

ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)

Doc. Type: Disposition of comments

Title: Disposition of comments on PDAM2.2 to ISO/IEC 10646 4th edition

Source: Michel Suignard (project editor)

Project: JTC1 02.10646.00.02.00.04

Status: For review by WG2

Date: 2015-01-22

Distribution: WG2

Reference: SC2 N4379 4387 WG2 N4651

Medium: Paper, PDF file

Comments were received from the following members: China, Germany, Ireland, Japan, Mongolia, UK, and USA. The following document is the disposition of those comments. The disposition is organized per country.

Note – With some minor exceptions, the full content of the ballot comments have been included in this document to facilitate the reading. The dispositions are inserted in between these comments and are marked in **Underlined Bold Serif text**, *with explanatory text in italicized serif*.

This disposition, by postponing the repertoires that were problematic (and behind many negative votes), could possibly have led to an enquiry ballot (Draft Amendment) for the next phase.

However, another consideration is the urgency to encode another set of Emoji characters that are currently in review by the Unicode Consortium to improve the diversity of the currently encoded Emoji. To give National Bodies a chance to also review these proposed additions, it is wise to conduct another round of proposed draft amendment (pdam2.3) including these additional Emoji characters along with the result of this disposition. This will also validate the subset created by this disposition and confirm the consensus required to conduct an enquiry level ballot.

Most of the repertoire removed from this amendment will be resubmitted as new proposed repertoire for the Committee Draft of the 5th edition of 10646 that will be initiated after this pdam2.3.

Changes to the repertoire of amendment 2 after disposition of comments:

Deletion (existing or remaining proposed block):

U+9FD1..U+9FE9 CJK UNIFIED IDEOGRAPH-9FD1..9FE9
U+1032F OLD ITALIC LETTER TTE
U+11CA8 MARCHEN SUBJOINED LETTER -A
U+16FE1 NUSHU ITERATION MARK

Deletion (new blocks, also removed)

11A00..11A44 Zanabazar Square (block: 11A00..11A4F)
1B100..1B28B Nushu (block: 1B100..1B28F)
2CEB0..2DDBB CJK Unified Ideographs Extension F (block: 2CEB0..2DDBF)

Move (new block, stays with updated value)

Move Tangut Radicals (renamed Tangut Components) from 18900..18BF0 to 18800..18AF0
The block is now 18800..18AFF.

There also name changes concerning 3 characters in the new Glagolitic Supplement block, and all Tangut Radicals.

Addition concerning Emoji characters arranged per block (existing or new)

The rationale for inclusion for the following characters are provided in links included in document WG2 N4654. These characters are currently under review by the Unicode Technical Committee. The characters are arranged per block (existing or new).

1. Miscellaneous Symbols and Pictographs (1F300..1F5FF)

U+1F32D HOT DOG
U+1F32E TACO
U+1F32F BURRITO
U+1F37E BOTTLE WITH POPPING CORK
U+1F37F POPCORN
U+1F3CF CRICKET BAT AND BALL
U+1F3D0 VOLLEYBALL
U+1F3D1 FIELD HOCKEY STICK AND BALL
U+1F3D2 ICE HOCKEY STICK AND PUCK
U+1F3D3 TABLE TENNIS PADDLE AND BALL
U+1F3F8 BADMINTON RACQUET AND BIRDIE
U+1F3F9 BOW AND ARROW
U+1F3FA AMPHORA
U+1F4FF PRAYER BEADS
U+1F54B KAABA
U+1F54C MOSQUE
U+1F54D SYNAGOGUE
U+1F54E MENORAH WITH NINE BRANCHES

