

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: Working Group Document
Title: Lanna ad-hoc report
Source: Chen Zhuang, Michael Everson, Martin Hosken, Wei Lin-Mei
Status: Ad-hoc report
Action: For consideration by JTC1/SC2/WG2
Date: 2006-09-27

The ad-hoc group on the encoding of the Lanna character set discussed the feedback documents from China in N3161 and from Myanmar in N3159. This document responds to the concerns raised there.

Chinese comment I. The ad-hoc group believes that the name “Lanna” is the best name for this script in the English language. The Lanna script is used to write four languages: Khün, Lao Tham, Northern Thai, and Tai Lue. The name proposed, “Old Tai Lue” is accurate in distinction to the “New Tai Lue” script already encoded, but would not be appropriate as a generic name for this script given its use in Myanmar, Thailand and Laos.

Chinese comment II. The Lanna script must be encoded distinctly from the New Tai Lue script, despite the resemblance between the characters. The two scripts function quite differently. If they were unified, the productive subjoiner SAKOT would be introduced into the New Tai Lue script. This would cause instability in texts using this script in China.

Chinese comment III. The Chinese experts reviewed N3121, which did not include GREAT SA. The revision, N3121R, does include this character, encoded at U+1A5E. The other glyphs shown are represented in the Lanna encoding as U+1A28 and U+1A60 followed by U+1A49.

Chinese comment IV. Section 2 of N3121 gives a list of subjoined consonants which is the mechanism used for handling these additive signs. Specifically *jawhoi* is encoded as U+1A60 (SAKOT) U+1A41 (LOW YA).

Hang characters. In addition to the subjoined mechanism, there are two ways that tails are used to create new characters. Where such a tail is used to create a new character, the new character receives its own encoding as shown in the chart (specifically U+1A30, U+1A4C and U+1A55). But for the first two glyphs shown (which are variants of the same thing), there is also a special character U+1A5D (SIGN HIGH RATHA OR LOW PA) may be used where the tail is not itself part of another character.

Chinese comment V. Again, the subjoined consonant mechanism is used for finals as per section 2. Specifically each of the characters listed follows 1A60 (SAKOT): 1A2F, 1A34, 1A36, 1A37, 1A3D, 1A29, 1A21, 1A23, 1A25, 1A26, 1A43, 1A27, 1A28.

Myanmar comments. The Lanna script is generally accepted as coming from Old Mon, which bears very little resemblance to the Mon script of today. Specifically, Lanna bears little resemblance to Burmese, Karen and Shan, even if some similarities can be seen, the behaviour of Lanna is much more complex. Specifically the process of subjoining in Lanna is used to represent: finals, medials, vowels and syllable chaining, of which only syllable chaining and medialisation is common to Burmese.

This complexity has resulted in an encoding model that only encodes a medial separately if there is a visual distinction. Specifically medial ya does not occur as a medial but is acting as a vowel and has no difference in shape to subjoined ya. Wa functioning as a vowel and wa functioning as a medial take the same shape, and are not formally distinguished in the writing system. Both are represented by the SAKOT + WA sequence.

The ad-hoc recommends to WG2 that the Lanna script be entered into the ballot for a new PDAM 4.