

Ned Holbrook

holbrook@unicode.org

Script Encoding WG

October 3, 2025

Feedback on Mwangwego encoding model

The current proposal for Mwangwego ([L2/25-135](#)) is for an encoding model using visual order, which complicates many text processes and is likely to result in a suboptimal experience for users of the script. In my opinion it would be best to adopt a logical encoding model to minimize barriers to adoption of the Mwangwego script.

According to [TUS 17.0.0 § 2.2.6](#): “The order in which Unicode text is stored in the memory representation is called logical order. This order roughly corresponds to the order in which text is typed in via the keyboard; it also roughly corresponds to phonetic order.” For Mwangwego such an ordering is called Siri-Musisi-Mutuyo, more generically Consonant-Vowel-Modifier (C-V-M). Mutuyo occur on their own as either spacing or non-spacing marks, or as multiple Mituyo, stacks of spacing Mutuyo that are encoded as 16 predefined combinations, or as a combination of one non-spacing Mutuyo and one Mituyo stack. A Mituyo stack is displayed to the left of the cluster [Mwangwego is a LTR script]. The question here is whether Mituyo stacks should be encoded in logical or visual order. Students are taught logical order and that is how keyboards are implemented, so already the choice of logical order seems obvious. It might help, however, to examine the problems that come with visual order before deciding.

The most obvious problem with visual order may be seen in the current chart, which is that malformed sequences of Mituyo stacks will not be validated by a shaping engine: this results in the absence of a dotted circle for a standalone Mituyo stack, as that behavior comes from a shaping engine. Even if implementers felt compelled to provide a dedicated (non-USE) shaping engine, realistically it is not helpful to assume one will come: New Tai Lue script, which has used visual order since Unicode 8.0, has no known dedicated shaping engine. Plus, there are other user-perceived behaviors that are directly influenced by the encoding model.

Segmentation, for one, relies on properties derived from the General Category of characters. Since spaces separate Mwangwego words we need not consider segmentation by line or word, but cluster segmentation will be degraded by the choice of visual order. Without some form of script-specific tailoring (or reconsidering the use of gc=Lm for Mutuyo and Mituyo stacks), text selection would treat a spacing Mutuyo and Mituyo stacks as independent clusters. And while it is likely that a Mwangwego keyboard would be capable of intelligently handling the deletion of a Mituyo stack, such behavior would not be available to other input methods.

Collation, too, is much easier to implement when the encoding order aligns with the expected sort order. Since a consonant with a Mituyo stack would sort after the bare consonant it seems likely that a logical order would avoid the need for contractions to properly order consonants relative to sequences with a preceding Mituyo stack it seems likely that logical order would be preferable, but I am not expert in collation and have not considered the interactions between a Mutuyo and a Mituyo stack in depth.

It would seem that the two main points in favor of visual order are ease of font implementation and user expectations, neither of which seem to be in evidence. Where font

implementation is concerned, the USE is capable of handling both the glyph reordering needed for a Mituyo stack and basic cluster validation. Where users are concerned, pedagogy and keyboards already favor logical order. It has been noted that script users may eventually progress to the point where they will write a Mituyo stack first, but using this as the basis for visual order is not defining; individual characters are not defined as a series of strokes, nor must an encoding accommodate every form of manuscript embellishment.

Even if there remain questions that are not entirely resolved by logical order, on balance it appears to be better aligned with user expectations and solves more problems than it causes. But more importantly, logical order is explicitly preferred by the Unicode Standard, and if the case study of New Tai Lue is any guide, visual order should only be an option when that choice is supported by a preponderance of evidence from the user community. In the absence of such evidence, I endorse logical order for the Mwangwego script.