2. Emoticons (1F600..1F64F)

U+1F643 UPSIDE-DOWN FACE
U+1F644 FACE WITH ROLLING EYES

3. Transport and Map Symbols (1F680..1F6FF)

U+1F6D0 PLACE OF WORSHIP,

4. Supplemental Symbols and Pictographs (**new block at 1F900..1F9FF**)

U+1F900 DHYANI BUDDHA

U+1F910 ZIPPER-MOUTH FACE

U+1F911 MONEY-MOUTH FACE

U+1F912 FACE WITH THERMOMETER

U+1F913 NERD FACE

U+1F914 THINKING FACE

U+1F915 FACE WITH HEAD-BANDAGE

U+1F916 ROBOT FACE

U+1F917 HUGGING FACE

U+1F918 SIGN OF THE HORNS

U+1F980 CRAB

U+1F981 LION FACE

U+1F982 SCORPION

U+1F983 TURKEY

U+1F984 UNICORN FACE

U+1F9C0 CHEESE WEDGE

Miscellaneous addition

The following character completes a range that is currently part of the amendment and have not received any negative comment. It was postponed for further study and has now been confirmed.

Block: Arabic Extended-A (08A0..08FF)

U+08B8 ARABIC LETTER TEH WITH SMALL TEH ABOVE

Dispositions per country follows:

China: Positive with comment

Technical comment

T1. CJK United Ideograph Extension F Code Charts

China is in favor of SC2N4379 ISO/IEC 10646:2014/PDAM 2.2 with comments.

China requires to remove 60 characters (61 sources) from CJK United Ideographs Extension F Code Charts, as recorded in IRGN2042 during IRG#43.

Proposed change by China

The following 11 characters in CJK_F are required to be removed as recorded in IRGN2042 during IRG#43.

Z_SAT01338 (CJKF:SN02537)  Z_SAT01338	KA-KC01963 (CJKF:03844)  03844
Z_SAT01096 (CJKF:03431)  Z_SAT01096	Z_SAT01722 (CJKF: 2D127)  00791
Z_SAT03456(CJKF:02198)  Z_SAT03456	Z_SAT04653 (CJKF: 2D834)  02976
Z_SAT04922(CJKF:02438)  Z_SAT04922	G_CY0697 (CJKF: 2D3AD)  G_CY0697
KA-KC01326(CJKF:02719)  KA-KC01326	
JMJ-059937(CJKF:04997)  JMJ-059937	
JMJ-058841(CJKF:07245)  JMJ-058841	

Besides, the following 50 (actual 49) characters listed in IRGN2041 are required to be removed. Note that one character (Z_SAT03090) is sharing U+2DB20 with a G character (G_Z4851402), G glyph and G source should be kept in CJK_F. This is also recorded in IRGN2042 during IRG#43.

Z_SAT00055	Z_SAT01869	Z_SAT03431	Z_SAT04809
Z_SAT00268	Z_SAT02066	Z_SAT03483	Z_SAT04810
Z_SAT00307	Z_SAT02114	Z_SAT03564	Z_SAT05302
Z_SAT00313	Z_SAT02160	Z_SAT03777	Z_SAT05701
Z_SAT00332	Z_SAT02396	Z_SAT03807	Z_SAT05796
Z_SAT00856	Z_SAT02486	Z_SAT03840	Z_SAT60012
Z_SAT00966	Z_SAT02590	Z_SAT03855	Z_SAT60046
Z_SAT00998	Z_SAT02802	Z_SAT03899	Z_SAT60095
Z_SAT01366	Z_SAT02889	Z_SAT03966	Z_SAT60177
Z_SAT01373	Z_SAT03076	Z_SAT04194	Z_SAT90141
Z_SAT01556	Z_SAT03090	Z_SAT04335	Z_SAT60123
Z_SAT01739	Z_SAT03269	Z_SAT04345	
Z_SAT01778	Z_SAT03388	Z_SAT04803	

Partially accepted

See also comment T16, T17, T18, and G19 from U.K, TE.2 from USA and their disposition. The ideograph corresponding to Z_SAT00856 is preserved with a new source UTC01155, thus resulting in 59 removals instead of 60. Note that per disposition of the comments T18 and G19 from UK, Extension F will be removed from this amendment.

