Tai Tham

a “hybrid” script

that challenges current encoding models

IUC 39
2015 OCT 26-28
Santa Clara, CA, USA
Introductions

Hello!

I’m Patrick Chew, presenting on behalf of two co-authors, Pichai Saengboon and Richard Wordingham, who were unfortunately unable to attend. I take full responsibility for any inaccuracies or mistakes in this presentation.

nb: transcription used in this presentation is based largely on the Library of Congress romanization schema: ư = w/ü ; đ = ə/oe ; ơ = ɔ ; æ = ε/ä, with some other modifications, e.g. ɓ = b, đ = d, etc.
“If you want to know, ask;
If you want to write nicely, practice.”

Lu’ proverb from M(u’)đng Yŏng
Why?

- Pandey and Anderson (2015-Aug-03)
  - “From Local Script to Global Standard: The Lifecycle of a Script in Unicode”
- present basic information about the Tai Tham script
- explore issues with(in) the current model
- contrast a native font/IME effort
- introduce new findings
Basics

What is Tai Tham?
Where is Tai Tham used?
Who uses Tai Tham?
When has Tai Tham been used?
How was Tai Tham used?
How is Tai Tham used?
What is Tai Tham?

Tai Tham is
- a descendant of Brahmi and Pallava
- a “complex script”
- a ‘recently’ Unicode encoded script
  - (Unicode 5.2 = 2009-OCT-01)
Where is Tai Tham used? (1/2)

Tai Tham user communities are locus located in:
- Thailand (Northern (Lanna) and Northeast (Isan))
- Laos
- Myanmar (Keng Tung Township, Shan State)
- PR China (Xīshuāngbǎnnà, Yúnnán Province)
Where is Tai Tham used? (2/2)
Who uses Tai Tham? (1/3)

Tai Tham is used by (a subset of) the communities of the following languages:

- Kammuang/Lanna (“Northern Thai“): 6.0m speakers
- Khun: 106k speakers
- Lu’: 551k speakers
- Lao: 3.3m speakers
- Isan (“Northeastern Thai“): 15m speakers
  \[ n \subset \sim 25m \]
Who uses Tai Tham? (2/3)

Linguistic classification (Wikipedia)

Two distinct branches

- Lao and Isan
- Thai, N.Thai (Lanna/Kammu'ang), Tai Lu', Tai Khun
Who uses Tai Tham? (3/3)

For purposes of examining writing systems ...

- **Consonant development**
  - *Std Thai* + Lao/Isan
  - Lanna + Khun/Lư ~ Shan

- **Vocalic development**
  - *Std Thai* + Lao/Isan + Lanna
  - Lư/Khun ~ Shan
Similarities and differences: Cyrillic

- Одного языка никогда недостаточно
- Однієї мови ніколи не досить
- Един език никога не е достатъчен
- Знати један језик није довољно
- same script
- similar languages
- variation in glyph choices
- script still behaves consistently
## Similarities and Differences: Tham region

<table>
<thead>
<tr>
<th>&quot;New Year&quot;</th>
<th>sankrānti</th>
<th>songkran</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thai</td>
<td>สงกรานต์</td>
<td>songkran</td>
</tr>
<tr>
<td>Lao</td>
<td>สวัสดี</td>
<td>songkān</td>
</tr>
<tr>
<td>Lanna</td>
<td>วันสงกรานต์</td>
<td>sangkhān</td>
</tr>
<tr>
<td>Khun</td>
<td>วันสงกรานต์</td>
<td>sangkhān</td>
</tr>
<tr>
<td>Lu</td>
<td>วันสงกรานต์</td>
<td>sangkhān</td>
</tr>
</tbody>
</table>
When has Tai Tham been used?

c. 1300–present

- 1262-1775 CE Lanna Kingdom
- 1277 written transcription of sutras in Sipsongpanna
- 1376 oldest dated Lanna Tham inscription (Ongsakul 2001)
- 1451 Lanna missionary activity to Keng Tung
- 1465 oldest Tham writing of Lanna Thai and Pali (Hartmann 1986)
- 1471 oldest Lanna palm leaf manuscript (Penth 2000)
- 1548 Lanxang (Laos) and Lanna unite
- 1559 Burmese suzerainty over Kengtung
- 1560 Tham script inscription in Lanxang (Hartmann 1986)
- 1600s use of Tham in Isan
How was Tai Tham used?

