

DATE: 10 June 2019
DOC TYPE: Working Group Document
TITLE: Comments on WG2 #68 documents
SOURCE: Deborah Anderson, Ken Whistler, Roozbeh Pournader, Liang Hai, Peter Constable, and Lisa Moore
STATUS: Expert contribution
ACTION: For consideration by WG2
DISTRIBUTION: ISO/IEC JTC1/SC2/WG2

The following are comments by a group of experts on Mongolian, Small Khitan, and other WG2 #65 documents.

JIANZI MUSICAL NOTATION

Document: [N5041](#) Preliminary proposal on encoding Jianzi Musical Notation and Jianzi Format Controls ([L2/19-107](#))

The proposal includes a total of 357 characters (with 21 format controls). The notation uses components that are put together in a manner similar to ideographs. The proposal was discussed at the May 2019 IRG meeting, but the IRG decided Jianzi is out of scope, except for confusables (with CJK).

The notation is still being analyzed. A doc focused on the structural analysis of the script and complete rationale behind the proposed encoding model is recommended.

LATIN

Tironian letters

Documents: [N5042](#) Proposal to add one or two Latin Tironian letters ([L2/19-172](#))-- Everson and West
Related:

[N4908](#) Proposal to add six Latin Tironian letters ([L2/17-359](#))

[N4885](#) Comments on WG2 #66 (Sept. 2017) documents ([L2/17-367](#))

[L2/17-384](#) Recommendations to UTC #153 October 2017 on Script Proposals

The main request of this proposal is to add an uppercase letter TIRONIAN ET, preferably with a casing relationship, either to U+204A TIRONIAN SIGN ET, requiring a change of the General Category property of U+204A from Punctuation, other (Po) to Letter, lowercase (Ll) (Option 1), or by adding a new case pair (Option 2). A third alternative, recommended in [N4885](#) ([L2/17-367](#)) and [L2/17-384](#) is to add a new non-casing TIRONIAN ET (Option 3).

As noted in this proposal, U+204A TIRONIAN SIGN ET was encoded already in Unicode 3.0 in the Punctuation block, with General Category property Po.

The following are comments from the experts:

- In our opinion, the number of users who would need uppercase is very small, compared to those using U+204A. (Note that U+204A is already described as “an infrequently used character” on page 3 of the proposal.) Hence, the requirement for having automatic uppercasing is not justified. How does not having automatic casing behavior cause a major problem for users? Medievalists would typically be entering a text as it appears on manuscripts or other

documents, rather than expecting automatic casing. Search could be tailored to find uppercase characters as needed.

- We agree that adding a new lowercase TIRONIAN ET (Option 2) would cause confusion for users, since the new lowercase character and U+204A would be nearly identical in shape.
- TIRONIAN ET is much like “&”: it acts as a mark of punctuation (“hot & cold”) and is letter-like (&c.). In Unicode, even though a character may act as a letter in some instances and in other cases as a mark of punctuation, only one value of the partition property `General_Category` can be assigned to the character, so a decision must be made. (AMPERSAND, like TIRONIAN ET, are both Po.)

In sum, we agree a new character is warranted, and suggest the name TIRONIAN SIGN CAPITAL ET with `General_Category=Po` and no case mapping. We recommend the code point U+2E52 in the Supplemental Punctuation block, since the earlier U+204A TIRONIAN SIGN ET is already in a punctuation block, and a cross-reference to U+204A TIRONIAN SIGN ET be added.

Four Latin Characters

Document: [N5044](#) Proposal for the addition of four Latin characters ([L2/19-179](#)) -- Everson and Lilley

This proposal requested four characters: two casing pairs for Latin letter s with stroke overlay and Latin letter d with stroke overlay. The characters were used to write the Gaulish language in Roman Gaul. Latin letter d with stroke overlay (“TAU GALLICUM”) is also attested in an orthography used to write the Moro language of Sudan.

In our view, the name TAU GALLICUM would not be a useful name to apply to an unrelated modern language. Instead, our preference is a name that reflects the archetype of the character, describing the diacritic. We recommend the following parallel names:

LATIN CAPITAL LETTER D WITH SHORT STROKE OVERLAY

LATIN SMALL LETTER D WITH SHORT STROKE OVERLAY

LATIN CAPITAL LETTER S WITH SHORT STROKE OVERLAY

LATIN SMALL LETTER S WITH SHORT STROKE OVERLAY.

If no characters are proposed in the positions U+A7C7...U+A7DA, we suggest those positions be used (instead of starting a new column U+A7D0...U+A7D3).

Two Characters for Middle Scots

Document: [N5045](#) Proposal to add two characters for Middle Scots ([L2/19-180](#)) – Everson

In the view of the experts, a ligature of <017F LATIN SMALL LETTER LONG S, 0073 LATIN SMALL LETTER S> would suffice, and users need not wait several years. Alternatively, users could employ U+00DF LATIN SMALL LETTER SHARP S (with uppercase 1E9E), making minor modifications to the font, if necessary. No evidence was provided for an uppercase character.

LISU

Document: [N5047](#) Proposal to Encode the Lisu Monosyllabic Scripts ([L2/19-208](#)) – China NB

This new proposal requests 989 characters.

The following were raised during discussion amongst experts:

- Provide background on the two documents appended to the proposal (i.e., starting pp. 76ff. and p. 91ff.), and their relation to one another.
- How are the characters ordered? Is there a native order or Romanized order? The order on page 76 appears to be left-to-right, but the document starting on p. 91 has the characters ordered vertically. Which is the traditional order?
- Because of the large repertoire, we recommend the proposal be reviewed carefully by experts.
- The glyphs in the “Lisu letter” column in the charts on pages 3-4 need to be embedded in the PDF to facilitate accurate review of the document.

TANGUT

Document: [N5064](#) Proposal to encode nine Tangut ideographs and six Tangut components ([L2/19-207](#))

– West

Related: [N5031](#) Investigation of Tangut unification issues ([L2/19-064](#)) – West and Zaytsev

This proposal reflects the consensus opinions of experts, who have reviewed the material in [N5031](#) and confirmed semantic differences between the characters, which are composed with different components. We agree the proposed nine Tangut ideograph characters and six Tangut components are justified. We also welcome the use of Ideographic Variation Sequences for minor glyph differences.

The proposed new block from U+18D00..U+18D7F should, in our view, should be named “Tangut Supplement,” since it appears not more than 128 characters are needed. The block name “Extended” tends to be used with the extensions -A, -B, etc., particularly in cases where many character additions are expected. This does not appear to be the case for Tangut.