Germany: Negative

Technical comments

T1. Page 151 ff. Clause 31 – Miscellaneous Symbols and Pictographs

T1: PDAM2.2 contains the following five characters in the block “Miscellaneous Symbols and Pictographs” (see p. 151, 157):

1F3FD1[sic] EMOJI MODIFIER FITZPATRICK TYPE-1-2

1F3FD 2[sic] EMOJI MODIFIER FITZPATRICK TYPE-3

1F3FD 3[sic] EMOJI MODIFIER FITZPATRICK TYPE-4

1F3FE 4[sic] EMOJI MODIFIER FITZPATRICK TYPE-5

1F3FF 5[sic] EMOJI MODIFIER FITZPATRICK TYPE-6

However, these characters are neither symbols nor pictographs, as their main purpose is not to depict squares framed with a zigzag or wavy outline and different hatches (or colors represented by them) inside. As their name indicates, their main purpose is to influence the color by which preceding (i.e. preceding in the input order) symbols should be represented. Thus, they act like variation selectors, except that they have a visible representation when in fact not being applied to the preceding symbol. By this, they constitute a character class of its own, related to variation selectors and combining characters. Putting such special characters in the midst of a large block of characters which otherwise require no special treatment by display engines is considered a design fault which invalidates the purpose of grouping related characters into blocks.

Proposed change by Germany

Germany requests the emoji modifiers to be grouped into a block of their own.

This block may be named “Modifier symbols” or otherwise, and may be placed at 1F9E0...1F9FF or elsewhere.

Not accepted

See also comment T3 from Ireland and T15 from UK.

The proposed code values in Amendment for these are 1F3B..1F3FF, not the values shown above.

The intent of these characters, as clearly indicated by the group header in the chart: Emoji Modifiers, is to only apply to Emoji characters, not be used as a general purpose color modifying scheme for arbitrary characters. Being in the same block as a large number of these emoji characters is practical.

It is quite common to have in the same block characters with very different features and that require very different display treatment (such as format characters in many script blocks). Another example is the Enclosed Alphanumeric Supplement block (1F100-1F1FF) which contains in addition to these enclosed characters, Regional Indicators which also have very a different display model than other characters in the same block.

There is no design or principle in ISO/IEC 10646 which requires characters with different display or processing need to be separated in different blocks. In this case, having the Emoji modifiers located near a large portion of the characters they ‘modify’ is convenient.

T2. Page 151 ff. Clause 31 – Miscellaneous Symbols and Pictographs

The five Emoji modifiers listed in German comment T1 are proposed in PDAM 2.2 with their code points in ascending order according to the Fitzpatrick numbers. While the actual code points have no technical relevance, this decision was disputed based on political arguments, as the order was interpreted as being charged with meaning by some non-professionals.

Proposed change by Germany

Germany suggests the five emoji modifiers to be assigned with code points in descending order relative to the Fitzpatrick numbers in their names. Thus, anybody may list them either by the code point or the Fitzpatrick order, thus being able to justify their preferred order on technical arguments in either case.

ISO and UTC gain a strong argument for non-technical discussions by implementing a technically equivalent solution.

Not accepted

See also comment T3 from Ireland and T15 from UK.

The same way that digits are typically encoded in their ascending order, or alphabets in their customary ascending order, there is no reason to introduce a new ordering precedent to give way to political or non-technical arguments (using terms used above by the German NB) that are out scope for this work.

Ireland: Negative

Ireland **disapproves** the draft with the technical and editorial comments given below. Acceptance of these comments and appropriate changes to the text will change our vote to approval.

Technical comments

T1. Page 34, Row 1030: Old Italic

The Irish NB continues to be of the opinion that the addition of the character 1032F OLD ITALIC LETTER TTE implies an implicit unification of Old Italic and North Italic scripts whose technical merit has not yet been agreed. We agree that a TTE should be encoded for writing North Italic, but do not believe that the ramifications of unification with Old Italic have been agreed by all of the stakeholders. This issue is similar to that of Phoenician and Hebrew. Ireland requests that the encoding of 1032F OLD ITALIC LETTER TTE be delayed for further study.

Accepted

The proposed character will be moved to the next amendment or committee draft.

T2. Page 48, Row 1700: Tangut

With reference to ISO/IEC JTC1/SC2/WG2 N4650 “Discussion of Tangut character L2008-4148”, Ireland requests that the glyph change and reordering of the character recommended in that document be implemented in the next version of the amendment.