Primarily as a liturgical script
- Tham (‘scripture’ < P. dhamma / S. dharma)
- contrasted with the Fak Kham and Nithet scripts
- written on palm leaves (bai-lan) or sa (mulberry) paper

Also secularly used
- historical chronicles,
- legal / astrological /medicinal treatises
- poetry, etc.
Palm leaf (bai-lan) manuscript
[Tham Lao]

Sa (mulberry) paper manuscript
[Tham Lanna]
How is Tai Tham used?

- Remains liturgical in Laos

- Coexists with
  - national Thai as a regional / identity script in N/NE Thailand
  - national Chinese & officially-reformed New Tai Lue script as an identity script
  - national Burmese and regional Shan as a regional / identity script (daily use)

- Online use
  - websites
  - music video subtitling
How is Tai Tham used?
How is Tai Tham used?
Basic features

- Tai Tham as a Brahmic/Indic script
  - abugida
  - independent vowels
  - conjuncts + overt vowel-killer
- Baseline
- Glyphs
Tai Tham as a Brahmic script

**abugida**
- consonants have a default /a/ vowel
- other vowels override default vowel
- vowel positioning all over (not strictly linear like Latn)

**independent vowels**
- conjuncts + overt vowel-killer
Tai Tham as a Brahmic script

abugida
- consonants have a default /a/ vowel
- other vowels override default vowel

| ka + å = kå  | ᴍ  | ʋ  | ʍ |
| ca + l = ci  | すこと  |  |  sổ |
| ta + u = tu  | tolist  |  | сос |
| ṭa + õ = bọ  | ʊ  | ɔ  | ɔ́ |

- vowel positioning all over (not strictly linear like Latn)

independent vowels

conjuncts + overt vowel-killer
Tai Tham as a “complex” Brahmic script

abugida
- consonants have a default /a/ vowel
- other vowels override default vowel
- vowel positioning all over (not strictly linear like Latn)
- unlike most other Indic scripts, Tham can have a complex vowel completely surrounding the consonant (ʔəʔ)
Tai Tham as a “complex” script

- **abugida**
- **independent vowels**

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>ā</th>
<th>i</th>
<th>ī</th>
<th>u</th>
<th>ā</th>
<th>ĕ</th>
<th>ĕ̀</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lanna</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
</tr>
<tr>
<td>trad. Lu</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
</tr>
<tr>
<td>cont. Lu / Khu</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
</tr>
<tr>
<td>Lao / Isan</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
<td>📬</td>
</tr>
</tbody>
</table>

- **conjuncts + overt vowel-killer**
Tai Tham as a “complex” Brahmic script

- abugida
- independent vowels
- conjuncts + overt vowel-killer
Baseline

Tai Tham has a “hanging” baseline
Tai Tham glyph exemplars

Lanna is the most ‘complex’
- open entries, high degree of swash, folded finals

Isan/Lao
- open entries, reduced swash, reduced folding in finals

tr. Lu’
- reduced entries, moderate swash, negligible finals

c. Lu’/Khun
- simplified entries, negligible swash, simplified finals; ‘sans serif’

<table>
<thead>
<tr>
<th>Tham</th>
<th>Lanna</th>
<th>Lao/Isaan</th>
<th>Lu’ trad</th>
<th>Lu’ (cont.)</th>
<th>Khun</th>
</tr>
</thead>
</table>
| Lanna  | ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფიძ ფिञ ფिञ ფिञ ფिञ ფिञ ფიძ ფიძ ფიძ ფიძ ფिञ ფिञ ფिञ ფिञ ფिञ ფิძ ფიძ ფიძ ფიძ ფि
<table>
<thead>
<tr>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myanmar</td>
<td>ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Khun</td>
<td>လူး (ဆိုင်) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Lu</td>
<td>လူး (နိုင်ငံ) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Lanna</td>
<td>လူး (လန်း) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Lao/Isaa</td>
<td>လူး (အိန္ဒိယ) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Lao</td>
<td>လူး (ဗီယလ်) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Thai</td>
<td>လူး (လူး) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
<tr>
<td>Khmer</td>
<td>လူး (ကချင်) ဗိုလ်မှူးများ အတွက် တိုက်ကြည်း သုံးစွဲနေသော ဗိုလ်မှူးများ အပေါ် တိုက်ကြည်း သုံးစွဲနေသော အသုံးအနှုန်းများ ကြည့်ရှုခြင်း</td>
</tr>
</tbody>
</table>
Tai Tham compared to neighbors (2/3)

