Lanna Unicode: A Proposal

L2/05-095

M. Hosken

Payap University and SIL International

Sociolinguistic Situation

Lanna[1] script is used for three living languages: Northern Thai (or GamMuang), Tai Lue and Khün. In addition the Lanna script is also used for Lao Dham[2] (or old Lao) and other dialect variants in Buddhist palm leaves and notebooks. The script is also known as Dham or Yuan script. There are 6 million speakers of Northern Thai of whom very few are literate in Lanna script, although there is some resurgent interest in the script among the young. There are 500,000 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 100,000 speakers of Khün for which Lanna is the only script.

Script Description

Subjoined Forms

Consonants, in Lanna script, take two forms: the base form (as listed in the code chart) and, for most consonants, a subjoined form. The subjoined form is used for writing medials, finals and for syllable chaining (the initial consonant of a subsequent syllable). U+1A80 LANNA SIGN SUBJOINER is used before a consonant to indicate a subjoined form. Notice also that the character may be used following a vowel, when indicating a final consonant[3]. For example:

ᩉᩣᩯ U+1A84 (LOW KA) U+1A80 (SUBJOINER) U+1AA2 (RA) U+1AB1 (MAI SAT) U+1A80 (SUBJOINER) U+1A98 (BA)

Words written in the Lanna script may sometimes be written in different ways, according to the desire of the author and the spelling school they follow. Thus the decision of when to use a subjoined form and when to use a base form is a spelling convention. For analysis purposes, therefore, a useful search technique is to ignore all U+1A80 LANNA SIGN SUBJOINER characters.

The following table lists the subjoined form for each base character[4].

	1A81	1A82	1A84	1A86	1A87	1A88	1A89	1A8A	1A8C	1A8D	1A8E
base	ᨁ	ᨂ	ᨄ	ᨆ	ᨇ	ᨈ	ᨉ	ᨊ	ᨌ	ᨍ	ᨎ
subjoined	ᩁ	ᩂ	ᩄ	ᩆ	ᩇ	ᩈ	ᩉ	ᩊ	ᩌ	ᩍ	ᩎ

	1A90	1A91	1A92	1A93	1A94	1A95	1A96	1A97	1A98	1A99	1A9A
base	ᨐ	ᨑ	ᨒ	ᨓ	ᨔ	ᨕ	ᨖ	ᨗ	ᨘ	ᨙ	ᨚ
subjoined	ᩐ	ᩑ	ᩒ	ᩓ	ᩔ	ᩕ	ᩖ	ᩗ	ᩘ	ᩙ	ᩚ

	1A9C	1A9F	1AA0	1AA2	1AA3	1AA4	1AA5	1AA6	1AA7	1AA8	1AAE
base	᨜	᨟	ᨠ	ᨡ	ᨢ	ᨣ	ᨤ	ᨥ	ᨦ	ᨧ	ᨮ
subjoined	ᩜ	᩟	᩠	ᩡ/ᩢ	ᩣ/ᩤ	ᩥ	ᩦ	ᩧ	ᩨ	ᩩ	ᩮ

There are two characters each that have two subjoined forms. In each case, the first form listed is the medial form used following a consonant. The second form is used for the final or syllable chained subjoined form of the letter, when not immediately following a consonant.

The subjoined form for U+1A99 LANNA LETTER HIGH PA is used to represent the Pali variant associated with U+1A9C LANNA LETTER LOW PA. The subjoined form is only used with the following characters: U+1A8E LANNA LETTER LATHA, U+1A92 LANNA LETTER LANA, U+1A9C LANNA LETTER LOW PA and U+1A9F LANNA LETTER MA.

Ligatures

There are 3 ligatures used in Lanna script:

U+1AA7 … U+1A80 U+1AA7	ᩧ᩠ᩧ
U+1A97 … U+1AB2	ᨗᩲ
U+1A8D … U+1A80 U+1A8D	ᨍ᩠ᨍ

Note that the ellipsis represents any sequence of diacritics and other subjoined forms which may occur in the ligature sequence. For example:

ᨗᩮᩲ U+1A97 (NA) U+1ABD (MAI KE) U+1AB1 (MAI SAT) U+1AB2 (MAI KAA)

Digits

There are two sets of numbers used in Lanna. The first is listed here in the range U+1AD0 .. U+1AD9. The second is the Myanmar set using the Myanmar block (U+1040 .. U+1049).

Encoding Order

The basic encoding order is[5]:

C (S C)* V1* T? V2* (S C)*

Where C is a consonant or independent vowel (from the range U+1A81 .. U+1AAE, U+1ADA .. U+1ADF); S is the subjoiner and T is a tone mark (from the range U+1AC3 .. U+1AC7). The class of vowels is split into two sets: those which occur before the tone mark and those which occur after. The set V1 consists of all non-spacing vowels and those that are rendered before the syllable (U+1AB1, U+1AB4 .. U+1ABB, U+1ABD .. U+1AC1). The set V2 consists of spacing vowels that follow the tone mark (including U+1AAA LANNA LETTER QA, and U+1A80 LANNA SUBJOINER U+1AA1 LANNA LETTER YA which occur as parts of vowel sequences): U+1AAA, U+1AB0, U+1AB2, U+1AB3, U+1ABC[6]. Vowel sequences from the set V1 are stored in the order left to right, bottom to top. Notice that vowels that are rendered before the initial consonant are stored following the initial consonant cluster as part of V1[7].