Accepted

See also comments T11 from UK.

Note that the comment T11 from UK corrects some spelling errors in the reordering requested by WG2 N4650. It should read: move 176CE to 176A1 and reorder 176A1..176CD to 176A2..176CE.

T3. Page 48, Row 1F20: Miscellaneous Symbols and Pictographs

The Irish NB considers that the names and glyphs for the characters at 1F3FB..1F3FF are optimal and does not favour any change to the amendment with regard to these characters.

Noted

See also comments T1 and T2 from Germany and T15 from UK.

Editorial comments

E.1 Page 19, Row 0D0: Malayalam.

Ireland requests that the glyph for the characters at 0D78 be displayed so it does not crash into its box. In addition, a check should be made to characters on the ballot to ensure they have the same style as the rest of the font. 0D76 for instance should likely have the same right vertical as 0D2E.

Accepted in principle

The first request is accommodated by a small change in chart production.

Concerning the second request (character checks), note that a single font contains all characters for the block, not just the characters added by this amendment. Therefore, variations between characters such as 0D78 and 0D2E are probably intended.

E2. Page 52, Row 104B: Osage.

Ireland recommends the use of a slab-serif Deja Vu style font rather than the Times-style font used in PDAM 2.2. In addition, Ireland notes that the horizontal bar in 104B6 and 104DE should be through the lower right of the character:



Accepted

A new font will be used for pdam2.3..

Japan: Negative

Technical/Editorial comments (T or E prefix):

T1. Page 31 Sub-clause 23.1 List of source references

It is proposed that “Replace all J1 and JA sources that are also included in JIS X 0213:2004 by the values defined in JIS X 0213:2004.”

a) Replacement of CJK source references is the significant technical change. Japan strongly objects the change of source references for J1 and JA sources’ characters. It causes the confusion to the existing users who need J1 and JA source information and who believe the compatibility is preserved on this standard.

Proposed change by Japan

Do not change J1 and JA sources’ references.

Remove NOTE 3 and NOTE 4.

Not accepted

The document WG2 N4620 (CJK Ideographs glyphs representation and sources references) goes in detail on why the change was proposed. In short, it has to do with using the more modern version of these references. The compatibility is preserved because the new sources are just updates to the original values (albeit in another JIS standard) and do not change the formal identity of these characters. It also clarifies the connection between a modern standard such as JIS X 0213:2004 and ISO/IEC 10646. The source values corresponding to J1 (JIS X 021-1990) and JA (Unified Japanese IT Vendors Contemporary Ideographs, 1993) are still available in the main source reference file (CJKSrc.txt) if they have not been superseded by JIS X 0213:2004. Furthermore, many J source glyphs as described until now in ISO/IEC 10646 have been obsoleted by JIS X 0213:2004, resulting in an imperfect view of characters used in Japan for these sources. Many of these glyph changes while not important enough to put in doubt their unification status are nevertheless significant.

It should also be noted that while the information concerning these superseded J1 and JA sources has been removed from the normative source reference file (CJKSrc.txt) by this proposed amendment, the information is still available through the existing collection 372 JAPANESE IDEOGRAPHICS SUPPLEMENT (J1 sources) and the newly added collection 373 JAPANESE IT VENDORS CONTEMPORARY IDEOGRAPHICS-1993 (JA sources).

b) See other reason as follows.

JIS X 0213:2014 is not included in the list of source standards on which the source separation rule is applied, as described in S.1.6. Therefore, if the source references for J1 are replaced with JIS X 0213:2014 references, some of existing source separation examples in S.3 are not valid anymore. This also affects the procedure to develop CJK Unified ideograph in the future.

Accepted in principle

This is a valid concern, however while the source references for J1 are replaced in the formal data file describing the source references, their original source standards are still correctly represented by these characters and again the information is still available through the collection 372. Therefore the introductory sentence of Annex S.1.6 still stands:

To preserve data integrity through multiple stages of code conversion (commonly known as “round-trip integrity”), any ideographs that are separately encoded in any one of the source standards listed below have not been unified.

