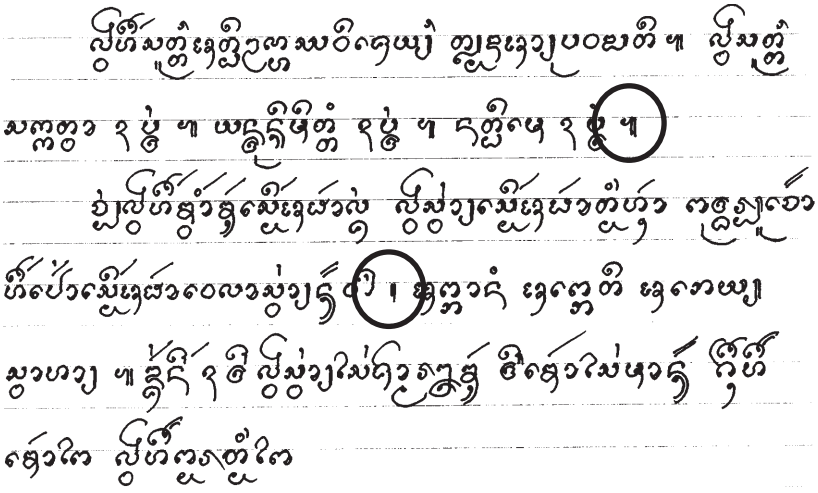


Figure 4. Sample from a reader in Northern Thai, showing five examples of ᩝ SIGN SATKAANKUU, and one example of ᩞ SIGN KAANKUU. Only one example of each is circled.

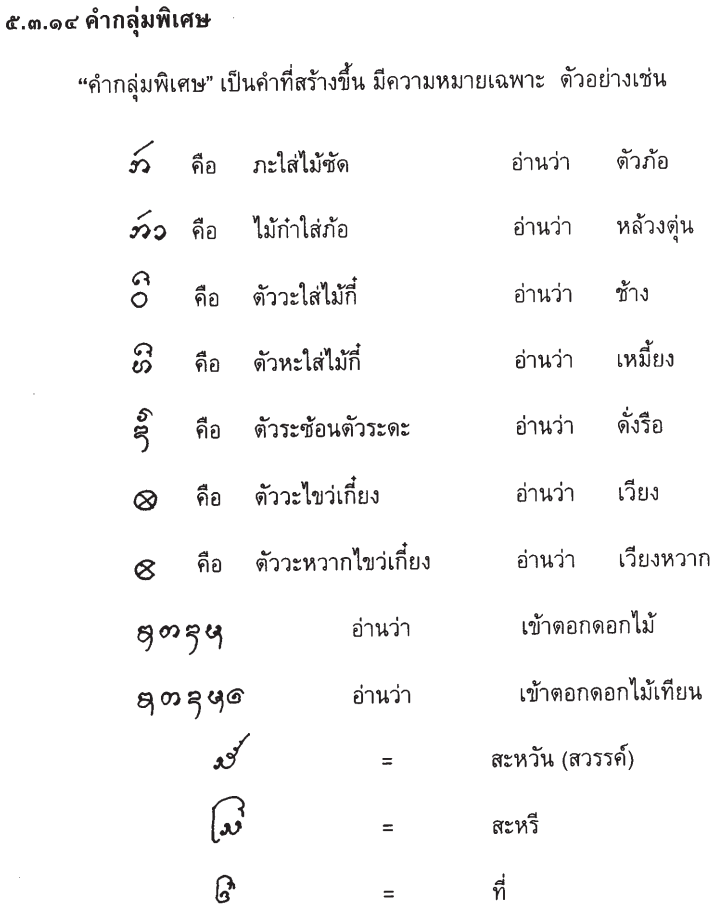


Figure 5. Sample from a reader in Northern Thai, showing ⊕ SIGN WIANG, ⊕ SIGN WIANGWAAK, and ᩮ SIGN SAWAN.

