Tai Tham

a "hybrid" script

that challenges current encoding models

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 = ɯ/ü ; ơ = ə/oe ; ọ = ɔ ; æ = ɛ/ä, with some other modifications, e.g. ɓ = ɓ, đ = ɗ, etc.

ꨅꨯꨟꨴꨕ ꨤꨳ ꨟ ꨤꨳ ꨟ ꨅ ꨟ

ꨅꨯꨟꨴ ꨟ ꨅꨯ ꨵ ꨤꨳ ꨟ

"If you want to know, ask;
If you want to write nicely, practice."

Lư proverb from M(ư)ơng Yọng

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 descendent of Brahmi and Pallava

◎ a "complex script"

◎ a 'recently' Unicode encoded script

  ○ (Unicode 5.2 = 2009-OCT-01)

	\<kʰ\>	\<g\>	\<gʰ\>	\<t\>
Proto-Sinaitic				
Phoenician	Φ / Ϡ	٦		✝
Aramaic	Ρ / Ϡ	⅄		η
Brāhmī	٦	Λ	Ⴑ	⅄
Pallava	౭	ర్	౮	ఞ
Tai Tham	౨	౧	౫	౬

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)
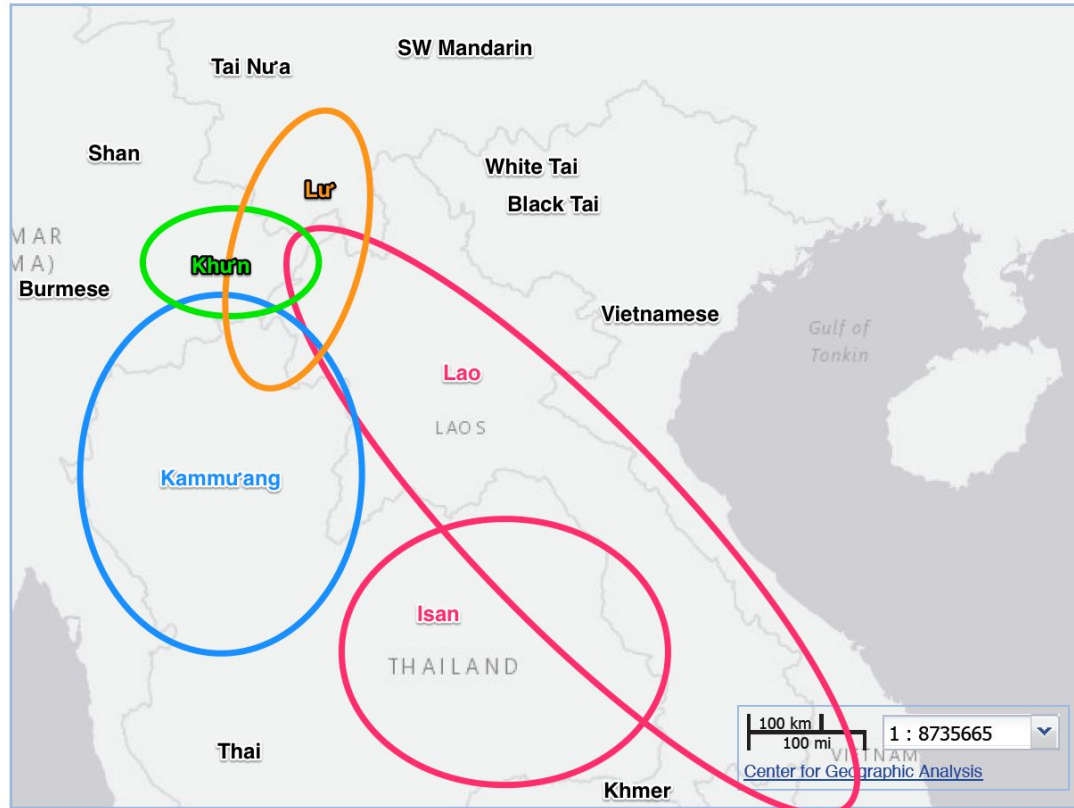
Who uses Tai Tham? (1/3)

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

◎ Kammɯang/Lanna ("Northern Thai") :   6.0m speakers
◎ Khɯn :   106k speakers
◎ Lɯ :   551k speakers
◎ Lao :   3.3m speakers
◎ Isan ("Northeastern Thai") :   15m speakers
   ○ $n \subset$ **~25m**

Linguistic classification (Wikipedia)

◎ Two distinct branches

    ○ Lao and Isan

    ○ Thai, N.Thai (Lanna/Kammưang), Tai Lư, Tai Khưn

Tai (Zhuang)	(Proto-Tai)		
	Northern	Standard Zhuang · Bouyei · Yei Zhuang · E · Saek · Tai Yo · Yoy	
	Central	Ningming · Nong Zhuang · Dai Zhuang · Min Zhuang · Yang Zhuang · Nung · Tày · Ts'ün-Lao	
	Southwestern (Thai)	Northwestern	Shan · Tai Ya · Tai Nüa · Tai Hongjin · Khamti · Tai Laing · Tai Phake · Tai Aiton · Khamyang · Ahom · Turung · Sapa
		Lao–Phutai	Lao · Phu Thai · Isan · Nyaw
		Chiang Saen	Thai (Siamese) · Northern Thai · Tai Daeng · Tai Dón · Tai Hang Tong · Tai Lü · Tai Dam · Khun · Phuan · Thai Song · Tày Tac
		Southern	Southern Thai
	(other)	Caolan	

For purposes of examining writing systems …

◎ Consonant development
  ○ *Std Thai* + Lao/Isan
  ○ Lanna + Khưn/Lư ~ *Shan*

◎ Vocalic development
  ○ *Std Thai* + Lao/Isan + Lanna
  ○ Lư/Khưn ~ *Shan*

PSW Tai	Std Thai	Lao / Isan	Lanna	Khưn / Lư	Shan
*k	k	k	k	k	k
*kʰ	kʰ	kʰ	kʰ ~ x	kʰ ~ x	x ~ kʰ
*x	kʰ	kʰ	kʰ ~ x	kʰ ~ x	x ~ kʰ
*g	kʰ	kʰ	k	k	k
*gʰ	kʰ	kʰ	kʰ ~ x	kʰ ~ x	x ~ kʰ
*ğ	kʰ	kʰ	kʰ ~ x	kʰ ~ x	x ~ kʰ
*e	e	e	e	e ~ i	e ~ i
*ia	ia	ia	ia	e	e
*o	o	o	o	o ~ u	o ~ u
*ua	ua	ua	ua	o	o
*ơ	ơ	ơ	ơ	ơ ~ ư	ơ ~ ư
*ưa	ưa	ưa	ưa	ơ	ơ
*ai	ai	ai	ai	ai	ai
*aư	ai	ai	ai	ai	aư