<table>
<thead>
<tr>
<th></th>
<th>&lt;s&gt;</th>
<th>&lt;b&gt;</th>
<th>&lt;t&gt;</th>
<th>&lt;u&gt;</th>
<th>&lt;w&gt;</th>
<th>&lt;r&gt;</th>
<th>&lt;h&gt;</th>
<th>&lt;n&gt;</th>
<th>&lt;p&gt;</th>
<th>&lt;y&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myanmar</td>
<td>๐</td>
<td>ຘ</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Khun</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Lu (cont.)</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Lu (trad)</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Lanna</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Lao/Isaan</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Lao</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>Thai</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>chrieng</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
<tr>
<td>muol</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
<td>๐</td>
</tr>
</tbody>
</table>
Tai Tham compared to neighbors (3/3)

<table>
<thead>
<tr>
<th>Lek hora</th>
<th>Lek tham</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6 7 8 9</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Myanmar</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Tham (Lur/Khun)</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Lanna</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Lao/Isan</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Lao</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Thai</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
<tr>
<td>Khmer</td>
<td>0 1 2 3 4 5 6 7 8 9</td>
</tr>
</tbody>
</table>
Script relationships suggested by presently-know dated texts

- **Old Khmer**
  - By 1292: Ramkhamhaeng script devised
  - By 1339: 'Classical SK'
  - By 1371: SK in Lamphun
  - By 1411: (Ayuthian Thai) SK > FK in North
  - By 1465: (used to write Tai)
  - By 1503: FK and LN in Nakhon Phanom

- **Old Mon**
  - (Indic prototype scripts)
  - (used to write Pali texts)
Tai Tham problematic points (?)

- ā vs tall-ā
- ṇ ~ ṇ (‘ng’)
- “conjoined” consonants
- medials
  - atomic encoding
- vowel representation
  - consonant? or vowel?
- placement of marks
  - “aesthetics” and regional preferences
default rule:
- Tall-ā follows certain letters to avoid potential ambiguity

<table>
<thead>
<tr>
<th></th>
<th>[g]</th>
<th>[b]</th>
<th>[d]</th>
<th>[dh]</th>
<th>[w]</th>
</tr>
</thead>
<tbody>
<tr>
<td>disambiguating form</td>
<td>౨</td>
<td>౧</td>
<td>౪</td>
<td>౩</td>
<td>౪</td>
</tr>
<tr>
<td>potentially ambiguous form</td>
<td>౨.thumbnail</td>
<td>౧.thumbnail</td>
<td>౪.thumbnail</td>
<td>౩.thumbnail</td>
<td>౪.thumbnail</td>
</tr>
<tr>
<td>look-alike</td>
<td>౨</td>
<td>౧</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conforms to standard Mon/Bamar orthographic rules.
Tai Tham ā and tall-ā (2/2) - “exceptions”

Myanmar encodes ā and tall-ā separately
- Mon / Bamar show complementary distribution
- S’gaw Karen only uses the tall-ā

Tai Tham encodes the two separately
- U+1A63 □ ; U+1A64 □
- despite normative complementary distribution, exceptions are found
- tall-ā follows other letters where potential ambiguity is not expected
- found primarily in Lanna palm-leaf manuscripts

<table>
<thead>
<tr>
<th></th>
<th>၃</th>
<th>သဝ</th>
<th>ဝ</th>
</tr>
</thead>
<tbody>
<tr>
<td>သီ</td>
<td>၃</td>
<td>သဝ</td>
<td>ဝ</td>
</tr>
<tr>
<td>သစ်</td>
<td>သဝ</td>
<td>သဝ</td>
<td>ဝ</td>
</tr>
<tr>
<td>သစ်</td>
<td>သဝ</td>
<td>သဝ</td>
<td>ဝ</td>
</tr>
</tbody>
</table>
Tai Tham -ŋ/ṁ : nga vs mai-kang / mai-kang-lai

- native

- Pali:
  - mai-kang (niggahit)
  - mai-kang-lai ("flowing mai-kang")