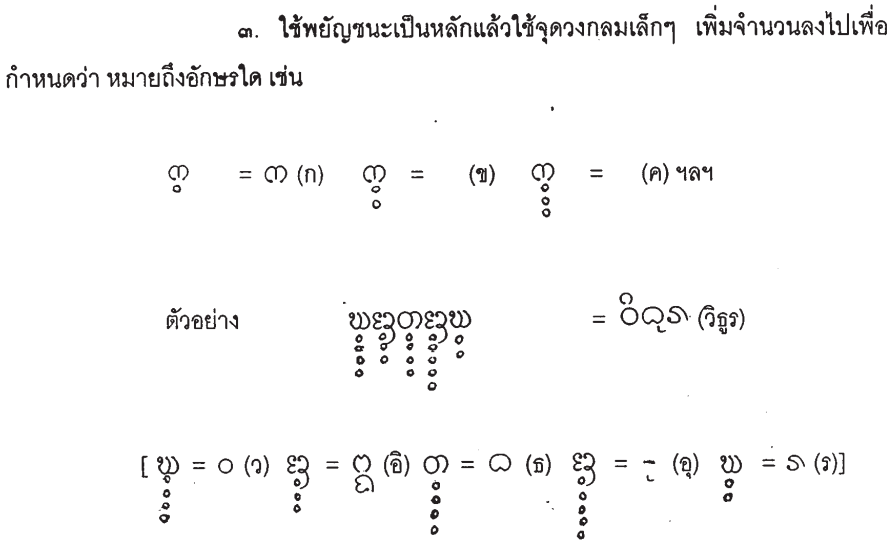
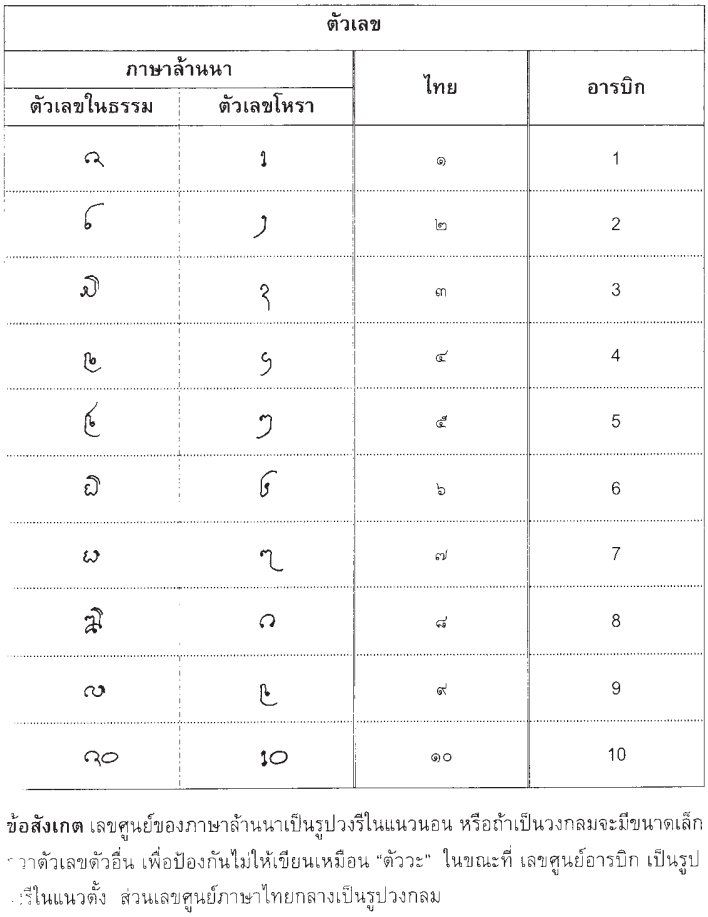


Figure 6. Sample from a reader in Northern Thai, showing examples of ◌᭼ COMBINING CRYPTOGRAMMIC DOT. The number of dots under a letter seem to indicate the number of places in a varga the reader should count to decode the intended letter.