Similarities and differences : Cyrillic

◎ Одного языка никогда недостаточно

◎ Одн**ієї** мови н**і**коли не досить

◎ Един език никога не е достатъчен

◎ Знати **је**дан **је**зик ни**је** дово**љ**но

   ◎ same script

   ◎ similar languages

   ◎ variation in glyph choices

   ◎ script still behaves consistently

Similarities and Differences : Tham region

"New Year"	saṅkrānti		
Thai	สงกรานต์		songkran
Lao	ມກຼາ໊ຕຖ	ສິງ� / ຫຼືcasu	songkān
Lanna	ᩈᨠᩕᩤᨶ᩠ᨲ᩼	ᩈᩴ᩵ᨦ	sangkhān
Khʉn		ᩈᩴ᩵ᨶ᩠ᨦᩩ	sangkhān
Lʉ		ᩈᩴ᩵ᨶ᩠ᨦᩩ	sangkhān

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]
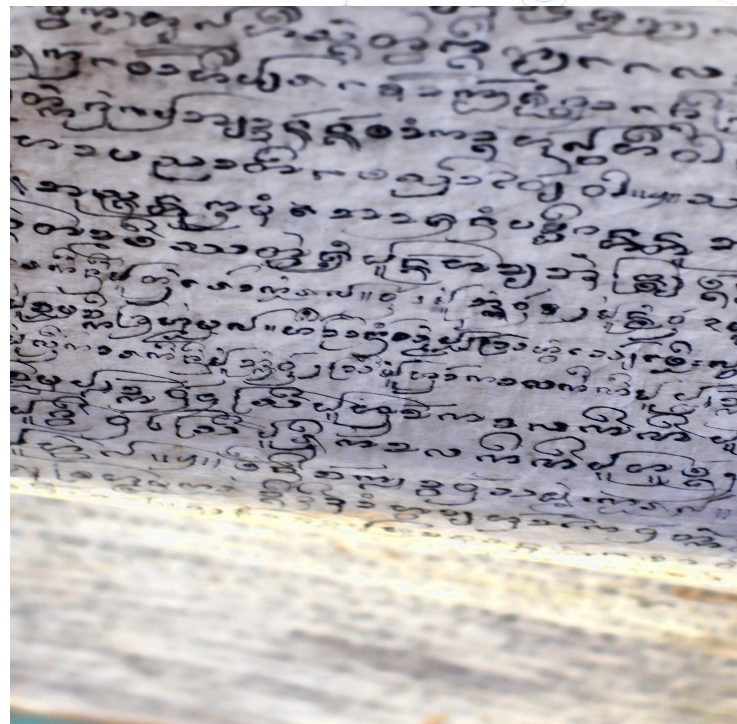
Sa (mulberry) paper manuscript
[Tham Lanna]

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

ေ	ဦ	ဥ	ဂ	ငိ	ဃ	ေ
<ka>	<kha>	<xa>	<ga>	<ğa>	<gha>	<ŋa>
ဧ	ဆ		ဇ	ဈ	ဈ	ဉ
<ca>	<cha>		<ja>	<za>	<jha>	<ńa>

　　○ 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 + ā = kā	ဢ	၂	က၁
ca + l = ci	၁	၆	၆
ta + u = tu	ဢ	‿	ဢ
ḅa + ǭ = bǭ	ပ	ံ	၀

  ○ 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

	a	ā	i	ī	u	ū	ē	ō
Lanna	ဧဒ	ဧဒာ	ၯၟ	ၯၟ	၉	၆	၉	ၜၶ
trad. Luʼ	ၮၟ	ၮၟာ	ၯၟ	ၯၟ	၉	၆	၉	ၜၮ
cont. Luʼ / Khu	ၮၟ	ၮၟာ	ၯၟ	ၯၟ	၉	၆	၉	ၜၮ
Lao/ Isan	ၮ	ၮာ	ၮ	ၮ	၉	၉	၉	ၮ

◎ conjuncts + overt vowel-killer

Tai Tham as a "complex" Brahmic script

◎ abugida
◎ independent vowels
◎ **conjuncts + overt vowel-killer**

	'elephant'	'fragrant'	teachings; dharma	'dharma'
	cāŋ	hǭm	tʰamma	tʰam
Lanna				
trad. Lɯ				
cont. Lɯ / Khɯn				
Lao/ Isan				
	T2 jā ŋ	h m ǫ	a dʰ m ma	a x dʰ m m(a)

Baseline

◎ Tai Tham has a "hanging" baseline
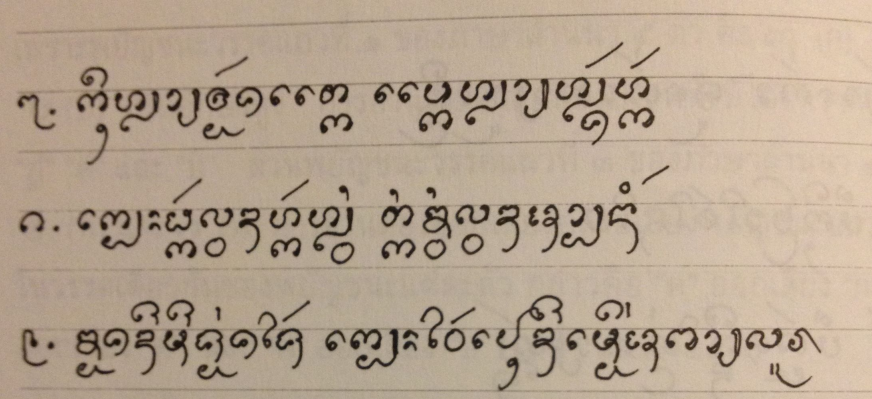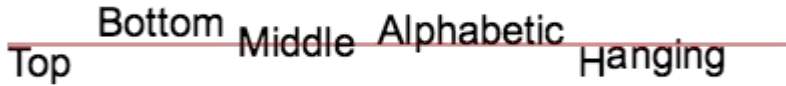
Tai Tham glyph exemplars

◎ Lanna is the most 'complex'
  ○ open entries, high degree of swash, folded finals
◎ Isan/Lao
  ○ open entries, reduced swash, reduced folding in finals
◎ tr. Lư
  ○ reduced entries, moderate swash, negligible finals
◎ c. Lư/Khưn
  ○ simplified entries, negligible swash, simplified finals ; 'sans serif'