- vis-à-vis Myanmar kinzi
Tai Tham -ŋ/m : nga vs mai-kang / mai-kang-lai

native/nativized vocabulary :
○ <jāŋ> » [tɕaːŋ] ‘elephant’
○ U+1A29 U+1A76 U+1A63 U+1A60 U+1A26
○ LOW_CA TONE-2 AA SAKOT NGA
Tai Tham -ŋ/ṁ: nga vs mai-kang / mai-kang-lai

Pali: mai-kang-lai ("flowing mai-kang")
- Pali spelling (word-internal): <saŋʰa> » [saŋkʰa?]
- U+1A48  U+1A58  U+1A25
- SA     MAI KANG LAI  LOW_KHA
Tai Tham -ŋ/ṁ: nga vs mai-kang / mai-kang-lai

Pali spelling (form inflection): `<budḍhaṁ` » `[putṭhaŋ]`
- U+1A3B U+1A69 U+1A34 U+1A60 U+1A35 U+1A74
- LOW_PA U LOW_TA SAKOT LOW_THA MAI KANG
Tai Tham -ŋ/m : nga vs mai-kang / mai-kang-lai

- vis-à-vis Myanmar kinzi
  - Myanmar kinzi specific rendering
  - NGA ASAT VIRAMA
  - U+1004 U+103A U+1039
Tai Tham -ŋ/ṁ : nga vs mai-kang / mai-kang-lai

vis-à-vis Myanmar kinzi
- MA   NGA   ASAT   VIRAMA   GA   LA
- U+1019  U+1004  U+103A  U+1039  U+1002  U+101C
Tai Tham -ŋ/ṁ: nga vs mai-kang / mai-kang-lai

Theoretically feasible to force MAI KANG LAI to behave like a kinzi, where overt vowel killer and virama are used in combination.

BUT - NGA+RA HAAM+SAKOT is a permissible sequence that would not produce the desired affect.
Tai Tham : mai-kang-lai

position: default = over following consonant

Lanna example

Lao example
Tai Tham: mai-kang-lai

position: conditionally shifted

Lanna example
(Chalieo Munchan 1983)
Tai Tham: mai-kang-lai

**position:** between

- modern typeset Lư/Khūn prefers medial placement

Khūn example (Peltier 1996)

Lư example
(西双版纳傣族自治州 2002)
Tai Tham: mai-kang-lai

position:
Luₘ/Khṳn counter-example

older handwritten Luₘ/Khṳn use the default

Khṳn example (1949)

Luₘ example (刀新华 1989)
Tai Tham : mai-kang-lai

position: completely un-shifted

Lanna example
(Chalieo Munchan 1983)
Tai Tham: mai-kang-lai

positioning

- traditional / handwritten varieties default to placement over the following consonant
  - conditionally offset
- modern typeset Lữ/Kh Além prefers medial placing

- Theppitak K. presents Lao examples that do not display offsetting
  - Lao/Isan font spacing is deeper
Final consonants and syllabicity

- Myanmar model always overtly marks
  - asat OR subjoined consonant

- Khmer model rarely marks
  - subjoined consonants in medial position
  - rare to use the tondakiat
  - otherwise unmarked (like Thai/Lao)
Tai Tham innovations, re: syllabicity

status quo stacking

When Tham consonants are written as subscripts

- final consonant of a syllable
- initial consonant cluster medial
- combined with <h>
- consonant symbol represents a vowel
  - y (= ia), w (= ua), ʔ (= ǝ)

\[\text{ exemplos de escrita de subscrita: } \text{ชื่อ} \text{ ไชย} \text{ ชัย} \]

\[\text{consonantes com h: } \text{ชื่อ} \text{ ไชย} \text{ ชัย} \]

\[\text{consonantes com y, w: } \text{ชื่อ} \text{ ไชย} \text{ ชัย} \]
Tai Tham innovations, re: syllabicity + positioning

default positioning is below (subjoined)
Tai Tham innovations, re: syllabicity + positioning

in the case of other characters below (vowels, clusters, medials) positioning below is blocked and spelling requires ‘baseline’ position
Tai Tham innovations, re: syllabicity + positioning

- with <ia>, Lu/Khuⁿ traditionally raise the final primarily -ng, -n (<r>), -w
Tai Tham innovations, re: syllabicity + positioning

- Raised final `<r>` visually looks like TAI THAM SIGN RA HAAM (U+1A7A)
- Traditionally size has differentiated, but graphic variations also have been used to disambiguate