Lanna script also uses the subjoined forms of U+1AA1 LANNA LETTER YA and U+1AA4 LANNA LETTER WA as vowels, in some contexts. For example:

ᩃᩮᩴ U+1A82 (HIGH KHA) U+1A80 (SUBJOINER) U+1AA4 (WA) U+1ABB (MAI KONG) U+1AC4 (MAI THO)[8]

ᨽᩮᩣ U+1A9E (LOW PHA) U+1AB4 (MAI KI) U+1AC3 (MAI EK) U+1A80 (SUBJOINER) U+1AA4 (WA)

ᨙᩮᩢ U+1A99 (HIGH PA) U+1ABD (MAI KE) U+1A80 (SUBJOINER) U+1AA1 (YA) U+1AB0 (MAI KA)

ᨠᩮᩬ U+1A84 (LOW KA) U+1ABB (MAI KOH) U+1AC3 (MAI EK) U+1A80 (SUBJOINER) U+1AA1 (YA)

Specific Characters

There are a number of characters in the encoding which have a variety of functions and different forms in different languages.

U+1AB1 LANNA VOWEL MAI SAT ◌᩵

This character is also used for the final k mai kak[9].

U+1AB2 LANNA MARK VOWEL KAA ◌᩶

If this character follows any of the characters U+1A84 (LOW KA), U+1A94 (HIGH TA), U+1A96 (LOW THA), U+1A98 (BA), U+1AA4 (WA) it takes a tall form (◌᩷). One school of Northern Thai spelling only renders a tall form when the word is of Pali origin. In the case where U+1AB2 would normally be rendered using a tall form, U+1AB2 may be followed by U+FE00 VARIANT SELECTOR-1 to inhibit the automatic rendering using the tall form.

U+1AC8 LANNA MARK LAHAAM ◌᩸

This mark has two possible roles:

- Marks that the character or characters it follows are not sounded. The precise range of characters not to be sounded is not defined, although it does not extend beyond one cluster.
- A final –n, in Tai Lue only.

U+1AC8 LANNA MARK LAHAAM is stored at the end of the syllable on which it occurs.

U+1AC9 LANNA MARK MAI SAM ◌᩹

This mark has a number of roles in Northern Thai:

- As a repetition mark. The character is stored following the word to be repeated.
- Used to disambiguate the use of a subjoined form between being a medial or final versus being the start of a new syllable. It is stored following the subjoined form to indicate the consonant being at the start of a new syllable.
- Used to mark double acting consonants. It is stored where the consonant would be stored if there were a separate consonant used.[10]

  ◌᩠◌᩹ U+1A93 (HIGH TA) U+1AC3 (MAI EK) U+1AB2 (MAI KAA) U+1A80 (SUBJOINER) U+1A87 (LOW NGA) U+1AC9 (MAI SAM)

  ◌᩠◌᩹ U+1A94 (HIGH THA) U+1A80 (SUBJOINER) U+1A97 (NA) U+1AC9 (MAI SAM) U+1ABA (MAI KONG) U+1AA2 (RA)

  ◌᩠◌᩹ U+1A82 (HIGH KHA) U+1ABD (MAI KE) U+1AB1 (MAI SAT) U+1AC4 (MAI THO) U+1AC9 (MAI SAM) U+1ABB (MAI KOH) U+1A87 (LOW NGA)

U+1ACA LANNA MARK MAI KANG ◌᩺

This mark has four meanings:

- Following U+1ABC LANNA VOWEL MAI KER, it acts as part of a vowel sequence.
- Following U+1AB5 LANNA VOWEL MAI KI it represents –ing, and following U+1AB9 LANNA VOWEL MAI KU it represents –ung.
- In conjunction with a following U+1AB2 LANNA VOWEL MAI KAA it represents the final –am. But this sequence is represented by U+1AB3 LANNA VOWEL MAI KAM[11] and the sequence should never be used.
- In all other cases it is a final –ang.

U+1ACB LANNA MARK MAI KANG LAI ◌᩻

This character is a final –ang. It takes a different form (◌ᩮ) in Khün.

U+1ACC LANNA LUE LETTER FINAL NGA ◌᩼

In Tai Lue, the final nga is rendered above the syllable, as shown.

U+1ACD LANNA KHUN LETTER FINAL NGA ◌᩽

In Khün the form of U+1A80 LANNA SIGN SUBJOINER U+1A87 LANNA LETTER LOW NGA differs from that used in U+1ACD LANNA LETTER FINAL NGA and both characters occur in the same text. The rendering behaviour of U+1ACD is for the character to subjoin if there is space below the base

character, otherwise it takes a full base character form. This character is the same as U+1ACC LANNA LETTER FINAL NGA but with radically different behaviour.