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

ข้อสังเกต เลขศูนย์ของภาษาล้านนาเป็นรูปวงรีในแนวนอน หรือถ้าเป็นวงกลมจะมีขนาดเล็ก
ว่าตัวเลขตัวอื่น เพื่อป้องกันไม่ให้เขียนเหมือน "ตัววะ" ในขณะที่ เลขศูนย์อารบิก เป็นรูป
รีในแนวตั้ง ส่วนเลขศูนย์ภาษาไทยกลางเป็นรูปวงกลม

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

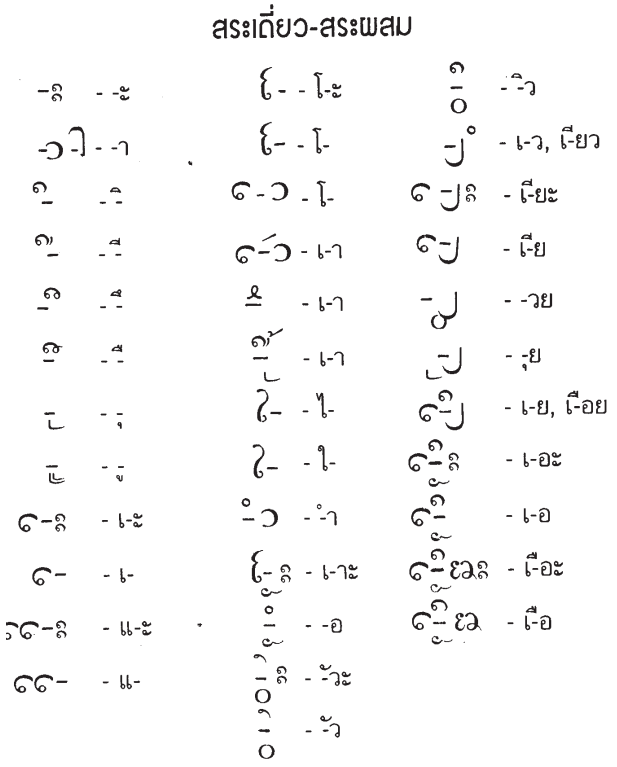
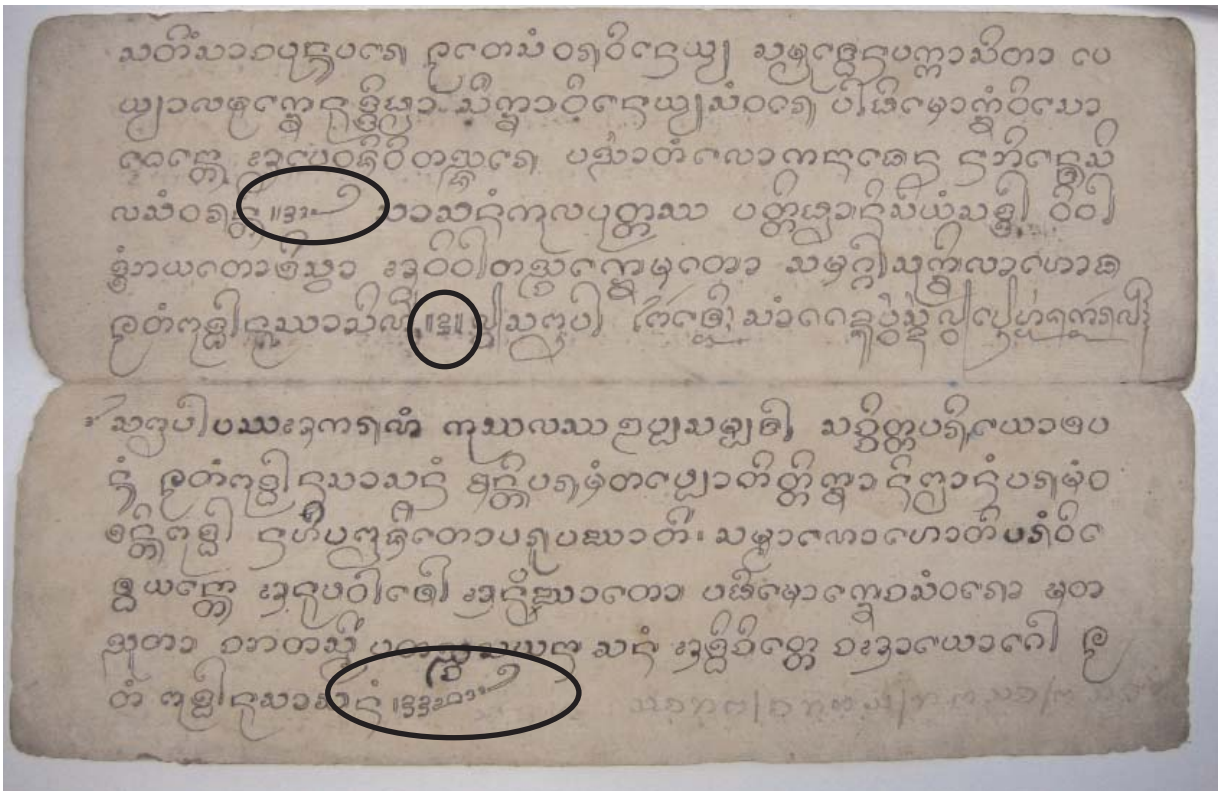