Concerning the examples in S.3, the source change only affects 7 pairs out of the 77 (approx.) entries which have J values, with a former J1 reference now superseded by J3 or J3A. However, because these characters have still J1 source (even if not documented in CJKSrc.txt), the examples in S.3 are still valid. And the procedure to develop new CJK Unified Ideograph is not affected.

For clarification, a new note can be inserted after NOTE 1 in S.1.6 with the following text:

NOTE 2 – The characters from the J source: JIS X 0212-1990 encoded in this International Standard are listed in the collection 372 JAPANESE IDEOGRAPHICS SUPPLEMENT.

c) If there is the need to add JIS X0213:2004 source information, the addition to existing source information, not the replacement, could be alternative. For example, add JIS X0213:2004 source information in CJKSrc.txt as follows.

U+4E02 kIRG_GSource G5-3021

U+4E02 kIRG_JSource J1-3021

U+4E02 kIRG_JSource J4-2122

U+4E02 kIRG_TSource T4-2126

U+4E02 kRSUnicode 1.1.

Not accepted

It is not possible to have multiple source references from the same tag value (such a kIRG_JSource) for a single code point. However, it is possible in the future to add a source field in the collection descriptions for collection 372 and 373 which would then capture the kIRG_JSource former value (instead of having a hex value per line such as <4E02>, we could have <4E02, TAB, J1-3021>).

If the mitigations are not considered sufficient by Japan, then The Japanese NB or other experts could propose other means to capture historical details of source references when national standards are superseded by other national standards, for possible incorporation into a future edition/amendment of 10646.

E2. Page 32 Sub-clause 23.2 Source references file for CJK Ideographs – Table 5

There are some errors in the following field format values for tag values “kIRG_GSource” which will be newly inserted.

GFC-dddddd, GGFZ-dddddd, GPLG-dddd, GXHZ-ddd, GXHZ-ddd, GZ-dddddd, and GZYS-ddddd.

Proposed change by Japan

Change to

GFC-ddd, GGFZ-dddddd, GPLG-dddd, GXHZ-ddd, GZ-dddddd, and GZYS-ddddd.

Accepted

See also comment E1 from UK.

In other words, ‘GFC-dddddd’ is replaced by ‘GFC-ddd’, and the duplicated ‘GCXHZ-ddd’ entry is removed.

E3. Page 34 Sub-clause 23.2 Source references file for CJK Ideographs – CJKSrc.txt file

(This is related comment to above.)

The G source reference for “GFC” should be in the format in “GFC-ddd”. However, the following source references in CJKSrc.txt file are wrong.

U+9FCE kIRG_GSource GFC-002200

U+9FCF kIRG_GSource GFC-005900

Proposed change by Japan

Change to

U+9FCE kIRG_GSource GFC-022

U+9FCF kIRG_GSource GFC-059.

Accepted

Some of the references were using the format ‘GFC-dddddd’ that is modified by accepting comment E2 above.

E4. Page 47 Clause 31 Code charts and lists of character names – CJK Unified Ideographs block

(This is related comment to above.)

With the same reason above, the source references of U+9FCE and U+9FCF in the code charts are wrong.

Proposed change by Japan

Change the source reference of U+9FCE

from GFC-002200 to GFC-022
and the source reference of U+9FCF
from GFC-005900 to GFC-059.

Accepted

Some of the references were using the format ‘GFC-ddddd’ that is modified by accepting comment E2 above.

T5. Page 47 Clause 31 Code charts and lists of character names – CJK Unified Ideographs Extension F block

Code chart of CJK Unified Ideographs Extension F is newly added. However, it seems this code chart is not the correct version (WG2 N4580) that IRG submitted to WG2 for inclusion. For example, the glyph shape of U+2D0E9 is different.

Proposed change by Japan

Please develop the code charts based on WG2 N4580.