Lu/Khun examples raised `-r ≠ rahaam`

Lanna examples raised `-r ≠ rahaam`
Tai Tham innovations, re: syllabicity + positioning

- raised final `<w>` visually looks like TAI THAM SIGN MAI KANG (U+1A74)
- size usually is the differentiating factor*

Lu/Khun examples
raised `-w ≠ maikang`
Tai Tham innovations, re: syllabicity + positioning

○ Tham Lao/Isan add further confusion to the mix
  ○ depending on the vowel
    ◦ some consonants (-p, -t, -k, -m, -w) behave as expected: subjoined raise to baseline
    ◦ -ng requires either subjoined OR raised
      ◦ raised final NGA and MAI KANG LAI are conflated
    ◦ some consonants (-y) always require subjoined forms
    ◦ some consonants (-n) require subjoined depending on the vowel
Tai Tham innovations, re: syllabic + positioning

Tham Lao/Isan -ng

 металл

 металл
Tai Tham vowels - conditioned allomorphy

Open vs closed syllable:

- khọ ≠ khọng
- hua ≠ huan
Final consonants and syllabicity

- **Myanmar model always overtly marks**
  - asat OR subjoined consonant

- **Khmer model rarely marks**
  - subjoined consonants in medial position
  - rare to use the tondakiat
  - otherwise unmarked

- **Tai Tham model**
  - use of rahaam is rare and sporadic (unmarked)
  - visual cues include
    - positional shifts
    - vowel composition
Orthography clashes

- maximal = 3 (?)
- subjoined stacks ≤ 3
  - <kluay>
Orthography clashes

- Usually stacks are two deep
  - <hluang>
- To mitigate illegibility, lowest stack moves
Orthography clashes

- Tai Tham KH font auto shifts lowermost stack (and changes shape)
  - logical ordering, exceptional placement
- logical ordering, but not logical semantics also occur
Two current Tham approaches

- LN Tilok/Monsaen
  - legacy as-yet-converted to Unicode
  - Lanna-style

- Tai Tham KH
  - Unicode implementation
  - Khưn-style (Lư, by extension)
LN Tilok / Monsaen

- pre-Unicode/legacy efforts ("hack")
  - LN Tilok = Lanna-style
  - LN Monsaen = traditional Lu’-style

- based on Thai (Kedmanee)
  - intent to facilitate ease in transliterated entry using Thai spelling

- combines OTF features + input
  - becomes a basic IME
LN Tilok / Monsaen

**Lan Na Open Type Font (OTF) Structure**

<table>
<thead>
<tr>
<th>Component</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone</td>
<td>วรรณศัพท์</td>
</tr>
<tr>
<td>Above Vowel</td>
<td>ซื่อแซ่</td>
</tr>
<tr>
<td>CONSONANT</td>
<td>ฟรีบินนิช</td>
</tr>
<tr>
<td>Final Consonant</td>
<td>สระสั้น</td>
</tr>
<tr>
<td>Cluster Consonant</td>
<td>สระไม่สลัก</td>
</tr>
<tr>
<td>Consonant as Vowel</td>
<td>ผู้สื่อสารภาษาไทยที่สระ</td>
</tr>
<tr>
<td>Below Vowel</td>
<td>สระต่ำ</td>
</tr>
</tbody>
</table>

**Lan Na Open Type Font (OTF) Process**

- **Tone**
  - eXtra
    - eXtra
      - eXtra
        - eXtra
          - eXtra
            - eXtra
              - eXtra
                - eXtra
                  - eXtra
                    - eXtra
                      - eXtra
                        - eXtra
                          - eXtra
                            - eXtra
                              - eXtra
                                - eXtra
                                 - eXtra
                                  - eXtra
                                   - eXtra
                                    - eXtra
                                     - eXtra
                                      - eXtra
                                       - eXtra
                                        - eXtra
                                         - eXtra
                                          - eXtra
                                           - eXtra
                                            - eXtra
                                             - eXtra
                                              - eXtra
                                               - eXtra
                                                - eXtra
                                                 - eXtra
                                                  - eXtra
                                                   - eXtra
                                                    - eXtra
                                                     - eXtra
                                                      - eXtra
                                                       - eXtra
                                                        - eXtra
                                                         - eXtra
                                                          - eXtra
                                                           - eXtra
                                                            - eXtra
                                                             - eXtra
                                                              - eXtra
                                                               - eXtra
                                                                - eXtra
                                                                 - eXtra
                                                                  - eXtra
                                                                   - eXtra
                                                                    - eXtra
                                                                ��ะ 다음과 ค่าต่อไปนี้

