

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

Doc Type: Working Group Document
Title: Additional Tangut Glyph Corrections
Source: Andrew West and Viacheslav Zaytsev
Status: Individual Contribution
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2019-12-21

The repertoire of encoded Tangut ideographs is primarily based on Li Fanwen's 2008 *Tangut-Chinese Dictionary*, and the code chart font is based on the font used in Li Fanwen's dictionary. However, recent research into Tangut orthography has identified a number of errors in the glyph forms shown in Li Fanwen's dictionary. The following documents written by West, Zaytsev, *et al.* have already resulted in a number of corrections in the Tangut code charts:

- WG2 N4588R2 / L2/14-209R (2014-10-01) identifies 76 corrections for Tangut ideographs
- WG2 N4723 / L2/16-112 (2016-04-21) identifies 3 corrections for Tangut ideographs
- WG2 N4850 / L2/17-313 (2017-09-07) identifies 31 corrections for Tangut ideographs and one Tangut component

As discussed at WG2 Meeting 68 at Redmond, Washington, USA in June 2019, there are still a number of outstanding glyph errors in the code charts (see L2/19-267 "Brief IR Report from WG2 Meeting #68" by Deborah Anderson). These are in addition to the systematic glyph errors discussed in N5031 (L2/19-064) and N5064 (L2/19-207).

The current document, which is the result of a careful comparison of all entries in the A and B editions of the *Homophones* (𐰚𐰍𐰏 ·êi leu = 同音 *Tóngyīn*) with the code chart glyphs, identifies a further 42 corrections for Tangut ideographs and four corrections for Tangut components. We hope that this document captures all or most of the outstanding glyph errors in the code charts (excluding those relating to N5031 and N5064), although of course we cannot exclude the possibility that other glyph corrections will be identified in the future. There are also a number of cases where the A and B editions of *Homophones* have somewhat different glyph forms, but it is not possible to be certain which is the correct glyph form to use in the code charts. In such cases we have not proposed a glyph change in this document, but anticipate that the differences will be registered in a future IVS collection for Tangut.

The detailed evidence and discussion for the proposed glyph changes are presented in UTN 42 version 2, and so the current document only provides a summary of the requested glyph changes in the tables below.










Table 1. Summary of Glyph Corrections for 42 Tangut Ideographs

Code Point	Original Glyph	New Glyph	Sample*	Original Radical/ Strokes	New Radical/ Strokes	UTN 42 Section
U+17097	𗀓	𗀓		2.12	2.12	3.2
U+1709D	𗀔	𗀔		2.12	2.13	3.9
U+171CA	𗀕	𗀕		26.7	26.7	3.10
U+172BA	𗀖	𗀖		36.9	36.9	3.4
U+172EA	𗀗	𗀗		36.11	36.11	3.4
U+17326	𗀘	𗀘		36.15	36.15	3.4
U+1732C	𗀙	𗀙		36.17	36.18	3.11
U+173C6	𗀚	𗀚		65.12	65.13	3.14
U+173D5	𗀛	𗀛		68.10	147.11	3.15

Code Point	Original Glyph	New Glyph	Sample*	Original Radical/ Strokes	New Radical/ Strokes	UTN 42 Section
U+174DF	薊	薊		75.14	75.14	3.16
U+174F7	薊	薊		75.14	75.14	3.4
U+17538	薊	薊		75.19	75.19	3.17
U+1753F	徧	徧		76.8	76.8	3.18
U+17540	徧	徧		76.8	76.8	3.18
U+177B7	薊	薊		106.18	106.18	3.4
U+177DA	薊	薊		107.11	107.10	3.6
U+1795D	徧	徧		141.11	141.11	3.10
U+17A87	徧	徧		141.15	141.14	3.5
U+17C20	徧	徧		177.11	177.11	3.4









Code Point	Original Glyph	New Glyph	Sample*	Original Radical/ Strokes	New Radical/ Strokes	UTN 42 Section
U+17CFC	𦉑	𦉑		194.13	194.12	3.21
U+17D15	𦉒	𦉒		200.9	200.9	3.22
U+17E7D	𦉓	𦉓		245.12	245.12	3.4
U+18111	𦉔	𦉔		308.15	308.16	3.17
U+1814A	𦉕	𦉕		316.12	316.12	3.25
U+18215	𦉖	𦉖		350.12	350.12	3.2
U+18216	𦉗	𦉗		350.12	350.12	3.2
U+18217	𦉘	𦉘		350.12	350.12	3.2
U+18259	𦉙	𦉙		376.13	376.13	3.10
U+182E6	𦉚	𦉚		412.10	301.9	3.28

Code Point	Original Glyph	New Glyph	Sample*	Original Radical/ Strokes	New Radical/ Strokes	UTN 42 Section
U+18380	𧯛	𧯛		431.13	431.14	3.29
U+18438	𧯜	𧯜		436.16	436.15	3.30
U+18493	𧯝	𧯝		458.15	458.15	3.32
U+185F2	𧯞	𧯞		553.12	553.12	3.3
U+18620	𧯟	𧯟		565.15	565.15	3.4
U+1865D	𧯠	𧯠		568.16	568.16	3.4
U+18702	𧯡	𧯡		651.11	651.10	3.5
U+18761	𧯢	𧯢		688.11	688.10	3.6
U+18762	𧯣	𧯣		688.15	688.14	3.6
U+187B0	𧯤	𧯤		719.16	719.15	3.33

Code Point	Original Glyph	New Glyph	Sample*	Original Radical/ Strokes	New Radical/ Strokes	UTN 42 Section
U+187D0				736.12	736.11	3.7
U+187D1				736.17	736.16	3.7
U+187D2				736.21	736.20	3.7

* The column headed “Sample” gives a sample image of the Tangut ideograph from an original woodblock printing of one of the *Homophones B* editions, except in the cases of U+187D2 and U+182E6 which show images from the *Homophones A* edition, and U+177DA, U+1865D, and U+18761 which show images from the *Sea of Writing*. See the referenced section of UTN42 for the exact source reference for the image.