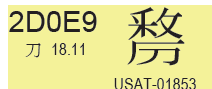
Accepted in principle

WG2 N4580 is a PDF file, it does not contain the information to create code charts. IRG members have provided fonts to the project editor that are supposed to follow WG2 N4580. If there are meaningful differences, it is up to the National Bodies to review and provide detailed feedback. For the case of U+2D0E9, the document WG2 N4580 shows:



Z_SAT01853

while Amendment 2 charts shows:



This seems to be a stylistic difference, however the project editor will gladly welcome an updated font from any IRG source which should be improved.

Note that per disposition of UK comment T18 and G19, CJK Ext F has been removed from Amendment 2.

T6. Page 47 Clause 31 Code charts and lists of character names – Zanabazar Square block

As described in WG2 N4653, Mongolian experts expressed their concern about the quality of the proposed character set. Japan suggests to form the consensus with Mongol national body before encoding this script.

Proposed change by Japan

Remove Zanabazar Square block.

Accepted

See comment T1 from Mongolia.

T7. Page 47 Clause 31 Code charts and lists of character names – Nushu block

The Nushu block contains the problematic characters listed in WG2 N4610. Japan thinks those character should be postponed.

Proposed change by Japan

Replace Nushu block with the chart proposed in WG2 N4652.

Partially accepted

See comment T14 from U.K and TE1 from USA and their disposition.

The UK provides rationale for keeping the characters with some reordering and glyph changes. Given the complexity of the issues at hand, it is preferable to postpone Nushu to the next amendment/committee draft. The repertoire as amended by UK comment T14 will be part of that document.

T8. Page 47 Clause 31 Code charts and lists of character names – Emoji modifiers in Miscellaneous Symbols and Pictographs block

The characters of EMOJI MODIFIER are currently in public review by UTC. They should be proposed at least after the consensus is made through the public review.

Proposed change by Japan

Remove EMOJI MODIFIER from U+1F3FB “EMOJI MODIFIER FITZPATRICK TYPE-1-2” to U+1F3FF “EMOJI MODIFIER FITZPATRICK TYPE-6”.

Not accepted

Typically UTC and ISO conducts repertoire review in parallel. In order to maintain synchronicity between the repertoire of Unicode and ISO/IEC 106464, it is important to keep the repertoire under ballot so that all interested parties can provide feedback as soon as possible.

E9. Page 2396 Annex A.5 Other collections – MOJI-JOHO-KIBAN IDEOGRAPHS-2014

It is found the error in “JMJKI-2014” file for new extended collection named MOJI-JOHO-KIBAN IDEOGRAPHS-2014. Could you please correct it ?

Proposed change by Japan

Change the line number 471

3626,<5207,E0103>

to

3626

And change the line number 6788

5207,<5207,E0101>,<5207,E0102>

to

5207,<5207,E0101>,<5207,E0102>,<5207,E0103>.

Accepted in principle

See US comment TE6. As a result the collection is removed from the amendment.

Mongolia: Negative

General comment:

We hope that encoding and international standardization of Soyombo and Zanabazar Square Script as unique heritage of Mongolian culture, should be considered and studied according to its traditional writing order, traits, dimensions and manuscript graphics of it and based on scientific research. Therefore, we would sincerely request you to consider our proposal of postponing of Soyombo and Zanabazar Square Scripts encoding standardization until necessary research is conducted. Thank you for your kind cooperation.

Yours sincerely,

Mr. Otgonbaatar.R (Research worker in The Institute of Language and Literature at Mongolian Academy of Sciences)

Mr. Demberel.S (Lecturer in the Department of Philosophy and Religious studies, NUM)

Mrs. Undraa.B (Secretary of ICT TC, senior officer of MASM)

Accepted

See also comment T6 from Japan.

The repertoire will be moved to the next amendment/Committee draft to allow further research and study of the repertoire.

There is clearly a need to communicate between Mongolian experts and the submitter of the current proposal.

What should be made clear is that ISO/IEC 10646 and Unicode are not transcription/transliteration standards.

While transcription in other writing systems is a useful mechanism to determine the appropriate encoding model and possibly additional code points to represent additional elements, it is only a partial view of the problem and solution.