Tham	Lanna	ᩈᩢᨣᩮᩖᩤ ᩈᩢᨲᩖᩤᩣ ᨁᩩᩅᨲᩣᩣ ᩉᩣᩣᨿᩩ ᨁᩢᨩᩭᩣᩅᨸᩖᩣᩣ ᩉᩣᩣᨿᩩ ᨁᨯᩢᨬᩣᩣ ᩉᩣᩣᨿᩩ ᩈᨬᩢᩥ ᨁᨲᩢᩢᨯ� ᨸᨶᩦᩉᩣᨬᩩ
	Lao/Isaan	ᨠᩬᩢᨣᩮᩖᩁ ᨠᩢᨲᩖᩣᩣ ᨣᩢᩅᨲᨬᩣ ᩉᩬᩣᩣᨬᩩ ᨣᩢᨩᩭᩣᨣᨯᩮᨲᩣ ᩉᩬᩣᩣᨬᩩ ᨣᨯᩥᨬᩣᩣ ᩉᩬᩣᩣᨬᩩ ᨠᨯᩥ ᨣᨲᩢᩥ ᨯᩢᨸᨶᩦᩉᩣᨬᩩ
	Lư (trad)	ᩈᩢᨣᩮᩖᩤ ᩈᩢᨲᩖᩤᩣ ᨷᩩᩅᨠᩮᩣ ᩉᩣᩣᨬᩩ ᨷᩩᨩᩭᩣᩅᨸᨲᩤᩣ ᩉᩣᩣᨬᩩ ᨷᩩᨣᨬᩣᩣ ᩉᩣᩣᨬᩩ ᩈᨬᩥᩢ ᨷᩩᨲᩢᩢᨯᩢ ᨸᨶᩦᩉᩣᨬᩩ
	Khưn Lư (cont.)	ᩈᩮᩣᨣᩮᩖᩤ ᩈᩢᨲᩖᩤᩣ ᨷᩮᩣᩅᨠᩣ ᨣᩮᩣᩣᨬᩩ ᨷᩩᨩᩭᩣᩅᨸᩖᨲᩤ ᨣᩮᩣᩣᨬᩩ ᨷᩩᨣᩥᨬᩣᩣ ᨣᩮᩣᩣᨬᩩ ᩈᨬᩥᨶᩦ ᨷᩩᨲᩢᩢᨯᩢ ᨸᨶᩦᩉᩣᨬᩩ

Tai Tham compared to neighbors (1/3)

Myanmar		သဗ္ဗေ သတ္တာ အဝေရာ ဟောန္တု၊ အဗျာပဇ္ဇာ အနီယာ သုခီ အတ္တာနံ ပရိဟရန္တု॥
Tham	**Khʉn Lʉ (cont.)**	သေဗ္ဗသတ္တာ ဘေဝရာ ေဟာဝ္တု ဘျာဝပျာ ေဟာဝ္တု ဘဝီဃာ ေဟာဝ္တု သုဃီ ဘျတ္တဝ့ ပဝီဟရဝ္တု
	Lʉ (trad)	သဗ္ဗ သတ္တာ ဘဝရာ ေဟာဝ္တု ဘျာဝပဇ္ဇာ ေဟာဝ္တု ဘဝီဃာ ေဟာဝ္တု သုဃီ ဘျတ္တဝ့ ပဝီဟရဝ္တု
	Lanna	သဗ္ဗ သတ္တာ ဃဝရာ ေဟာဝ္တု ဃျာဝပဇ္ဇာ ေဟာဝ္တု ဃဝီဃာ ေဟာဝ္တု သုဃီ ဃတ္တဝ့ ပဝီဟရဝ္တု
	Lao/Isaan	ສຯຕ ລຕຕຯ ຫຍຍຯ ຫຍຍຫຯ ຫຍຍຍຯ ຫຍຍຫຯ ຍຍຫຯ ຫຍຍຫຯ ຍຍ ຫຍຍ ຍຫຍຍຍຫຍຍຍ
Lao		ສັບເພ ສັດຕາ ຂະເວຣາ ໂຫນຕຸ ຂະພະຍາປັດຊາ ໂຫນຕຸ ຂະນີຕາ ໂຫນຕຸ ສຸຂີ ອັດຕານັ້ງ ປະຕິຮາຣັນຕຸ
Thai		สพฺเพ สตฺตา อเวรา โหนฺตุ อพฺยาปชฺฌา โหนฺตุ อนีฆา โหนฺตุ สุขี อตฺตานํ ปริหรนฺตุ
Khmer	**chrieng**	សព្វេ សត្តា អវេរា អព្យាបជ្ឈា អនីឃា សុខី អត្តានំ បរិហរន្តុ។

Tai Tham compared to neighbors (2/3)

		<s>	<b>	<t>	<?>	<w>	<r>	<h>	<n>	<j>	<y>	<kʰ>
Myanmar		သ	ဗ	တ	အ	ဝ	ရ	ဟ	န	ၕ	ယ	ခ
Tham	Khưn Lư (cont.)	သ	ၜ	တ	ၒ	ဝ	ၖ	ဟ	ၕ	ၕ	ယ	ၑ
	Lư (trad)	သ	ၜ	တ	ၒ	ဝ	ၖ	ဟ	ၕ	ၕ	ယ	ၑ
	Lanna	သ	ၜ	တ	ၒ	ဝ	ၖ	ဟ	ၕ	ၕ	ယ	ၑ
	Lao/Isaan	ບ	ၜ	တ	ၒ	ဝ	ຣ	ຫ	ၕ	ຍ	ယ	ຂ
Lao		ສ	ພ	ຕ	ອ	ວ	ຣ	ຫ	ນ	ຢ	ຍ	ຂ
Thai		ส	พ	ต	อ	ว	ร	ห	น	ช	ย	ข
Khmer	chrieng	ស	ព	ត	អ	វ	រ	ហ	ន	ជ	យ	ខ
	muol	ស	ព	ត	អ	វ	រ	ហ	ន	ជ	យ	ខ

Tai Tham compared to neighbors (3/3)

		Lek hora										Lek tham									
		0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
Myanmar		၀	၁	၂	၃	၄	၅	၆	၇	၈	၉										
Tham	Lur/Khurn	ᦐ	ᦑ	ᦒ	ᦓ	ᦔ	ᦕ	ᦖ	ᦗ	ᦘ	ᦙ	ᦐ	ᦑ	ᦒ	ᦓ	ᦔ	ᦕ	ᦖ	ᦗ	ᦘ	ᦙ
	Lanna	ᰀ	ᰁ	ᰂ	ᰃ	ᰄ	ᰅ	ᰆ	ᰇ	ᰈ	ᰉ	ᰀ	ᰁ	ᰂ	ᰃ	ᰄ	ᰅ	ᰆ	ᰇ	ᰈ	ᰉ
	Lao/Isan	ᩐ	ᩑ	ᩒ	ᩓ	ᩔ	ᩕ	ᩖ	ᩗ	ᩘ	ᩙ	᩠	ᩡ	ᩢ	ᩣ	ᩤ	ᩥ	ᩦ	ᩧ	ᩨ	ᩩ
Lao												໐	໑	໒	໓	໔	໕	໖	໗	໘	໙
Thai												๐	๑	๒	๓	๔	๕	๖	๗	๘	๙
Khmer												០	១	២	៣	៤	៥	៦	៧	៨	៩

Diller 1988

[3] Script relationships suggested by presently-know dated texts

(Indic prototype scripts)

Old Khmer   Old Mon

By 1292   Ramkhamhaeng script   (used to write
               devised   Pali texts)

By 1339   ' Classical SK '