Table 2. Summary of Glyph Corrections for 4 Tangut Components

Code Point	Component	Original Glyph	New Glyph	Original Strokes	New Strokes	UTN 42 Section
U+18A28	553			7	7	3.3
U+18A8A	651			8	7	3.5
U+18AAF	688			9	8	3.6
U+18ADF	736			11	10	3.7

Potential Tangut Component Disunifications

UTN 42 Sections 3.2 and 3.4 respectively discuss the separation of Tangut Component 350 𐰇 (U+1895D) and Tangut Component 565 𐰇 (U+18A34) into two different components, as shown in the table below. The second glyph form of each component may be proposed for encoding in a future proposal for additional Tangut components.

Table 3. Summary of Potential Tangut Component Disunifications

Code Point	Component	Glyph	Used in
U+1895D	350	𐰇	𐰇𐰇𐰇𐰇𐰇𐰇𐰇
		𐰇	𐰇𐰇𐰇𐰇
U+18A34	565	𐰇	𐰇𐰇𐰇𐰇𐰇𐰇𐰇𐰇𐰇
		𐰇	𐰇𐰇𐰇𐰇𐰇𐰇𐰇𐰇 𐰇𐰇𐰇𐰇𐰇𐰇𐰇𐰇

Tangut Data Changes

18 of the proposed glyph corrections for Tangut ideographs involve a change in radical and/or total stroke count, which if accepted would necessitate a change in the Tangut data files for the Unicode and ISO/IEC 10646 standards. The revised lines in the data files are listed below.

```

U+1709D   kRSTUnicode   2.13
U+1732C   kRSTUnicode   36.18
U+173C6   kRSTUnicode   65.13
U+173D5   kRSTUnicode   147.11
U+177DA   kRSTUnicode   107.10
U+17A87   kRSTUnicode   141.14
U+17CFC   kRSTUnicode   194.12
U+18111   kRSTUnicode   308.16
U+182E6   kRSTUnicode   301.9
U+18380   kRSTUnicode   431.14
U+18438   kRSTUnicode   436.15
U+18702   kRSTUnicode   651.10
U+18761   kRSTUnicode   688.10
U+18762   kRSTUnicode   688.14
U+187B0   kRSTUnicode   719.15
U+187D0   kRSTUnicode   736.11
U+187D1   kRSTUnicode   736.16
U+187D2   kRSTUnicode   736.20

```

References

- L2/19-267.** Deborah Anderson, IR. *Brief IR Report from WG2 Meeting #68 (Redmond, WA, June 2019)*. L2/19-267. 2019-07-03. <http://www.unicode.org/L2/L2019/19267-ir-report-wg2.pdf>
- N4588R2.** Andrew West, Viacheslav Zaytsev, Sun Bojun, Michael Everson. *Tangut glyph corrections*. ISO/IEC JTC1/SC2/WG2 N4588R2 (L2/14-209R). 2014-10-01. <http://www.unicode.org/L2/L2014/14209r-n4588.pdf>
- N4723.** Andrew West, Viacheslav Zaytsev, Sun Bojun. *Glyph Corrections for 3 Tangut ideographs*. ISO/IEC JTC1/SC2/WG2 N4723 (L2/16-112). 2016-04-21. <http://www.unicode.org/L2/L2016/16112-n4723-tangut-glyph-corr.pdf>
- N4850.** Andrew West and Viacheslav Zaytsev. *Corrections for 31 Tangut ideographs and one Tangut component*. ISO/IEC JTC1/SC2/WG2 N4850 (L2/17-313). 2017-09-07. <http://www.unicode.org/L2/L2017/17313-n4850-tangut.pdf>
- N5031.** Andrew West and Viacheslav Zaytsev. *Investigation of Tangut unification issues*. ISO/IEC JTC1/SC2/WG2 N5031 (L2/19-064). 2019-02-10. <http://www.unicode.org/L2/L2019/19064-tangut-n5031.pdf>
- N5064.** Andrew West, Viacheslav Zaytsev (Institute of Oriental Manuscripts, Russian Academy of Sciences), Jia Changye (Ningxia Academy of Social Sciences), Jing Yongshi (Beifang University of Nationalities), Sun Bojun (Institute of Ethnology and Anthropology, Chinese Academy of Social Sciences). *Proposal to encode nine Tangut ideographs and six Tangut components*. ISO/IEC JTC1/SC2/WG2 N5064 (L2/19-207). 2019-05-27. <http://www.unicode.org/L2/L2019/19207-n5064-tangut.pdf>
- UTN 42.** Andrew West and Viacheslav Zaytsev. Unicode Technical Note No. 42: *Tangut Character Additions and Glyph Corrections*. Version 2. 2019-12-21. <https://www.unicode.org/notes/tn42/>