Technical comments

(Note that references for sections and pages are related to the document WG2 N4541, not the amendment which contains only a subset of elements described in N4541, so many aspects of these comments are out of scope for the disposition of comments.)

T1. Zanabazar Square

a) Section 4.7 (4.7.1 and 4.7.4), page 8, line 8-17, 20-22

There is some dimensional faulty of Soyombo and Zanabazar Square script found in Anshuman Pandey's project.

The manuscript graphics of Soyombo and Zanabazar Square scripts totally wrong in this project.

Proposed change by Mongolia

We should consider, conduct scientific research on traditional writing order, traits, dimensions, manuscript graphics, proportion of Zanabazar Square script, when encoding it.

Noted

The glyphs for the proposed repertoire shown in page 8 are called 'manuscript graphics' by the reviewers. The commenters are encouraged to provide specifics to the author (Anshuman Pandey) in order to create code charts and representative glyphs that satisfy their requests. Note that glyphs are informative and can be modified in future iterations of the standard.

b) Section 4.7.3, 4.7.4 Page 10, line 3-19

According to our observation Dr Anshuman Pandey mainly compared Zanabazar square script with Sanskrit and Tibetan vowels and consonants, but he didn't emphasize phonetics, word-lore and Mongolian grammar in his research.

Proposed change by Mongolia

He should consider how to transcribe consonants (г, ж/ з/, ч/ц) in Mongolian alphabet by soyombo and Zanabazar Square script.

Noted

What is apparently missing in that table is the transcription of Mongolian written in Cyrillic.

c) Section 4.7.3, 4.7.4 Page 10, line 3-19

Zanabazar square script being created for writing down literary works of Sanskrit, Tibetan and Mongolian in perfect way, hence should allocate/equivalent/ phonetics and word-lore of those languages.

Proposed change by Mongolia

- If Mongolian words are written by Zanabazar Square script, there should be the right transcription of Mongolian script adhered
- If Sanskrit words are written by Zanabazar square script, there should be right transcription of Sanskrit script adhered
- If Tibetan words are written by Square script, there should be the right transcription of Tibetan script adhered. /This issues should be also considered in Zanabazar Soyombo script/.

Noted

Transcription is not part of ISO/IEC 10646. See disposition of General comment above.

d) Section 4.8.3, Page 12, line 30-36

There is only Mongolian word example written by the horizontal and vertical by Zanabazar Square script./not enough study did on Sanskrit and Tibetan words/

There should be more study on Sanskrit and Tibetan word written by the horizontal and vertical by Zanabazar's soyombo and Square scripts.

Proposed change by Mongolia

Research should be conducted on the horizontal and vertical writing methods of Zanabazar Square script, when Mongolian words written by it. There is another incident should be considered that Zanabazar Square script written by either head letter or without head letter, when transcript Sanskrit tantric manuscripts by it. We would like to seek for advice and comments on encoding Zanabazar square script from local experts and international scholars.

Noted

However vertical text handling is not a character encoding matter, but is done at a higher presentation level.

UK: Negative

General/Technical/Editorial comment (G, T, or E prefix):

E1. Clause 23 – Source reference presentation for CJK Unified Ideographs

The list of kIRG_GSource format values contains a duplicate value:

"GFC-dddddd, GGFZ-dddddd, GPGLG-dddd, GXHZ-ddd, GXHZ-ddd, GZ-dddddd, and GZYS-ddddd."

Proposed change by UK:

Remove duplicate "GXHZ-ddd".

Accepted

See also comment E2 from Japan.

E2. Sub-clause 24.2 – Source reference file for Tangut Ideographs

"The three fields are delimited by a LINE TABULATION control character (000B)."

The delimiter is actually a CHARACTER TABULATION control character (0009).

Proposed change by UK:

Change to "The three fields are delimited by a CHARACTER TABULATION control character (0009)."

Accepted

Note that the sub-clause 23.2 needs a similar fix and other occurrences in the text of the standard where the term 'TAB character' is used should be modified for consistency.

