

TO: Unicode Technical Committee

FROM: Debbie Anderson, SEI, UC Berkeley

DATE: 10 October 2007

RE: Feedback on L2/07-299 Preliminary Proposal for Northeastern Yunnan Simple Miao (WG2 N3335)

The following are comments from a few UTC members regarding the new preliminary proposal on NE Yunnan Simple Miao by China.

This proposal contains some useful information, but questions were raised about the overall encoding model. Please see comments below.

1. The proposal requests that this be called "Northeastern Yunnan Simple Miao" script. That is fine as an alternative name or alias for the script, but it is much too unwieldy as a formal script name for the standard. The character names proposed in fact use "SIMPLE MIAO" for the names, which would then be the script and block names. That is much better. In fact, just "MIAO" would be even better, leaving the option for the future that some old script might be uncovered, and named "OLD MIAO" or some other variant.

2. The proposal, in passing, requests that ISO 15924 change the registration for Pollard to "Northeastern Yunnan Simple Miao". That is simply impossible, based on the stability requirements for ISO 15924 registrations. The name also has the same problem mentioned in #1. "Simple Miao" or just "Miao" would be a better registration. What *can* be done is for aliases to be added to ISO 15924 for the existing registration. In any case, that is something to be taken up with the ISO 15924 registrar, and is not an issue for WG2.

3a. Status of precomposed initials. This is not sufficiently argued. I don't see any strong argument for the prenasalized digraphs to be encoded as single characters. The proposal is encoding the *phoneme* chart from Wang Yangcai 2005, rather than encoding the graphemes necessary for writing text.

3b. The proposal makes an apparently unwarranted distinction between the grapheme for the prenasalization and the grapheme for the independent initial "NR". This is also based on an analytic decision based on phonological considerations in Wang Yangcai 2005, but is not actually reflected in printed text examples.

3c. The argument for encoding the digraphic initial consonants with apostrophe "Y", etc. as units is stronger, I think, because the placement of vowels is with respect to the base letter, and not with respect to the apostrophe. Letter + apostrophe + vowel trigraphs could be handled as a sequence of 3 elements, but it is probably simpler to build rendering systems that deal with the vowel placement correctly if the apostrophe diacritic letters are simply encoded as single units.

3d. The dotted series of "stressed" letter forms looks to me like a matter of phonological analysis. I don't see any evidence of those in written forms, and without indication of how they would be used, encoding an entire set of them precomposed doesn't make sense. Attestation of usage outside of Wang Yangcai 2005 is needed. And the only usage I spotted there, in the row of tonal letter examples, suggests to me that the way these are used with respect to vowels and tone marks are fundamentally different. Is this a matter of formal specification for the 1988 recension?

4a. Vowels. The proposal suggests encoding each single vowel symbol, which is probably the correct thing to do. However, it is silent about the "complex vowels" cited in the second row of vowels in Wang Yangcai 2005, many of which are -VN type rhymes. It is presumed that those are simply produced with sequences of the vowel letters, but that introduces some serious rendering issues in the model which the proposal says nothing about.

4b. Beyond that, there are some issues of identity of the final nasal vowel forms. The second row, with complex vowels, shows a final nasal not obviously identified in the first row. I'm going to guess that this is an alloglyph for -ang. This might explain the bizarre duplicate entry at 0x86 in the proposal, for SIMPLE MIAO LETTER ANG. Cf. 0x7D, also called SIMPLE MIAO LETTER ANG. Unless I'm missing something, I see no reason why these shouldn't be unified as a single character with a positional rendering distinction.

4c. Also for the complex vowels, Illustration 3 shows a common digraphic vowel with the second part being simply a right-facing cup. but would be accounted for if this were an alloglyph of -an (making a complex rhyme, -in). This kind of graphic variation in very common forms needs to be explicitly explained in the proposal, instead of simply encoding the list from Wang Yangcai 2005 without explanation.

5a. The glyphs for the "tone value markers" are wrong, and need to be redone. It is inappropriate to use full-size serifed Latin letters here, when the examples clearly show these are small forms on the baseline, comparable in size and style to the vowel letters.

5b. The encoding of an unmarked tone, 0x90 SIMPLE MIAO TONE VALUE 1, annotated as having "no graphical value" and showing a dotted circle glyph, is simply a mistake. Once again, this is attempting to encode the *tonemes* of the system, rather than the *graphemes*. This needs to be removed.

6. The encoding and explanation of the "tone category markers" is unexplained. As five graphic markers, implicitly indicated as combining marks (and all given gc=Mn general category property, confirming that intent), they make no sense -- and would in fact be confused with the vowel letter -a. I *think* the intent here is to encode five format control characters that are to indicate the position where the vowel is rendered, so as to indicate the tone for the syllable. (Presumably this is for the 1936 recension, so-called.)

The section in the proposal, "Character storage order" claims that "the placement of the tone marker is arbitrary", but then goes on to make the case that it *isn't* arbitrary, and that both the 1936 recension and the 1988 recension should be stored in CV(V)T order. If that is the case, and if the "T" in the 1988 recension is an explicit visible tone letter -- the 7 "tone value markers" -- then the cleanest representation for the Simple Miao script would at this point be to start from the tone letters for the 1988 recension, and simply map them onto the five positions when rendering for the 1936 recension. This positional mapping should be handled as a *font* (rendering system) issue, rather than a respelling of the text with 5 dubious positional "tone category markers", which would require the same complex font and rendering system support in any case.

This needs to be worked out in much more detail in the proposal before it is mature -- including explicit examples of Simple Miao text, showing the representation that would be used in terms of sequences of the proposed encoded characters, both for the 1936 recension and the 1988 recension -- so that it will be clear how to actually use the proposed encoding to represent text.

Other comments: Remove cross-references to UCAS.