By Mr. Pichai Saneeboon, 22-Oct-2009
Center for the Promotion of Arts and Culture,
Chiang Mai University, THAILAND.
LN Tilok / Monsaen

- based on Thai (Kedmanee)
- allows for ease of typing from Thai into Tham
- codepoints used were the fuller Thai inventory, the font allowed for visually homographic codepoints
LN Tilok (paiyannoï)

- Standard Thai does not stack or ligate
- "variation selector" built in
  - use the lesser-used paiyannoï (ʻ) to choose variants (up to three layers)
    - positional
    - glyph choice(s)
LN Tilok (paiyannoi)

- 1st layer of variation = subjoined
- 2d-3d layer of variation = glyph and positional variants
  - substitutions for spacing
  - regional variations

n.b. <r> doesn’t show the r-medial; it occurs only in certain combinations
### LN Tilok vs current Unicode

<table>
<thead>
<tr>
<th>Tai Tham KH</th>
<th>Copdepoints</th>
<th>LN Tilok</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>render</strong></td>
</tr>
<tr>
<td>NGA</td>
<td>ก</td>
<td>NGA (U+1A26)</td>
</tr>
<tr>
<td></td>
<td>SAKOT (U+1A60) + NGA (U+1A26)</td>
<td>ง</td>
</tr>
<tr>
<td></td>
<td>Final NGA (U+1A59)</td>
<td>งงงง</td>
</tr>
</tbody>
</table>
Atomic encodings vs Variations

<table>
<thead>
<tr>
<th>Name (Codepoint)</th>
<th>Tai Tham KH</th>
<th>LN Tilok</th>
<th>Name (Codepoint)</th>
<th>Tai Tham KH</th>
<th>LN Tilok</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEDIAL RA (U+1A55)</td>
<td>[ ]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA TANG LAI (U+1A57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LAI KANG LAI (U+1A58)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FINAL NGA (U+1A59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOW PA (U+1A5A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIGH Ratha or LOW PA (U+1A5B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MA (U+1A5C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA (U+1A5D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SA (U+1A5E)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIGN OA BELOW (U+1A6C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIGN OY (U+1A6D)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Unlike other Indic scripts (e.g. Malayalam chillaksharam), Tham users mentally associate variants with the original
“Aesthetics” vs ordering

Tham re-orders glyphs “aesthetically” at the writer’s discretion

semantically they are identical

- consonant cluster (HA + SAKOT + MA)
- two-part vowel (MAI KANG + OA BELOW)
- tone (TON-2)

adjusting to visually allow for both results in divergent semantic analysis
“Aesthetics” vs ordering

○ LN Tilok begins visually
  ○ ordering of the upper elements
  ○ HA + MAI KANG + TONE-2

○ last two elements are ordered semantically identical, but ‘variation’ selection adjusts
  ○ MA (subscripted) + A (subscripted + adjusted)
“Aesthetics” vs ordering
LN fonting insights

- current Unicode encoding encodes separate codepoints for positional variants
- LN Tilok glyph choices based on semantic relatedness from most expected to lesser expected within the same font/system
  - Lanna-centric; LN Monsaen somewhat addresses Lự/Khữn preferences, but is not comprehensive
- extremely visually-based
Tai Tham KH

- [ ] [http://www.kengtung.org](http://www.kengtung.org)
- iOS implementation (font + keyboard)
- full (?) repertoire of current Unicode Tai Tham block
- Khṳ̀n-centric (by extension, Lṳ́)
Tai Tham KH

- base layer = most frequent (?)
Tai Tham KH

- shift = less common (?)
Tai Tham KH

3d layer = numbers, symbols, punctuation
Tai Tham KH

handwritten Lư and Khunj simplify the strokes in a TONE-2 (U+1A76) to look like that of a MAI SAT (U+1A62)
- Tai Tham KH keyboard has TONE-2 in the 3d layer and is overlooked in visual representation
- LUE KARAN and raised final RA contrast
  visually homographic; semantically mismatched
Tai Tham KH

U+1A2F U+1A55 U+1A6A U+1A63
◎ logical ordering (d ū r ā) yields incorrect rendering

U+1A2F U+1A6A U+1A55 U+1A63
◎ entry according to visual rendering required to achieve desired rendering
rishida on typography

“An ordinary web author who is trying to add [tatweels] to manually justify the text may not know how to apply these rules.”