Figure 8. Sample from a reader in Northern Thai, showing dependent vowels.



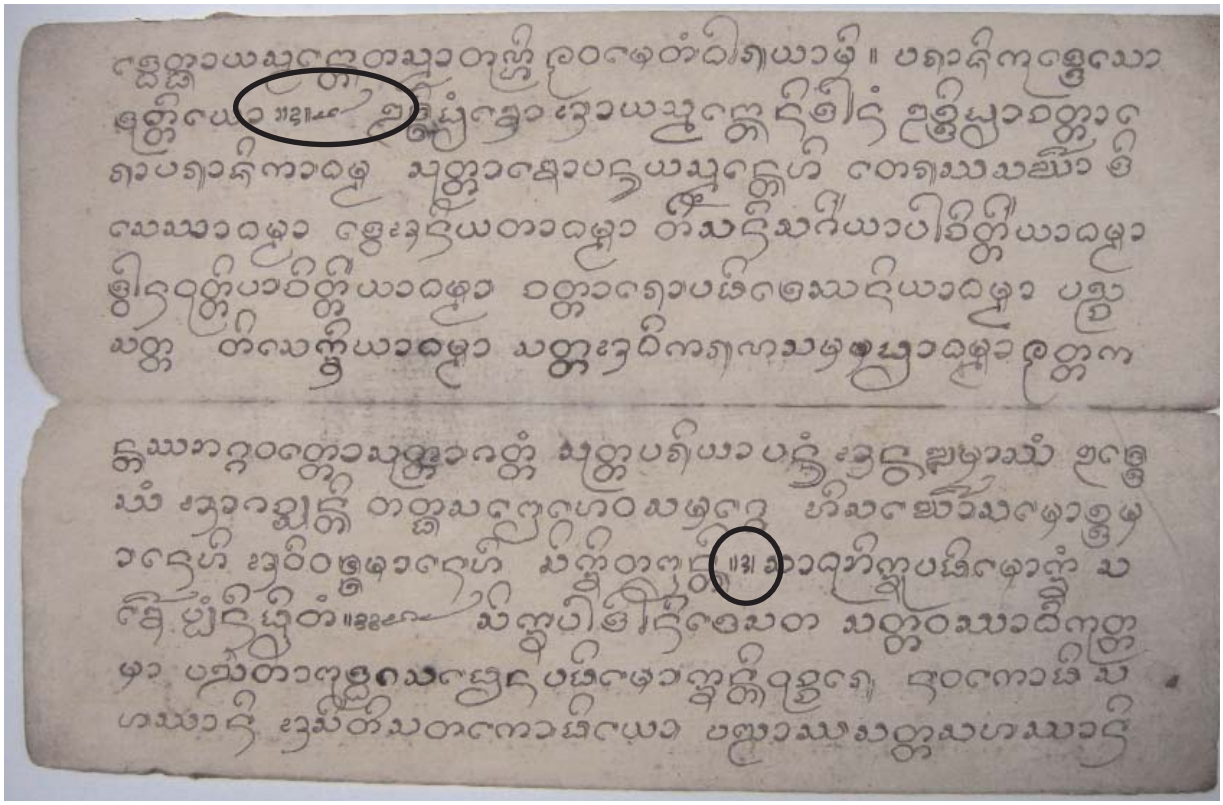
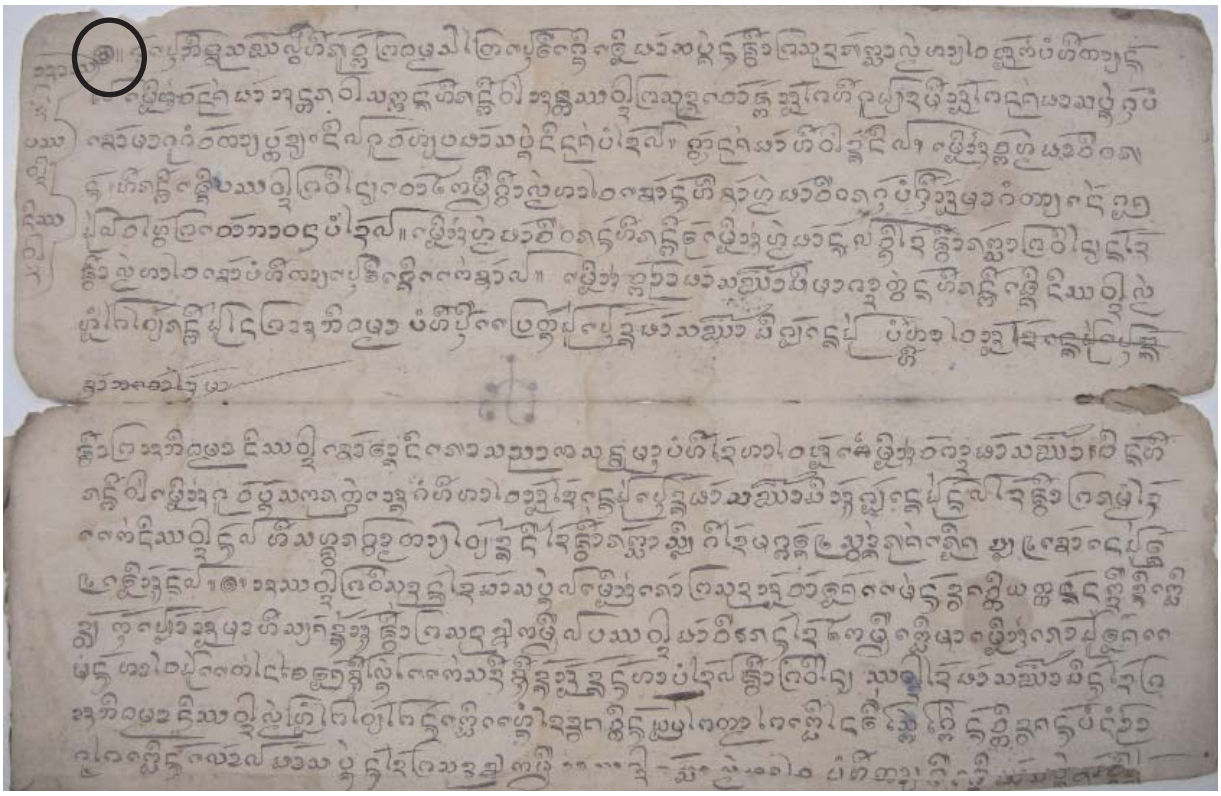


Figure 9a above and 9b below. Sample from a Northern Thai manuscript, showing a variety of marks built up out of individual punctuation marks: ᩞ᩠᩵, ᩞ, ᩞᩞ᩠᩵, ᩞ᩠᩵, ᩞ.