U+1ACE LANNA LUE LETTER FINAL NA ᳎

This character is a rare character used only in Tai Lue. It is stored before the consonant it occurs over since it is a final of the previous syllable.

U+1ACF LANNA LUE LETTER FINAL PA ᳏

Like U+1ACE LANNA LUE LETTER FINAL NA, this rare character is used only in Tai Lue and is stored before the consonant it occurs over.

Bibliography

บุญคิด วัชรศาสตร์, 1995 *คนเมือง อู้คำเมือง* (เชียงใหม่: ธาราทองการพิมพ์)
(Boonkit Wacharasaat, 1995 *Khon Muang Oo Kham Muang* (Chiang Mai: Taratong Garn Pim))

บุญคิด วัชรศาสตร์, 2005 *ภาษาเมืองล้านนา* (เชียงใหม่: ธาราทองการพิมพ์)
(Boonkit Wacharasaat, 2005 *Pasa Muang Lanna* (Chiang Mai: Taratong Garn Pim))

ผู้ช่วยศาสตราจารย์ ลมูล จันทน์หอม, 1999 *อักษรล้านนา* (เชียงใหม่)
(Asst. Prof. Lamun Janhom, 1999 *Ackson Lanna* (Chiang Mai))

ศาสตราจารย์ ดร.อุดม รุ่งเรืองศรี, 2547 *พจนานุกรมล้านนา-ไทย: ฉบับแม่ฟ้าหลวง*
(Prof. Dr Udom Rungrueangsri, 2004 *Potjananugrom Lanna-Thai: Chababmaefaluang, 2nd Edition)*

สถาบันราชภัฏเชียงใหม่, 1996 *พจนานุกรมภาษาภาคเหนือ* (เชียงใหม่: สถาบันราชภัฏเชียงใหม่)
(Chiangmai Rajabhat Institute, 1996 *Potjananugrom Pasaa Paknua* (ChiangMai: Sataban Rajabhat Chiangmai))

อินสม ไชยชมภู, 1991 *เรียนหนังสือล้านนา ด้วยตนเอง* (เชียงราย: ร้านบุญผดุง)
(Insom Chayshompu, 1991 *Rian Nangsue Lanna Duay Ton Eng* (Chiang Rai: Raan Bunpadung))

Davis, Richard, c. 1970 *A Northern Thai Reader* (Bangkok: The Siam Society)

Peltier, Anatole-Roger, 1996 *Khun Reader* (Chiang Mai)

Sai Kam Mong, 2004 *The History and Development of the Shan Scripts* (Chiang Mai: Silkworm Books)

1A80 **Lanna** **1AEF**

	1A8	1A9	1AA	1AB	1AC	1AD	1AE
0	◌ 1A80	᪐ 1A90	᪠ 1AA0	᫰ 1AB0	◌ᩰ 1AC0	᪰ 1AD0	
1	᪁ 1A81	᪑ 1A91	᪡ 1AA1	◌᫁ 1AB1	◌ᩱ 1AC1	᪱ 1AD1	
2	᪂ 1A82	᪒ 1A92	᪢ 1AA2	᫲ 1AB2	◌ᩲ 1AC2	᪲ 1AD2	
3	᪃ 1A83	᪓ 1A93	᪣ 1AA3	◌᫳ 1AB3	◌ᩳ 1AC3	᪳ 1AD3	
4	᪄ 1A84	᪔ 1A94	᪤ 1AA4	◌᫴ 1AB4	◌ᩴ 1AC4	᪴ 1AD4	
5	᪅ 1A85	᪕ 1A95	᪥ 1AA5	◌᫵ 1AB5	◌᩵ 1AC5	᪵ 1AD5	
6	᪆ 1A86	᪖ 1A96	᪦ 1AA6	◌᫶ 1AB6	◌᩶ 1AC6	᪶ 1AD6	
7	᪇ 1A87	᪗ 1A97	ᪧ 1AA7	◌᫷ 1AB7	◌᩷ 1AC7	᪷ 1AD7	
8	᪈ 1A88	᪘ 1A98	᪨ 1AA8	◌᫸ 1AB8	◌᩸ 1AC8	᪸ 1AD8	
9	᪉ 1A89	᪙ 1A99	᪩ 1AA9	◌᫹ 1AB9	◌᩹ 1AC9	᪹ 1AD9	
A	᪊ 1A8A	᪚ 1A9A	᪪ 1AAA	◌᫺ 1ABA	◌᩺ 1ACA	᪺ 1ADA	
B	᪋ 1A8B	᪛ 1A9B	᪫ 1AAB	◌᫻ 1ABB	◌᩻ 1ACB	᪻ 1ADB	
C	᪌ 1A8C	᪜ 1A9C	᪬ 1AAC	◌᫼ 1ABC	◌᩼ 1ACC	᪼ 1ADC	
D	᪍ 1A8D	᪝ 1A9D	᪭ 1AAD	◌᫽ 1ABD	◌᩽ 1ACD	᪽ 1ADD	
E	᪎ 1A8E	᪞ 1A9E	᪮ 1AAE	◌᫾ 1ABE	◌᩾ 1ACE	᪾ 1ADE	
F	᪏ 1A8F	᪟ 1A9F		◌᫿ 1ABF	◌᩿ 1ACF	ᪿ 1ADF	