By 1371   SK in Lamphun

By 1411   (Ayuthian Thai)   SK > FK in North

                                                (used to write Tai)
By 1465

                                      LN in North

By 1503   FK and LN in Nakhon Phanom
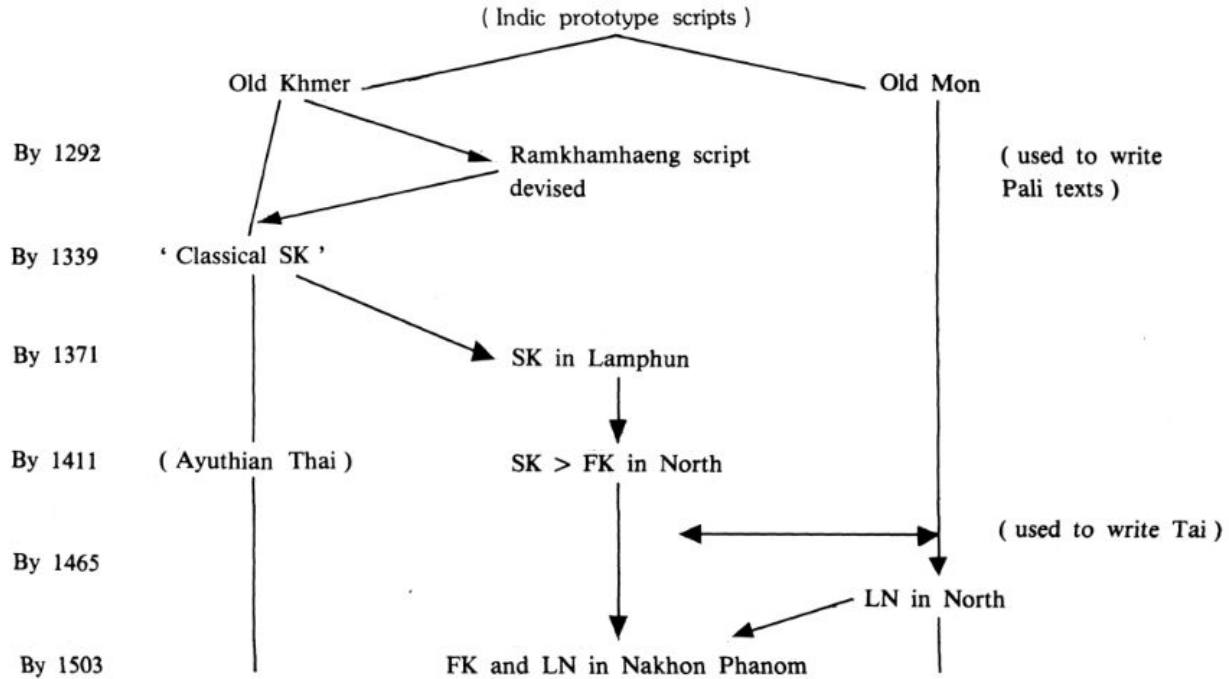
Tai Tham problematic points (?)

- ā vs tall-ā
- ṅ ~ ŋ ('ng')
- "conjoined" consonants
- *medials*
  - *atomic encoding*
- *vowel representation*
  - *consonant? or vowel?*
- *placement of marks*
  - *"aesthetics" and regional preferences*

Tai Tham ā and tall-ā (1/2) - "default"

default rule:

◎ tall-ā follows certain letters to avoid potential ambiguity

	[g]	[b]	[d]	[dh]	[w]
disambiguating form	�	�	ဝwon	ဝwon	ဝwon
potentially ambiguous form	ဝၵ	ၯ		ဝၵ	
look-alike	ၵ	ၯ		ၵ	
	[k]	[h]		[t]	

◎ 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

<ṇ>	ᩉ	ᩉᩣ	ᩉᩤ
<s>	ᩈ	ᩈᩣ	ᩈᩤ
<ss>	ᩔ	ᩔᩣ	ᩔᩤ

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/ṁ : 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) : <saṅgʰa> » [saŋkʰaʔ]
   ○ 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) : <buddᶠaṁ> » [puttʰaŋ]

- 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/ṁ : 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

မင်ႅလ ᥜ်ႅဂလ

ᥜ်ꧥဂလ

ᥜ်ꧥဂလ

◎ 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

ᥜꧡ်လ

Tai Tham : mai-kang-lai
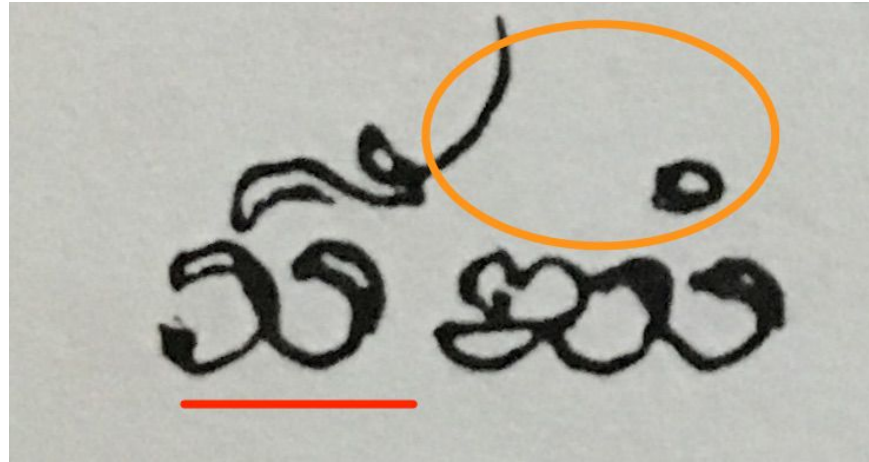
position: default = over following consonant
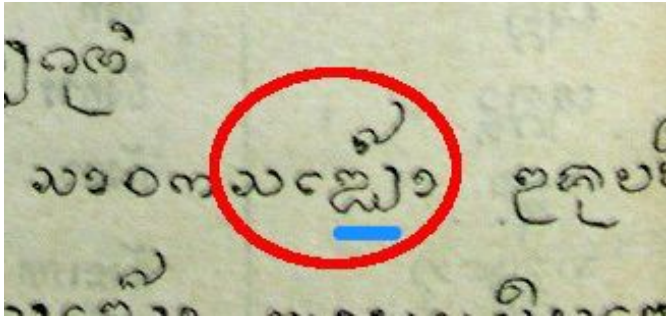
Lanna example
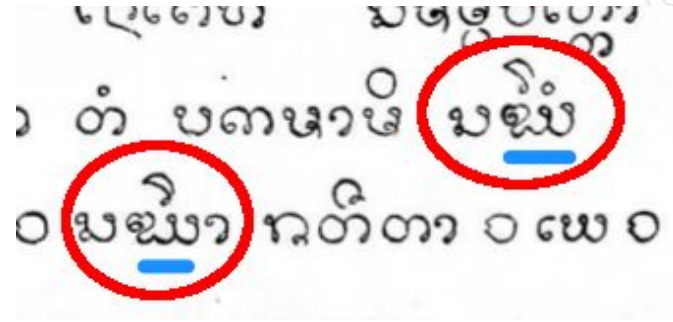
Lao example

