WG2 N4939
2018-02-20

Thai-Noi Transliteration

L2/18-068

M. Hosken,
Payap University,
NRSI SIL International

This is a response document to ISO/TC 46/WG 3 CD20674_4 N2631 also Unicode L2/18-042.

Unfortunately I do not have a font to reference characters of interest. I will therefore use the No. column in the chart in section 6, the Romanization table.

Introduction

N2631 describes a transliteration scheme for representing Thai Noi using the Thai script. It is not aimed at encoding Thai Noi using the Thai script block of Unicode. This difference is significant in that the mapping, therefore, does not need to conform to the general shaping and encoding requirements needed to encode a script fully. The only requirement for transliteration is that the data can be round tripped. It doesn't actually matter which characters characters are transliterated into, although it helps significantly if the mapping can be largely self describing to a human reader.

At first glance, Thai Noi seems to fall within a script continuum including Tai Tham and Thai. Some shapes resemble Thai and some Lanna (a particular instance of a Tai Tham script). This makes it tempting to try to encode Thai Noi using the Thai script block (or the Tai Tham block). But in either case it does not fit directly without changing the block. Of the two, Thai Noi would fit more naturally within Tai Tham than Thai. This is because Thai has lost much of the behavioural complexity that is in Thai Noi, whereas Tai Tham supports the behaviour (and much more).

But the objective here is not to encode Thai Noi but to transliterate it into Thai script. As such, the aim should be to not have to extend the Thai script, if possible. Any additions to the Thai script are going to confuse readers since they will not be familiar with the additions; fonts will not support the extensions and keyboards will not be able to type the transliteration. This document examines how Thai Noi could be transliterated into Thai script without extending the character repertoire.

A note of caution needs to be raised regarding Thai Noi. The script clearly exists, but the precise specification of the script is not well attested. Through lack of data, it has not been possible to assess whether all the characters are truly contrastive or whether in some cases there is glyph and character variation involved. In particular I would value evidence of the following contrasts:

- a (char 37) vs a̲ (char 80)
- The existence of l̤ (char 34)
- The fact that Thai Noi has an extra ssa (chars 32, 74)
- I disagree that there is a 3 way contrast of subjoined ya (48, 71, 79) and the 48 and 79 are the same character, or perhaps some other merging. But that there is only a 2 way contrast

between subjoined ya. Char 78 is suspect too. This whole collection of characters needs further analysis.

- The extra tha (chars 14 vs 69). No other Thai/Tai script has such a contrast.
- The extra la (chars 28 vs 73). No other Thai/Tai script has such a contrast.

We proceed taking the Thai Noi characters on face value.

We concentrate on the characters proposed for extending the Thai script block. These are characters numbers 46-48, 54-80.

Behaviour

Subjoined Consonants

The first big difference between Thai Noi and Thai script is that Thai Noi makes limited use of subjoined characters. Such characters are prevalent in Tai Tham, but Thai Noi shows evidence of the transition towards the simpler Thai script. As such Thai Noi only uses subjoined characters in a few limited contexts:

- Subjoined n and m syllable initially.
- Medial and final y.
- Medial r and l.
- Subjoined s.

N2631 proposes adding new characters to the Thai script block. But there is evidence of other subjoined consonants, and given the wide variance in this script, a more generic mechanism for representing subjoined characters in transliteration would be suitable. Here I propose the use of the phintuu dot on the parent character. Thus, for example:

Character number	Romanization	Thai transliteration
47	ḷ′	◌ฺล
48, 79	ẏ	◌ฺย
64	h̄.n	ห◌ฺน
67	k̄h.m	ฃ◌ฺม

Just as with the romanization, the Thai transliteration is legible in the same way. If the text is read ignoring the dots, it will sound the same as the original Thai Noi. The dot is included to ensure round tripping.

This behavioural model accounts for the following characters: 47, 48, 64-68, 70, 72, 75-77 and is extensible into other subjoined characters not listed in N2631.

Characters

Alternatives

There are a few proposed characters that can take standard Thai characters, even if they are not ideal. But since this is transliteration and not encoding, so long as the characters used bear some appropriate phonetic resemblance to the characters being transliterated, they are an option.

Character Number	Romanization	Thai transliteration
69	t̄h′	ฑ
73	l′	ฦ

Subjoins

Some characters are not obviously subjoined characters but transliterate well as if they were:

Character Number	Romanization	Thai transliteration
78	x′y	อฺย
74	s̄′	อฺส
71	xy′	◌ฺอฺย
80	a̲	ะ̲
46	a′	◌ั

The logic here is based on the first entry in this table. As in Tai Tham, there is no subjoined character after an o ang. As such o ang, like ho hip, is used for changing a character's class. This can be used to modify character classes. The phintuu dot here is used not for subjoining but for marking an alternative character. This also applies to sara a. The marking of mai hanakat is done on the preceding consonant.

Digits

Given this is transliteration, there is no reason for Thai digits to occur in Thai Noi text, and so Thai Noi digits may be transliterated using Thai digits. This covers characters 54-63.

Conclusion

It seems odd that there is a desire to transliterate a script before it has even been encoded. Is this an attempt to encode it via the back door? It would help everyone if a well researched orthography statement can be written from which the analysis of both encoding and transliteration can be done.

The Royal Academy is encouraged to relate to other experts in the area of Thai scripts as they do their work.