Consonants

1A80 ◌ LANNA SIGN SUBJOINER
- not rendered, no shape

1A81 ᨡ LANNA LETTER HIGH KA
1A82 ᨢ LANNA LETTER HIGH KHA
1A83 ᨣ LANNA LETTER HIGH KHAA
1A84 ᨤ LANNA LETTER LOW KA
1A85 ᨥ LANNA LETTER LOW KHA
1A86 ᨦ LANNA LETTER LOW KHAA
1A87 ᨧ LANNA LETTER LOW NGA
1A88 ᨨ LANNA LETTER HIGH JA
1A89 ᨩ LANNA LETTER HIGH SA
1A8A ᨪ LANNA LETTER LOW JA
1A8B ᨫ LANNA LETTER LOW SA
1A8C ᨬ LANNA LETTER LOW SAA
1A8D ᨭ LANNA LETTER HIGH NYA
1A8E ᨮ LANNA LETTER LATA
1A8F ᨯ LANNA LETTER HIGH LATA
1A90 ᨰ LANNA LETTER DA
1A91 ᨱ LANNA LETTER LOW LATA
1A92 ᨲ LANNA LETTER LANA
1A93 ᨳ LANNA LETTER HIGH TA
1A94 ᨴ LANNA LETTER HIGH THA
1A95 ᨵ LANNA LETTER LOW TA
1A96 ᨶ LANNA LETTER LOW THA
1A97 ᨷ LANNA LETTER NA
1A98 ᨸ LANNA LETTER BA
1A99 ᨹ LANNA LETTER HIGH PA
1A9A ᨺ LANNA LETTER HIGH PHA
1A9B ᨻ LANNA LETTER HIGH FA
1A9C ᨼ LANNA LETTER LOW PA
1A9D ᨽ LANNA LETTER LOW FA
1A9E ᨾ LANNA LETTER LOW PHA
1A9F ᨿ LANNA LETTER MA
1AA0 ᩀ LANNA LETTER LOW NYA
1AA1 ᩁ LANNA LETTER YA
1AA2 ᩂ LANNA LETTER RA
1AA3 ᩃ LANNA LETTER LA
1AA4 ᩄ LANNA LETTER WA
1AA5 ᩅ LANNA LETTER HIGH SAA
1AA6 ᩆ LANNA LETTER HIGH SAA2
1AA7 ᩇ LANNA LETTER HIGH SAA3
1AA8 ᩈ LANNA LETTER HIGH HA
1AA9 ᩉ LANNA LETTER LAA
1AAA ᩊ LANNA LETTER QA
1AAB ᩋ LANNA LETTER LOW HA
1AAC ᩌ LANNA LETTER RU
1AAD ᩍ LANNA LETTER LU
1AAE ᩎ LANNA LETTER LAE

Dependent Vowel Signs

1AB0 ◌ᩫ LANNA VOWEL MAI KA
1AB1 ◌ᩬ LANNA VOWEL MAI SAT
1AB2 ◌ᩭ LANNA VOWEL MAI KAA
1AB3 ◌ᩮ LANNA VOWEL MAI KAM
1AB4 ◌ᩯ LANNA VOWEL MAI KI
1AB5 ◌ᩰ LANNA VOWEL MAI KII
1AB6 ◌ᩱ LANNA VOWEL MAI KUE
1AB7 ◌ᩲ LANNA VOWEL MAI KUUE
1AB8 ◌ᩳ LANNA VOWEL MAI KU
1AB9 ◌ᩴ LANNA VOWEL MAI KUU
1ABA ◌᩵ LANNA VOWEL MAI KONG
1ABB ◌᩶ LANNA VOWEL MAI KOH
1ABC ◌᩷ LANNA VOWEL MAI KOY
1ABD ◌᩸ LANNA VOWEL MAI KE
- stands to the left of the consonant

1ABE ◌᩹ LANNA VOWEL MAI KEE
- stands to the left of the consonant

1ABF ◌᩺ LANNA VOWEL MAI KO
- stands to the left of the consonant

1AC0 ◌᩻ LANNA VOWEL MAI KAI
- stands to the left of the consonant

1AC1 ◌᩼ LANNA VOWEL MAI KAI2
- stands to the left of the consonant

1AC2 ◌᩽ LANNA VOWEL MAI KAW

Tone Marks

1AC3 ◌᩾ LANNA TONE MAI EK
1AC4 ◌᩿ LANNA TONE MAI THO
1AC5 ◌᪀ LANNA KHUN[12] TONE THREE
1AC6 ◌᪁ LANNA KHUN TONE FOUR
1AC7 ◌᪂ LANNA KHUN TONE FIVE