Tai Tham : mai-kang-lai
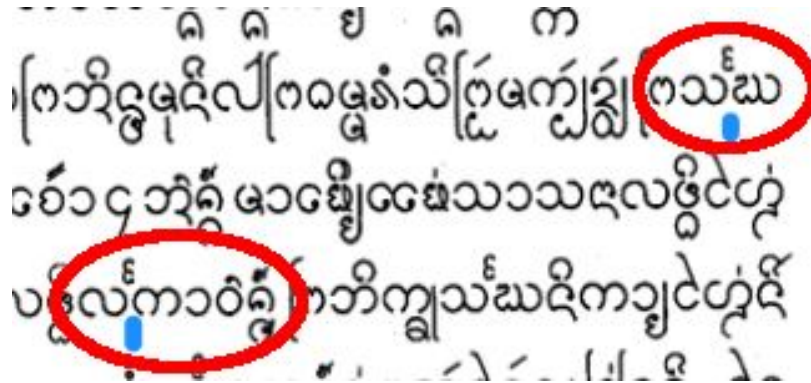
position: conditionally shifted
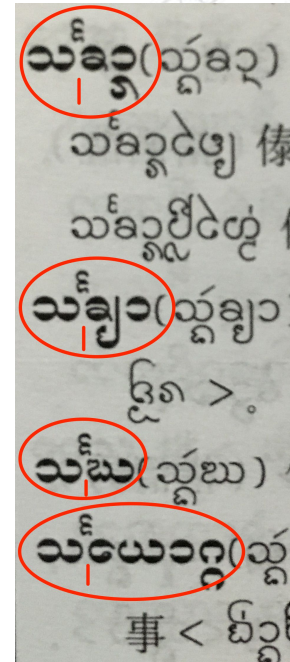

Lanna example
(Chalieo Munchan 1983)

Tai Tham : mai-kang-lai

position : between

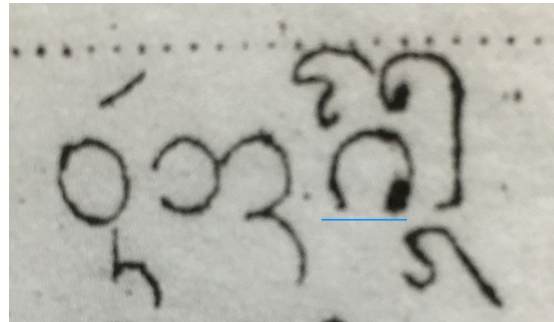◎ modern typeset Lư/Khưn prefers medial placement

Khưn example (Peltier 1996)

Lư example
(西双版纳傣族自治州 2002)

position:
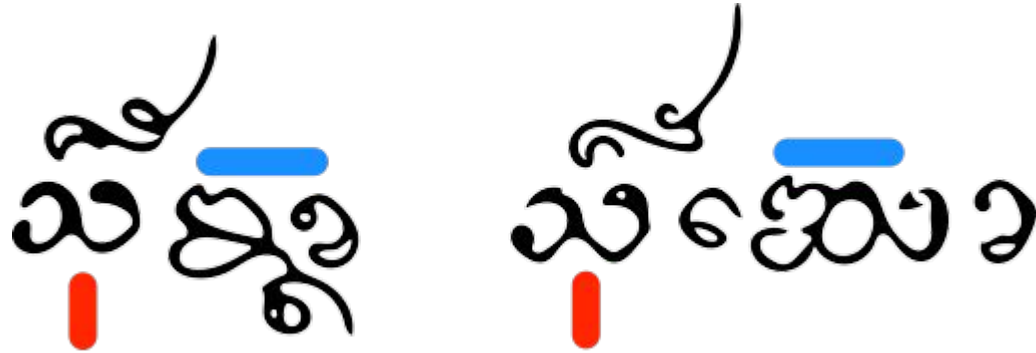# Lư/Khưn counter-example

◎ older handwritten Lư/Khưn use the default

Khưn example (1949)

Lư example
(刀新华 1989)

position: completely un-shifted

Lanna example
(Chalieo Munchan 1983)

positioning

◎ traditional / handwritten varieties default to placement over the following consonant
  ○ conditionally offset
◎ modern typeset Lư/Khưn 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>
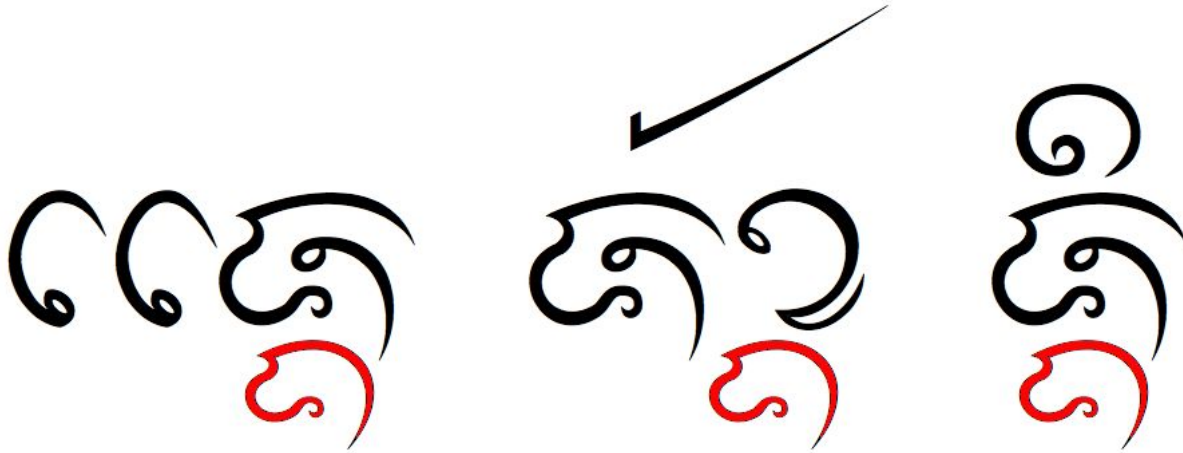- consonant symbol represents a vowel
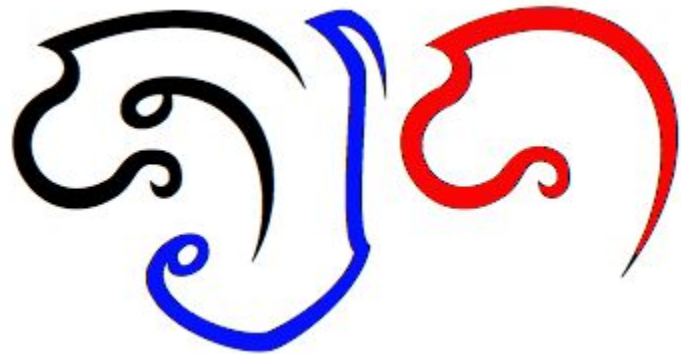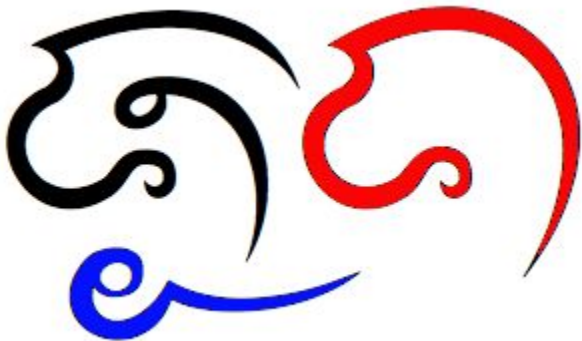 - y (= ia), w (= ua), ʔ (= ǫ)