Tham users are digitizing according to what is visually familiar, despite mismatches in expectations from encoding.
Current issues

- “abbreviations”
- TAI THAM CONSONANT SIGN LA TANG LAI (U+1A57)
- TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)
- TAI THAM SIGN MAI SAM (U+1A7B)
### Abbreviations

<table>
<thead>
<tr>
<th>భ్రమా</th>
<th>భ్రమి</th>
<th>కు భ్రమి</th>
<th>'also not have'</th>
</tr>
</thead>
<tbody>
<tr>
<td>కుం</td>
<td>కుంతి</td>
<td>కు వాం</td>
<td>'also say(s)'</td>
</tr>
<tr>
<td>భం</td>
<td>భంతి</td>
<td>భ్రమం</td>
<td>'to not come'</td>
</tr>
<tr>
<td>భ్రమి</td>
<td>భ్రమిను</td>
<td>భ్రమిం</td>
<td>ibid.</td>
</tr>
<tr>
<td>భ్రమి</td>
<td>భ్రమిలో</td>
<td>భ్రమిలో</td>
<td>'to not have'</td>
</tr>
<tr>
<td>ప్యా</td>
<td>ప్యాం</td>
<td>ప్యాంం</td>
<td>'comings and goings'</td>
</tr>
</tbody>
</table>
<tanglai> used as a plural marker, usually abbreviated

TAI THAM CONSONANT SIGN LA TANG LAI (U+1A57)

- partial abbreviation atomically encoded
- only/gallery and Khun show this form
Lao example shows abbreviation process

Lanna shows multiple variations
“Abbreviations” (inclusive of TANGLAI)

questions:
◎ are the inventories regional or universal?
◎ visual input expectation vs logical input expectation?
◎ contrast with Latin : & (et) % (per cent) ; ibid. etc.
◎ handled by font? » rendering engine needs?
◎ dictionary? » decomposability?
Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

U+1A5B = subjoined HIGH RATHA (ṭh) | subjoined LOW PA (b)

Lanna and documents and fonts (inclusive of Khùn) indicate non-distinct homography
Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

- Lanna / Khùn : subjoined ṭha = subjoined ba
- Lữ / Lao / Isan : subjoined ṭha ≠ subjoined ba

Lao example of graphic distinction (Luang Mahasena Phui, 1951)
Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

- Lanna / Khūn
  - subjoined ṭha = subjoined ba

- Lư’ / Lao / Isan
  - subjoined ṭha ≠ subjoined ba

Isan example of graphic distinction (Sawing Bunčhôm, 2009)
Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

Lanna / Khün: subjoined ṭhəa = subjoined ba
Lư / Lao / Isan: subjoined ṭhəa ≠ subjoined ba

Lu' example of graphic distinction (刀新华, 1989)

Lu' example of graphic distinction (Khotsimeuang, 2001)
Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

Tham
- Lanna / Khuôn: subjoined ṭha = subjoined ba
- Lư / Lao / Isan: subjoined ṭha ≠ subjoined ba

Myanmar
- Homographic subjoined consonants not documented (yet?)

Khmer
- Subjoined ṭa = subjoined ta
Lek-tho/Mai-sam (U+1A7B) : a scoped relationship

Unicode Standard documented three cases
○ repetition mark for the whole of a word
○ disambiguate function of a subjoined letter
○ “double-acting” consonants

One other also exists
○ indicates a long-a is present
<table>
<thead>
<tr>
<th>Lek-tho/Mai-sam (U+1A7B) : a scoped relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Table" /></td>
</tr>
<tr>
<td><img src="image" alt="Table" /></td>
</tr>
</tbody>
</table>
As yet documented for handling

- use of other numbers
- hāŋg-læm “tail” letters
- other vowel forms
- akkhara piset
More than MAI-SAM/LEK-THO

- Numbers other than “2” (□) are used for abbreviation notation.