Various Signs

1AC8 ◌᪃ LANNA MARK LAHAAM
- cancellation mark

1AC9 ◌᪄ LANNA MARK MAI SAM
- repetition mark

Final Consonant Signs

1ACA ◌᪅ LANNA MARK MAI KANG
1ACB ◌᪆ LANNA MARK MAI KANG LAI
1ACC ◌᪇ LANNA LETTER FINAL NGA
1ACD ◌᪈ LANNA KHUN LETTER FINAL NGA
1ACE ◌᪉ LANNA LUE LETTER FINAL NA
1ACF ◌᪊ LANNA LUE LETTER FINAL PA

Digits

1AD0 ᪐ LANNA DIGIT ZERO
1AD1 ᪑ LANNA DIGIT ONE
1AD2 ᪒ LANNA DIGIT TWO
1AD3 ᪓ LANNA DIGIT THREE

1AD4 ᪔ LANNA DIGIT FOUR

1AD5 ᪕ LANNA DIGIT FIVE

1AD6 ᪖ LANNA DIGIT SIX

1AD7 ᪗ LANNA DIGIT SEVEN

1AD8 ᪘ LANNA DIGIT EIGHT

1AD9 ᪙ LANNA DIGIT NINE

Independent Vowels

1ADA ᪚ LANNA VOWEL LETTER I

1ADB ᪛ LANNA VOWEL LETTER II

1ADC ᪜ LANNA VOWEL LETTER U

1ADD ᪝ LANNA VOWEL LETTER UU

1ADE ᪞ LANNA VOWEL LETTER EE

1ADF ᪟ LANNA KHUN VOWEL LETTER OO

Character Database

```
1A80;LANNA SIGN SUBJOINER;Cf;0;BN;;;;N;;;;;
1A81;LANNA LETTER HIGH KA;Lo;0;L;;;;;N;;;;;
1A82;LANNA LETTER HIGH KHA;Lo;0;L;;;;;N;;;;;
1A83;LANNA LETTER HIGH KHAA;Lo;0;L;;;;;N;;;;;
1A84;LANNA LETTER LOW KA;Lo;0;L;;;;;N;;;;;
1A85;LANNA LETTER LOW KHA;Lo;0;L;;;;;N;;;;;
1A86;LANNA LETTER LOW KHAA;Lo;0;L;;;;;N;;;;;
1A87;LANNA LETTER LOW NGA;Lo;0;L;;;;;N;;;;;
1A88;LANNA LETTER HIGH JA;Lo;0;L;;;;;N;;;;;
1A89;LANNA LETTER HIGH SA;Lo;0;L;;;;;N;;;;;
1A8A;LANNA LETTER LOW JA;Lo;0;L;;;;;N;;;;;
1A8B;LANNA LETTER LOW SA;Lo;0;L;;;;;N;;;;;
1A8C;LANNA LETTER LOW SAA;Lo;0;L;;;;;N;;;;;
1A8D;LANNA LETTER HIGH NYA;Lo;0;L;;;;;N;;;;;
1A8E;LANNA LETTER LATA;Lo;0;L;;;;;N;;;;;
1A8F;LANNA LETTER HIGH LATA;Lo;0;L;;;;;N;;;;;
1A90;LANNA LETTER DA;Lo;0;L;;;;;N;;;;;
1A91;LANNA LETTER LOW LATA;Lo;0;L;;;;;N;;;;;
1A92;LANNA LETTER LANA;Lo;0;L;;;;;N;;;;;
1A93;LANNA LETTER HIGH TA;Lo;0;L;;;;;N;;;;;
1A94;LANNA LETTER HIGH THA;Lo;0;L;;;;;N;;;;;
1A95;LANNA LETTER LOW TA;Lo;0;L;;;;;N;;;;;
1A96;LANNA LETTER LOW THA;Lo;0;L;;;;;N;;;;;
1A97;LANNA LETTER NA;Lo;0;L;;;;;N;;;;;
1A98;LANNA LETTER BA;Lo;0;L;;;;;N;;;;;
1A99;LANNA LETTER HIGH PA;Lo;0;L;;;;;N;;;;;
1A9A;LANNA LETTER HIGH PHA;Lo;0;L;;;;;N;;;;;
1A9B;LANNA LETTER HIGH FA;Lo;0;L;;;;;N;;;;;
1A9C;LANNA LETTER LOW PA;Lo;0;L;;;;;N;;;;;
1A9D;LANNA LETTER LOW FA;Lo;0;L;;;;;N;;;;;
1A9E;LANNA LETTER LOW PHA;Lo;0;L;;;;;N;;;;;
1A9F;LANNA LETTER MA;Lo;0;L;;;;;N;;;;;
1AA0;LANNA LETTER LOW NYA;Lo;0;L;;;;;N;;;;;
1AA1;LANNA LETTER YA;Lo;0;L;;;;;N;;;;;
1AA2;LANNA LETTER RA;Lo;0;L;;;;;N;;;;;
1AA3;LANNA LETTER LA;Lo;0;L;;;;;N;;;;;
1AA4;LANNA LETTER WA;Lo;0;L;;;;;N;;;;;
1AA5;LANNA LETTER HIGH SAA;Lo;0;L;;;;;N;;;;;
1AA6;LANNA LETTER HIGH SAA2;Lo;0;L;;;;;N;;;;;
1AA7;LANNA LETTER HIGH SAA3;Lo;0;L;;;;;N;;;;;
1AA8;LANNA LETTER HIGH HA;Lo;0;L;;;;;N;;;;;