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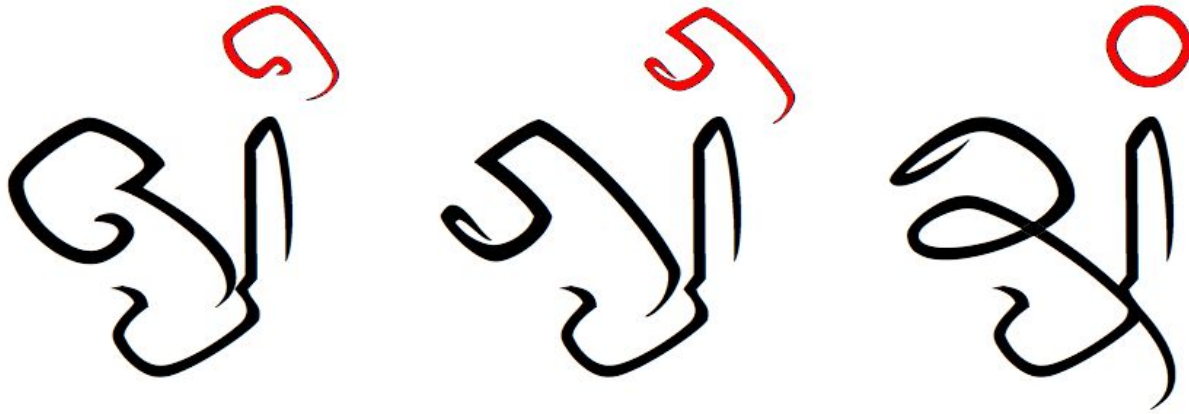
◎ default positioning is below (subjoined)

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>, Lư/Khưn 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
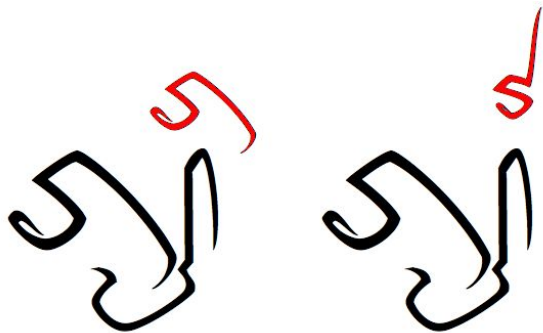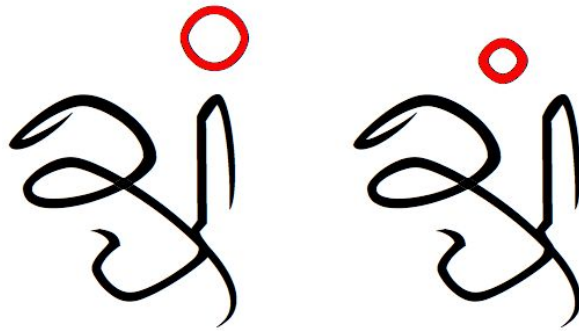
Lư/Khư̈n examples
raised -r ≠ rahaam

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*

Lư/Khưn 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: syllabicity + positioning

◎ Tham Lao/Isan -ng

open vs closed syllabl

◎ khǭ ≠ khǭng

◎ hua ≠ huan
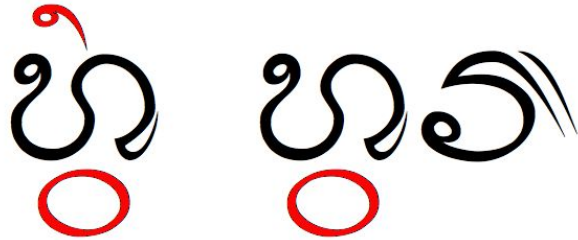
Final consonants and syllabicity

◎ 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
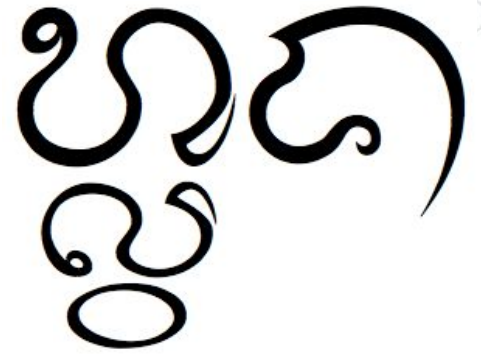
Orthography clashes

◎ Tai Tham KH font auto shifts lowermost stack (and changes shape)
  ○ logical ordering, exceptional placement
◎ logical ordering, but not logical semantics also occur

ᩈᩣᨠᩖ᩠ᨯ = ᩈᩣᨠᩖ᩠ᨯ

Two current Tham approaches

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

◎ Tai Tham KH
  ○ Unicode implementation
  ○ Khửn-style (Lử, by extension)

LN Tilok / Monsaen

◎ pre-Unicode/legacy efforts ("hack")
  ○ LN Tilok   = Lanna-style
  ○ LN Monsaen   = traditional Lɯ-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

กำหนดให้

Tone	=	วรรณยุกต์
Above Vowel	=	สระบน
CONSONANT	=	พยัญชนะ
Final Consonant	=	ตัวสะกด
Cluster Consonant	=	อักษรควบ
Consonant as Vowel	=	พยัญชนะทำหน้าที่สระ
Below Vowel	=	สระล่าง

x = ปุ่มพิเศษ (eXtra)
ทำหน้าที่ขยาย หรือดึงความสามารถ
ของอักษรใดๆ ที่ x ได้กำกับเอาไว้
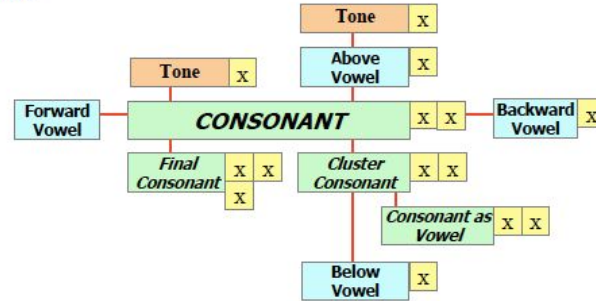
Lan Na Open Type Font (OTF) Process
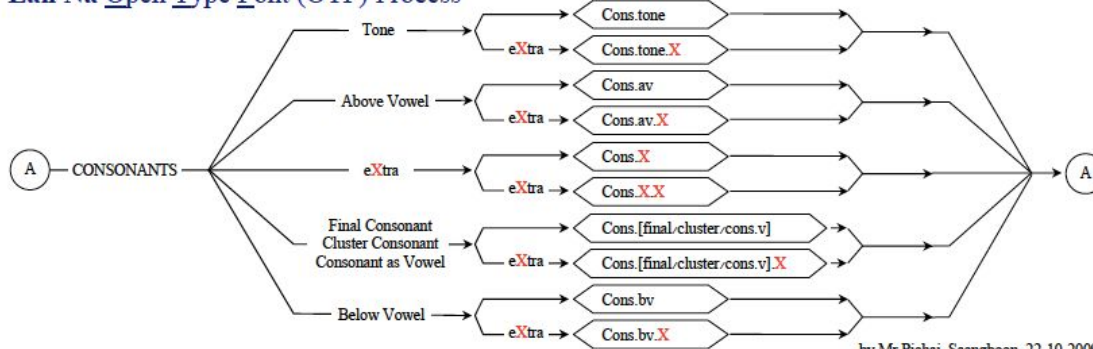
by Mr.Pichai Saengboon, 22-10-2009
Center for the Promotion of Arts and Culture,
Chiang Mai University, THAILAND.