- Numbers “3” (□), “4” (□), “5” (□), and “8” (□) are attested.
  - These indicate how many words are contained in a compact ‘abbreviation’.
More than MAI-SAM/LEK-THO

kā kin kluay klāng kṥ = the crow eats bananas in the middle of the grove

- as with the vowel-killing ra-haam, these are optional, c.f. graphic on title page
“long tail” variants

- Myanmar/Khmer models diverge
- Tham incorporates both

<table>
<thead>
<tr>
<th>Padauk</th>
<th>Lucida Sans</th>
<th>Tai Tham KH</th>
<th>LN Tilok</th>
</tr>
</thead>
<tbody>
<tr>
<td>ျ ျ</td>
<td>ပ် ပ်</td>
<td>ဝါ ဝါ</td>
<td>မှ မှ</td>
</tr>
<tr>
<td>ျ ျ</td>
<td>ပ် ပ်</td>
<td>ဝါ ဝါ</td>
<td>မှ မှ</td>
</tr>
<tr>
<td>ျ ျ</td>
<td>ပ် ပ်</td>
<td>ဝါ ဝါ</td>
<td>မှ မှ</td>
</tr>
</tbody>
</table>
“long tail” variants

- “long tail” in higher frequency on particular letters
- some claim only “aesthetics”
- some claim there is semantic difference
  - “long-tail” indicates a syllable initial

<table>
<thead>
<tr>
<th>图片 1</th>
<th>plā</th>
<th>'fish'</th>
</tr>
</thead>
<tbody>
<tr>
<td>图片 2</td>
<td>talāt</td>
<td>'market'</td>
</tr>
</tbody>
</table>
New vowel forms (?)

◎ a full-sized ‘2’ represents the long-æ vowel

◎ new representations for /au/
  ○ /i+/u/
  ○ TAI THAM VOWEL SIGN OA ABOVE (U+1A73)
  ○ ‘subscript-qa’ above for /au/
“Akkhara piset” (special characters)

- some are visually derivable; some are not
- cryptogrammic equivalents for default glyphs
  - “U+1A7F tai tham 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.”
  - variation in directionality in layout
Tai Tham
- exhibits similarities with neighboring writing systems, as expected of related scripts, but shows “hybrid” nature exhibiting features found in one or another or incorporating multiple
- shows extreme variance that confounds contemporary programming efforts
- despite relatively uniform behavior, regional differences manifest in a number of ways
- more content and data indicate that there’s much more to investigate
Summary 2/2

high-level topics

- user experience; actual use; flow
- interoperability
- storage, parsing, normalization
- orthography; lexicography; fonting
- standardization
Resources

Digitized manuscripts
- Chiang Mai University DBMF Project (http://202.28.27.140/)
- École Française d’Extrême Orient (http://www.efeo.fr/lanna_manuscripts/)
- Digital Library of Lao Manuscripts (http://www.laomanuscripts.net/en/index)
- NIU SEASite (http://www.seasite.niu.edu/tai/TaiLue/manuscript/group1/manuscript1.htm)

Websites
- 西双版纳傣文网／新闻网 (http://www.dw12.com/)
- 莽 Triangle (http://www.kengtung.org/)
Attributions

Fonts used in this presentation include:

- LN Tilok / LN Monsaen
  - https://drive.google.com/folderview?id=0B3hdjepXeOuaRGVKVVIETXVTcGM&usp=sharing
- Tai Tham KH
  - http://www.kengtung.org/download-font/
- UBWManut
  - http://www.isangate.com/word/fontthum.rar
- Da Lekh
  - http://homepage.ntlworld.com/richard.wordingham/lanna/dalekh.ttf
- Mulawarman Pallava by Aditya Bayu Perdana
  - https://www.behance.net/gallery/17197405/Pallava-font-Mulawarman
- NotoSans (Aramaic, Brahmi, Phoenician, Tai Tham)
  - https://www.google.com/get/noto/

Photo image attributions:

- Ashley Veasly
  - https://www.flickr.com/photos/ducatiash/
- NIU SEAsite, Laos
  - http://www.seasite.niu.edu/lao/Lao_galleries/LaoManuScriptsProgramme/LaoManuscript/manuscript2.jpg
References

Unicode Standard:
- Ch. 16.7
- code charts

Bibliography:
- http://bit.ly/1O0cN0B
Thank you very much!

Patrick Chew (patrick.chew@gmail.com)
Pichai Saengboon (porlor@gmail.com)
Richard Wordingham (richard.wordingham@ntlworld.com)