1AA9;LANNA LETTER LAA;Lo;0;L;;;;;N;;;;;
1AAA;LANNA LETTER QA;Lo;0;L;;;;;N;;;;;
1AAB;LANNA LETTER LOW HA;Lo;0;L;;;;;N;;;;;
1AAC;LANNA LETTER RU;Lo;0;L;;;;;N;;;;;
1AAD;LANNA LETTER LU;Lo;0;L;;;;;N;;;;;
1AAE;LANNA LETTER LAE;Lo;0;L;;;;;N;;;;;

1AB0;LANNA VOWEL MAI KA;Lo;0;L;;;;;N;;;;;
1AB1;LANNA VOWEL MAI SAT;Mn;0;L;;;;;N;;;;;
1AB2;LANNA VOWEL MAI KAA;Lo;0;L;;;;;N;;;;;
1AB3;LANNA VOWEL MAI KAM;Mn;0;L;<compat>1AB2
                1ACA;;;;N;;;;;
1AB4;LANNA VOWEL MAI KI;Mn;0;L;;;;;N;;;;;
1AB5;LANNA VOWEL MAI KII;Mn;0;L;;;;;N;;;;;
1AB6;LANNA VOWEL MAI KUE;Mn;0;L;;;;;N;;;;;
1AB7;LANNA VOWEL MAI KUUE;Mn;0;L;;;;;N;;;;;
1AB8;LANNA VOWEL MAI KU;Mn;0;L;;;;;N;;;;;
1AB9;LANNA VOWEL MAI KUU;Mn;0;L;;;;;N;;;;;
1ABA;LANNA VOWEL MAI KONG;Mn;0;L;;;;;N;;;;;
1ABB;LANNA VOWEL MAI KOH;Mn;0;L;;;;;N;;;;;
1ABC;LANNA VOWEL MAI KOY;Mn;0;L;<compat>1ABA 1A80
       1AA1;;;;N;;;;;
1ABD;LANNA VOWEL MAI KE;Mc;0;L;;;;;N;;;;;
1ABE;LANNA VOWEL MAI KEE;Mc;0;L;;;;;N;;;;;
1ABF;LANNA VOWEL MAI KO;Mc;0;L;;;;;N;;;;;
1AC0;LANNA VOWEL MAI KAI;Mc;0;L;;;;;N;;;;;
1AC1;LANNA VOWEL MAI KAI2;Mc;0;L;;;;;N;;;;;
1AC2;LANNA VOWEL MAI KAW;Mn;0;L;<compat>1ABD 1AB3
       1AB4;;;;N;;;;;
1AC3;LANNA TONE MAI EK;Mn;0;L;;;;;N;;;;;
1AC4;LANNA TONE MAI THO;Mn;0;L;;;;;N;;;;;
1AC5;LANNA KHUN TONE THREE;Mn;0;L;;;;;N;;;;;
1AC6;LANNA KHUN TONE FOUR;Mn;0;L;;;;;N;;;;;
1AC7;LANNA KHUN TONE FIVE;Mn;0;L;;;;;N;;;;;
1AC8;LANNA MARK LAHAAM;Mn;0;L;;;;;N;;;;;
1AC9;LANNA MARK MAI SAM;Mn;0;L;;;;;N;;;;;
1ACA;LANNA MARK MAI KANG;Mn;0;L;;;;;N;;;;;
1ACB;LANNA MARK MAI KANG LAI;Mn;0;L;;;;;N;;;;;
1ACC;LANNA LETTER FINAL NGA;Mn;0;L;;;;N;;;;;
1ACD;LANNA KHUN LETTER FINAL NGA;Mn;0;L;<compat>
       1ACC;;;;N;;;;;
1ACE;LANNA LUE LETTER FINAL NA;Mn[13];0;L;<compat>
       1A92 1A80;;;;N;;;;;
1ACF;LANNA LUE LETTER FINAL PA;Mn;0;L;<compat>
       1A9C 1A80;;;;N;;;;;
1AD0;LANNA DIGIT ZERO;Nd;0;L;;0;0;0;N;;;;;
1AD1;LANNA DIGIT ONE;Nd;0;L;;1;1;1;N;;;;;
1AD2;LANNA DIGIT TWO;Nd;0;L;;2;2;2;N;;;;;
1AD3;LANNA DIGIT THREE;Nd;0;L;;3;3;3;N;;;;;
1AD4;LANNA DIGIT FOUR;Nd;0;L;;4;4;4;N;;;;;
1AD5;LANNA DIGIT FIVE;Nd;0;L;;5;5;5;N;;;;;
1AD6;LANNA DIGIT SIX;Nd;0;L;;6;6;6;N;;;;;
1AD7;LANNA DIGIT SEVEN;Nd;0;L;;7;7;7;N;;;;;
1AD8;LANNA DIGIT EIGHT;Nd;0;L;;8;8;8;N;;;;;
1AD9;LANNA DIGIT NINE;Nd;0;L;;9;9;9;N;;;;;
1ADA;LANNA VOWEL LETTER I;Lo;0;L;;;;;N;;;;;
1ADB;LANNA VOWEL LETTER II;Lo;0;L;;;;;N;;;;;
1ADC;LANNA VOWEL LETTER U;Lo;0;L;;;;;N;;;;;
1ADD;LANNA VOWEL LETTER UU;Lo;0;L;;;;;N;;;;;
1ADE;LANNA VOWEL LETTER EE;Lo;0;L;;;;;N;;;;;
1ADF;LANNA KHUN VOWEL LETTER OO;Lo;0;L;<compat>
       1AAA 1ABF;;;;N;;;;;
```