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

k	ก	ᩀ		ḍ̣	ᨮ	ᩈ		đ	ด	ᩈ
kh	ข	ᩈ		ṭ	ᨯ	ᩅ		t	ต	ᩈ
x	ฃ	ᩈ		ṭh	ᨒ	ᩈ		th	ถ	ᩈ
g	ค	ᩏ		ḍ	ᨮ	ᩈ		d	ท	ᩈ
ğ	ฅ	ᩆ		dh	ᨒ	ᩈ		dh	ธ	ᩈ
gh	ฆ	ᩈ		ṇ	ᨱ	ᩈ		n	น	ᩈ
ng	ง	ᩈ								

LN Tilok (paiyannoi)

◎ Standard Thai does not stack or ligate
◎ "variation selector" built in
  ○ use the lesser-used paiyannoi (ๆ) 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

	Tai Tham KH	Copdepoints	LN Tilok	
			render	entry
NGA	ᨦ	NGA (U+1A26)	ᨦ	ꩦ
	ᨦ	SAKOT (U+1A60) + NGA (U+1A26)	ᨦ	ꩦ�
	ᨦ	FINAL NGA (U+1A59)	ᨦ	ꩦ��

Atomic encodings vs Variations

Name (Codepoint)	Tai Tham KH	LN Tilok		Name (Codepoint)	Tai Tham KH	LN Tilok	
MEDIAL RA (U+1A55)	ᩕ	ᨠᩕ	กรๆ	HIGH RATHA OR LOW PA (U+1A5B)	ᩛ	ᩛ /ᩛ	ฐๆๆ / พๆๆ
MEDIAL LA (U+1A56)	ᩖ	ᨠᩖ / ᨠᩖ	กลๆ / กลๆๆ	MA (U+1A5C)	ᩜ		
LA TANG LAI (U+1A57)	ᩗ			BA (U+1A5D)	ᩝ	ᩝ	บๆ
MAI KANG LAI (U+1A58)			◌ๆ	SA (U+1A5E)	ᩞ	ᩞ	สๆ
FINAL NGA (U+1A59)	ᩙ	ᩙ	งๆๆๆ	SIGN OA BELOW (U+1A6C)	ᩬ	ᩬ	อๆ
LOW PA (U+1A5A)	ᩚ	ᩚ	พๆๆๆ	SIGN OY (U+1A6D)	ᩭ		

◎ unlike other Indic scripts (e.g. Malayalam chillaksharam), Tham users mentally associate variants with the original

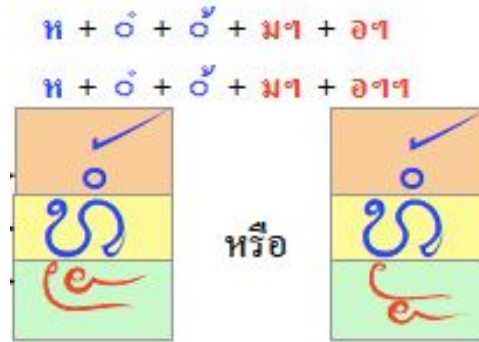"Aesthetics" vs ordering
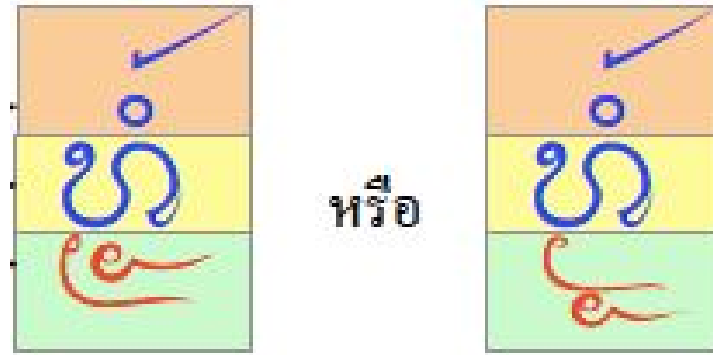
◎ Tham re-orders glyphs "aethetically" at the writer's discretion

◎ semantically they are identical
  ○ consonant cluster (HA + SAKOT + MA)
  ○ two-part vowel (MAI KANG + OA BELOW)
  ○ tone (TONE-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 <u>within the same font/system</u>

　　○ Lanna-centric ; LN Monsaen somewhat addresses Lư/Khưn preferences, but is not comprehensive extremely visually-based

Tai Tham KH

◎ http://www.kengtung.org

◎ iOS implementation (font + keyboard)

◎ full (?) repertoire of current Unicode Tai Tham block

◎ Khʉn-centric (by extension, Lʉ)

᩠ᩈᨣᩮᩬᩮ ᩈᨲ᩠ᨲᩤᩚ ᨷᩩᨾᩮᩤᩚ ᩮᨲᩣᩢᩥᩊ ᨷᩩᩉᨿᩚᨦᩊ᩠ᨿᩣᩚ
ᩮᨲᩣᩢᩥᩊ ᨷᩩᨫᩥᩬᨰᩣᩚ ᩮᨲᩣᩢᩥᩊ ᩈᩩᨲᩬᩥ ᨷᩩᨲ᩠ᨲᨦᩴ
ᨷᩢᩥᨲᩉᩬᩮᩊ

Tai Tham KH

◎ base layer = most frequent (?)

Tai Tham KH

◎ shift = less common (?)

Tai Tham KH

◎ 3d layer = numbers, symbols, punctuation

Tai Tham KH

U+1A4B U+1A20 U+1A60 U+1A21 U+1A41 || U+1A34 U+1A71 || U+1A35
U+1A62 U+1A3E U+1A60 U+1A3E U+1A7C || U+1A21 U+1A60 U+1A3F U+1A7A
|| U+1A2F U+1A71 U+1A62 || U+1A37 U+1A73 U+1A75

◎ handwritten Lư and Khừn 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
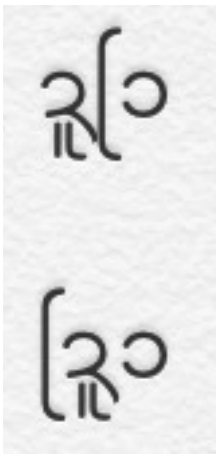
U+1A2F U+1A6A U+1A55 U+1A63

◎ logical ordering (d ū r ā) yields incorrect rendering

U+1A2F U+1A55 U+1A6A U+1A63

◎ entry according to visual rendering required to achieve desired rendering

rishida on tyopgraphy

◎ "<u>An ordinary web author</u> who is trying to add [tatweels] to <u>manually justify the text may not know how to apply these rules</u>."