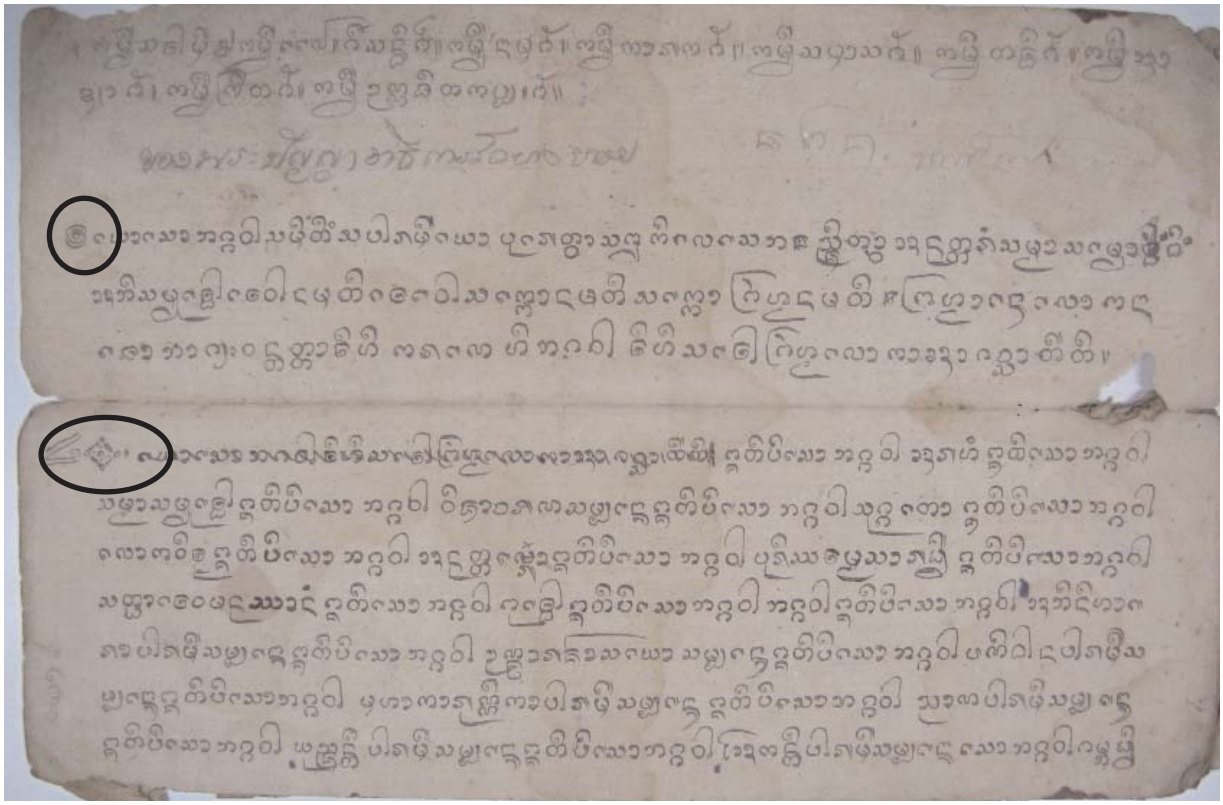


Figure 10a above and 10b below. Sample from a Northern Thai manuscript, showing a variety of marks built up out of individual punctuation marks: ꧁, ๏, and ꧂.