Proposal Form

A. Administrative

1. **Title:** ___**Lanna**_____
2. Requester's name: ___Martin Hosken_____
3. Requester type (Member body/Liaison/Individual contribution): __Individual contribution_____
4. Submission date: ___21 April 2005____
5. Requester's reference (if applicable): _____
6. (Choose one of the following:)

This is a complete proposal: ____yes_____

or, More information will be provided later: _____

B. Technical - General

1. (Choose one of the following:)
   a. This proposal is for a new script (set of characters): _yes_____
      Proposed name of script: _____Lanna_____
   . b. The proposal is for addition of character(s) to an existing block: _____
      Name of the existing block: _____
2. Number of characters in proposal: __95_____
3. Proposed category (see section II, Character Categories): _A_____
4. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000): ____3__
   Is a rationale provided for the choice? ____no_____
      If Yes, reference: _____
5. Is a repertoire including character names provided? ____yes_____
   a. If YES, are the names in accordance with the 'character naming guidelines in Annex L of ISO/IEC 10646-1: 2000? ____yes_____
   b. Are the character shapes attached in a legible form suitable for review? __yes___
6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? _____Martin Hosken_____

   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: ____martin_hosken@sil.org_____
7. References:
   a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? _yes_
   b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? _see bibliography_____
8. Special encoding issues:

   Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

   _____presentation, sorting, searching_____

9. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? __Yes_____

   If YES explain ___L2/04-351. This is a final proposal based on that document_____
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? ___yes_____

If YES, with whom? ___Thai user communities of N.Thai, Lue & Khün._____
Particularly with Dr. Udom Rungrueangsri of ChiangMai University, a foremost expert in the script.

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? _____yes_____

Reference: _____

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

Reference: _____

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

If YES, where? Reference: ___See bibliography post 1990_____

6. After giving due considerations to the principles in *Principles and Procedures document* (a WG 2 standing document) must the proposed characters be entirely in the BMP? ____yes_____

If YES, is a rationale provided? __Already roadmapped for BMP__

If YES, reference: _____

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

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

If YES, is a rationale for its inclusion provided? ____yes_____

If YES, reference: _____

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

If YES, is a rationale for its inclusion provided? ____yes_____

If YES, reference: _____

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

If YES, is a rationale for its inclusion provided? _____

If YES, reference: _____

11. 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_____

If YES, is a rationale for such use provided? _____yes_____

If YES, reference: _____

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

If YES, reference: _____

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

If YES, describe in detail (include attachment if necessary) ____included___

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

If YES, is the equivalent corresponding unified ideographic character(s) identified? _____

If YES, reference: _____

Notes

[1] A script name of Dham is suggested as being the least political, giving no particular emphasis to one language group over another. But the common English language name for this script is widely accepted to be Lanna, and the association between the term Lanna and Northern Thai is not so strong as to exclude the other languages.

[2] While care has been taken to ensure that this proposal is not incompatible with Lao Dham, a complete analysis of that writing system has not been made and a further proposal may well be required to complete the needs for that language. An extra column (1AEx) is suggested as being sufficient for any additions that may be needed in the future.