◎ 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"

ဂ့ံ	ဂ်ဂ့ံ	kǫ ɓǫ	'also not have'
ဂ္ဂာ	ဂ်ဒ္ဓ	kǫ wā	'also say(s)'
ဗ့ံာ	ဗ့ံဧာ	ɓǫ mā	'to not come'
ဗ္ဗ	ဗ့ံဌ	ɓǫ mī	ibid.
ဗ္ဂ	ဗ့ံဌ	ɓǫ mī	'to not have'
ပ္ပာ	ဗ္ဗဧာ	pai mā	'comings and goings'

TANGLAI

◎ <tanglai> used as a plural marker, usually abbreviated
◎ TAI THAM CONSONANT SIGN LA TANG LAI (U+1A57)
  ○ partial abbreviation atomically encoded
  ○ only Lư and Khưn show this form
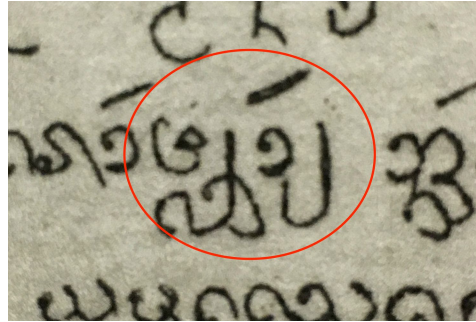
TANGLAI

◎ 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.

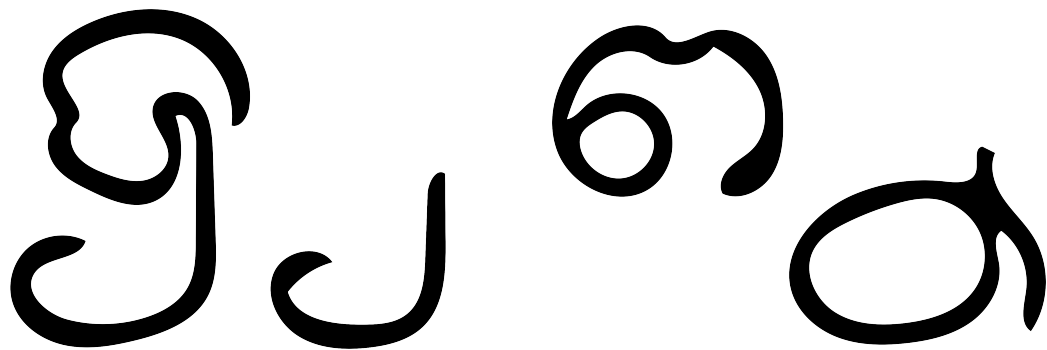◎ handled by font? » rendering engine needs?

◎ dictionary? » decomposability?

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

◎ Lanna and documents and fonts (inclusive of Khʉn) indicate non-distinct homography

Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

◎ Lanna / Khʉn : subjoined ṭha = subjoined ba

◎ Lʉ / 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 / Khửn
  ○ subjoined ṭha = subjoined ba

◎ Lử / Lao / Isan
  ○ subjoined ṭha ≠ subjoined ba

Isan example of graphic distinction
(Sawing Bunčhŏm, 2009)

Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

◎ Lanna / Khɯn : subjoined ṭha = subjoined ba
◎ Lɯ / Lao / Isan : subjoined ṭha ≠ subjoined ba

Lɯ example of graphic distinction
(刀新华, 1989)

Lɯ example of graphic distinction
(Khotsimeuang, 2001)

Disunification? TAI THAM CONSONANT SIGN HIGH RATHA OR LOW PA (U+1A5B)

◎ Tham
   ○ Lanna / Khưn : subjoined ṭha = subjoined ba
   ○ Lư / Lao / Isan : subjoined ṭha ≠ subjoined ba
◎ Myanmar
   ○ homographic subjoined consonants not documented (yet?)
◎ Khmer
   ○ subjoined ṭa = subjoined ta

glyphs	codepoints	unicode names
ដ្ដ	178A 17D2 178A	da coeng da
ត្ត	178F 17D2 178F	ta coeng 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

Lek-tho/Mai-sam (U+1A7B) : a scoped relationship

ဂွံ့	(ဂွံ့)ˊ = ဂွံ့ ဂွံ့	khọy khọy	'gradually, little by little'
လှွဲ့က	သ ( ) လှွ	să-wằng (*swằng)	'to pursue, seek for'
မီ့ည	မီ (မ)ည	mī māk	'to have a lot'
တဏှ့	တဏှ (၁)	tanhā	'loneliness'

As yet documented for handling

◎ use of other numbers

◎ hāng-læm "tail" letters

◎ other vowel forms

◎ akkhara piset

◎ 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 kǭ =
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

Padauk	Lucida Sans	Tai Tham KH	LN Tilok	

"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

	plā	'fish'
	talāt	'market'

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
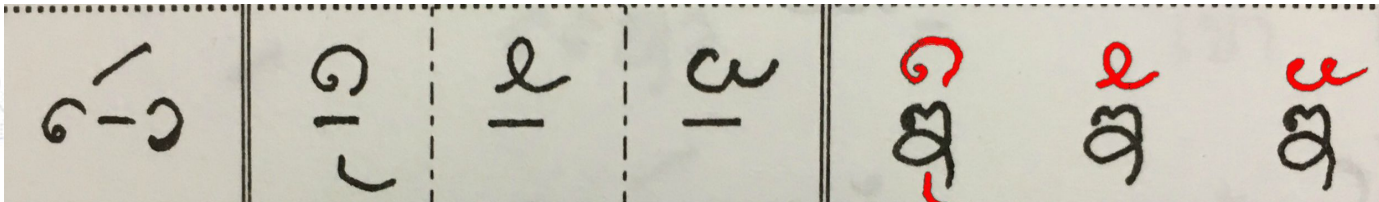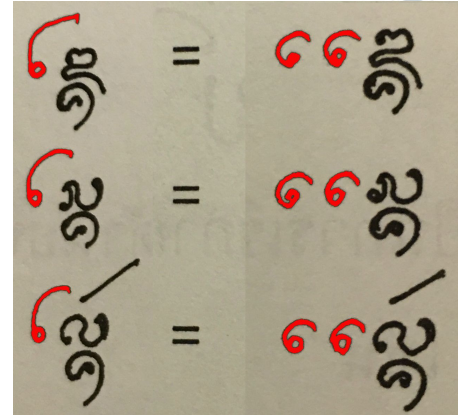
Summary 1/2

◎ 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

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/)
◎ □□□□□□□ (http://www.kengtung.org/)

Attributions

Fonts used in this presentation include:

- ◎ LN Tilok / LN Monsaen
    - ○ https://drive.google.com/folderview?id=0B3hdjepXeOuaRGVKVVlETXVTcGM&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
  ○ http://www.unicode.org/versions/Unicode8.0.0/ch16.pdf
◎ code charts
  ○ http://www.unicode.org/charts/PDF/U1A20.pdf

Bibliography:

◎ http://bit.ly/1O0cN0B

Thank you very much !

ယူႅႅႄ   ꪨꪚꪆꪙꪲꫀ   ꪁꪳꪬꪚ꫁ꪚꪲꪚ

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