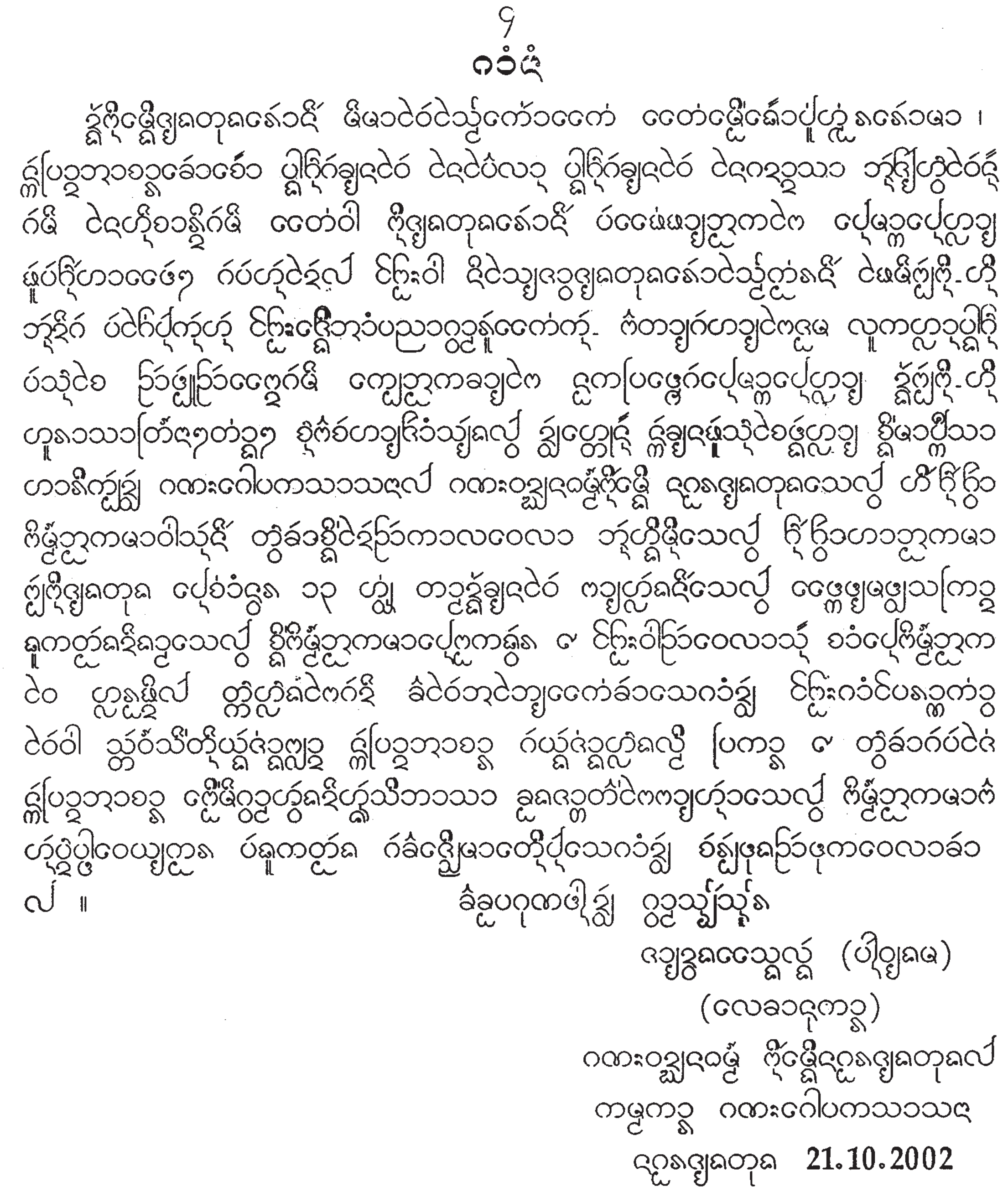


Figure 11. Sample from a reader in Khün.