[3] Lanna differs from other Indic scripts in that it has no visual representation for killing the inherent vowel. In fact the use of the inherent vowel in Lanna is rare and phonetically occurs only when declustering an initial consonant cluster. Therefore a true virama has little meaning in Lanna script. Its only use would be to indicate a possible subjoined relationship that may follow. In addition, Lanna makes heavy use of subjoined final consonants, not found in other Indic scripts. Using a strict virama/halant model for Lanna, therefore is problematic, due to the lack of visual representation of the presence of such a killer. For example, consider the sequence U+1A81 (HIGH KA) U+1ABA (MAI KONG) U+1A81 (HIGH KA) (VIRAMA) U+1AA2 (RA) U+1AB2 (MAI KAA) U+1A98 (BA) (VIRAMA) could either be rendered as: ᨠᩫᨠ᩠ᩁᩣ᩠ᨷ or as ᨠᩫ᩠ᨠᩁᩣ᩠ᨷ which are both legal Northern Thai word sequences. It is clear that some other mechanism that gives a more precise indication of the subjoining relationship is needed. Other problems that arise if no clear subjoining relationship is marked are: complexity of rendering rules for deciding when a marked final is to be rendered subjoined or as a full base character and lack of user feedback when inserting a virama after a non-subjoined final character, and therefore the likelihood of the virama not being added, resulting in misspelling.

Therefore we introduce the somewhat artificial notion of a subjoining letter. In effect, this code merely saves having explicit codepoints for each subjoined form (cf Tibetan). Implementation would be simplified if such an approach of encoding each subjoined form were followed. On the other hand, due to the variety of spellings used when writing using the Lanna script, having an easy mechanism to relate base form and subjoined form can aid analysis. For example, it is possible to ignore the U+1A80 (SUBJOINER) particularly when searching.

[4] Those listed here are those with clear textual attestation. There is still debate over whether subjoined forms exist for other characters, for example U+1A9E (LOW PHA).

[5] Following regular expression conventions, ? indicates an optional item, * indicates zero or more items.

[6] It is tempting to desire that the tone mark occur after all the vowels. Unfortunately taking such an approach results in possible ambiguities regarding U+1AAA being at the end of one syllable or at the start of the next, and consequently where a tone mark should be rendered. For example, whether ᨯᩮᩋ᩶ᩡ (2 syllables) is rendered ᨯᩮ᩶ᩋᩡ (1 syllable) is dependent purely on deep knowledge of the vowel system in Northern Thai. If tones always followed vowels, both sequences would be identically encoded as: U+1A90 (DA) U+1ABD (MAI KE) U+1AAA (QA) U+1AB0 (MAI KA) U+1AC4 (MAI THO). Treating U+1AAA LANNA LETTER QA as a final is not an option since it may be followed by U+1AB0 LANNA VOWEL MAI KA, thus forming a complete syllable (as in the first of the two examples shown) rather than being part of the previous syllable (as the second example shows).

[7] User preference is usually to type pre-vowels before the initial consonant. Storing in such a position introduces significant problems when it comes to sorting. For example: ᩈ᩠ᩅᩮ᩵ may be transliterated as either [swe] `push` or [sew] `insert` depending on which word is being used. The two words sort into different positions since the vowel either occurs before the [w] or after. Storing the vowel after the initial consonant cluster (after or before the [w] depending on the word) resolves this ambiguity, but presents a difficulty in implementation. How is someone who wishes to type an initial [e] before the [s] to indicate whether the vowel is to be stored before or after the following [w]? Implementation issues are resolvable, sorting issues are not, and for this reason prevowels are stored following the initial consonant cluster.

The Thai and Lao scripts get around the problem by having strict sort rules that say that a prevowel is reordered to follow the following letter when sorting, regardless of whether that letter is part of a consonant cluster or not.

[8] The relative ordering between U+1A80 (SUBJOINER) U+1AA4 (WA) and U+1ABB (MAI KONG) would be arbitrary in this case. This example establishes a conventional encoding order for this vowel based on storing diacritics from bottom to top.

[9] No separate code is given for mai kak since there is no visual distinction between it an mai sat, and therefore the character is ambiguous within the script. The ambiguity is carried into the encoding since users will not necessarily be able to distinguish between the two letters at the time of data entry.

[10] The precise rendering rules for U+1AC9 LANNA MARK MAI SAM are not clear. For the most part, the mark is rendered above the point where it occurs, unless there is already an upper diacritic in that position, in which case it is moved to be above the next spacing character in the syllable, if there is one.

[11] U+1AB3 LANNA VOWEL MAI KAM has been included since it is the only example where a spacing vowel is followed by a final consonant, but where the final consonant occurs as a diacritic on the base consonant that precedes the spacing vowel. Not including this character would greatly increase the complexity of U+1ACA LANNA MARK MAI KANG.

[12] Where a language name occurs in a letter name, this indicates that the character only occurs in that language.

[13] There is no appropriate general category for U+1ACE (LUE FINAL NA) or U+1ACF (LUE FINAL PA). `Mn` has been used in the database as being the closest to what is needed, but these characters combine with the following character and not the previous character. They therefore are not grapheme extending but cause the following character to be grapheme extending. The behaviour is identical to that of U+1039 (VIRAMA) U+1004 (NGA) in Myanmar, for which no general category exists since it is a sequence.