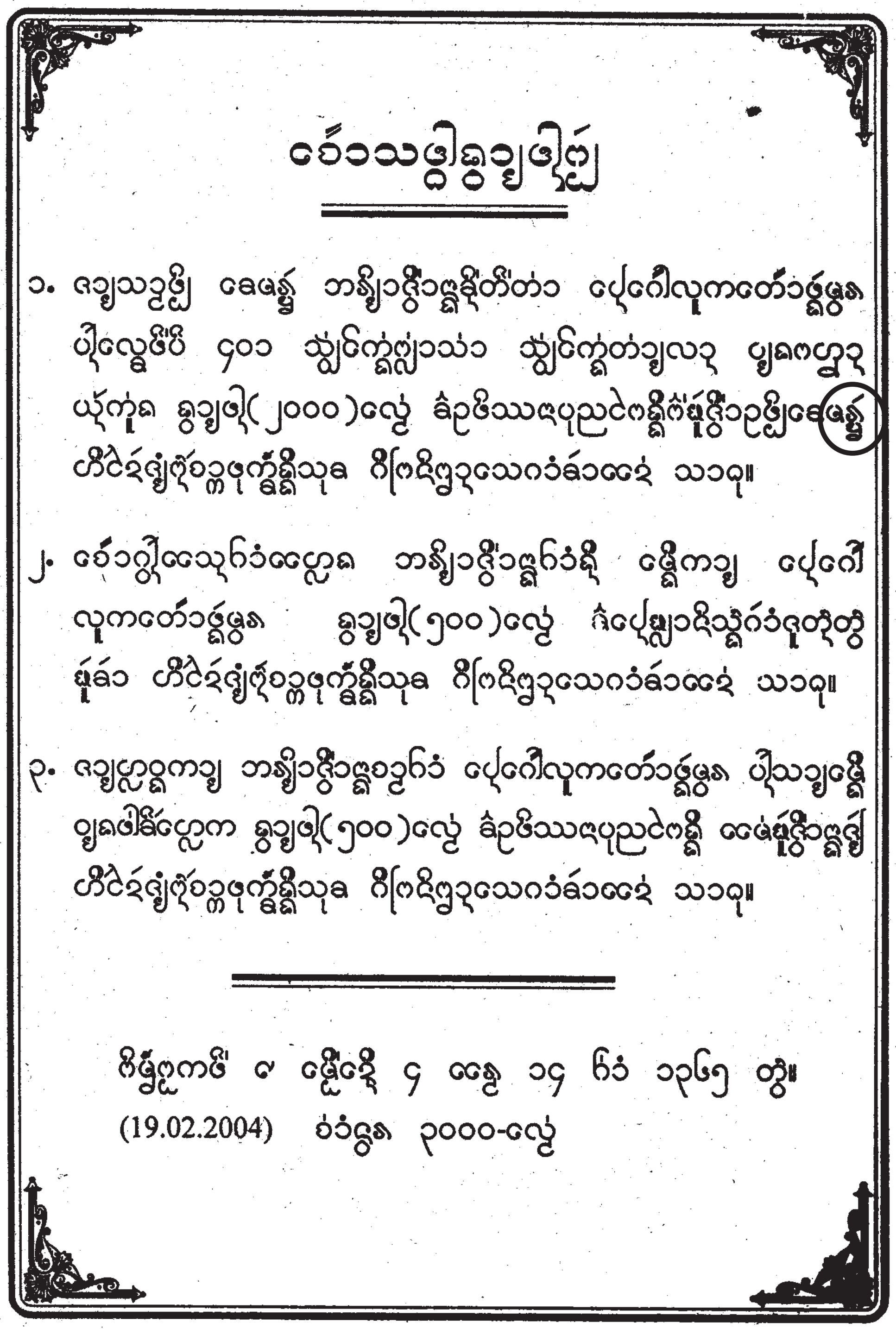


Figure 12. Sample from a reader in Khün. Note that in the circled example, the sequence is 1A44 ᩄ RA + 1A62 ᩢ MAI SAT + 1A60 ᩠ SAKOT + 1A30 ᨰ RATA + 1A5E HIGH RATHA OR LOW PA ᩞ = ᩄᩢ᩠ᨰᩞ; it is not 1A44 ᩄ RA + 1A62 ᩢ MAI SAT + 1A60 ᩠ SAKOT + 1A31 ᨱ HIGH RATHA (= *ᩄᩢ᩠ᨱ) since U+1A31 is not used in Khün.

A. Administrative

1. Title

Revised proposal for encoding the Lanna script in the UCS.

2. Requester's name

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

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

Liaison contribution.

4. Submission date

2007-03-21

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), Richard Wordingham
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?
Yes.
9b. If YES, is a rationale for its inclusion provided?
Yes.
9c. If YES, reference
See page 4 above.
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?