E3. Sub-clause 24.2 – Source reference file for Tangut Ideographs

"The format definition uses 'd' as a decimal unit, 'x' as an alphanumerical unit (0 to 9 and A to Z), t and <space> as the SPACE character."

There is a spurious 't' after the comma.

The space character is not used in the definition.

Proposed change by UK:

Remove 't' after the comma.

Remove " and <space> as the SPACE character".

Accepted

E4. Annex A.1 Collections – Tangut Radicals

"1101 TANGUT RADICAL 18900-18BFF"

"TANGUT RADICAL" should be "TANGUT RADICALS".

Proposed change by UK:

Change to "1101 TANGUT RADICALS 18900-18BFF".

Accepted in principle

See also comment TE3 from US and its disposition.

The new name is TANGUT COMPONENTS.

E5. Annex A.6 – Unicode collections

UNICODE 8.0 collection Plane 00 Row 20 has the following range of characters:

"00-64 66-71 74-8E 90-9C A0-BD D0-F0"

This omits 20BE LARI SIGN.

Proposed change by UK:

Change range for Plane 00 Row 20 to "00-64 66-71 74-8E 90-9C A0-BE D0-F0".

Also add Plane 00 Row 20 Value BE to the Unicode 8.0 ranges given in the following note.

Accepted

Note that the Unicode 8.0 repertoire is not yet stable and will probably change in the next version of this amendment.

E6. Annex G – Alphabetically sorted list of character names

"Page 2419, Annex G"

Annex G is on page 2421 of ISO/IEC 10646:2014.

Proposed change by UK:

Change to "Page 2421, Annex G".

Accepted

E7. Annex I – Ideographic description characters

I.2 specifies that a Description Component (DC) may be a coded ideograph or a coded radical or FF1F or a PUA character or another IDS.

It does not specify what ranges of characters a coded ideograph or a coded radical encompasses.

As this amendment adds Tangut ideographs and Tangut radicals the text should clarify what a coded ideograph and a coded radical are.

Proposed change by UK:

Suggest adding a note for the first paragraph that an ideographic character may be a CJK ideograph or a Tangut ideograph.

Suggest adding a note that a coded ideograph may be any coded character in the following blocks:

- CJK COMPATIBILITY IDEOGRAPHS
- CJK COMPATIBILITY IDEOGRAPHS SUPPLEMENT
- CJK UNIFIED IDEOGRAPHS
- CJK UNIFIED IDEOGRAPHS EXTENSION A
- CJK UNIFIED IDEOGRAPHS EXTENSION B
- CJK UNIFIED IDEOGRAPHS EXTENSION C
- CJK UNIFIED IDEOGRAPHS EXTENSION D
- CJK UNIFIED IDEOGRAPHS EXTENSION E
- CJK UNIFIED IDEOGRAPHS EXTENSION F
- TANGUT

Suggest adding a note that a coded radical may be any coded character in the following blocks:

- KANGXI RADICALS
- CJK RADICALS SUPPLEMENT
- TANGUT RADICALS

Accepted in principle

See also comment TE3 from US and its disposition.

TANGUT RADICALS is changed to TANGUT COMPONENTS.

T8. Annex I – Ideographic description characters

I.2 specifies that a Description Component (DC) may be a coded ideograph or a coded radical or FF1F or a PUA character or another IDS.

It would be useful to add stroke characters to the list of allowable description components, as IDS sequences for CJK ideographs often need to include CJK strokes. For example, 3514 𠂇 may be described as 𠂇加○, using 31E3 CJK STROKE Q as there is no corresponding coded ideograph for this stroke.

Proposed change by UK:

Add the following bullet point to the list of Description Components, following "a coded radical":

- a coded stroke

Add a note that a coded stroke may be any coded character in the following blocks:

- CJK STROKES.

Accepted

T9. Clause 31 (32) CJK Unified Ideographs

9FD1 through 9FE9 do not need to be encoded in the BMP (or fast-tracked to Unicode 8.0) as they are low frequency characters that are neither included in published national character lists, nor required in order to correct a unification error. As this amendment now incorporates CJK UNIFIED IDEOGRAPHS EXTENSION F, it would be appropriate to move these characters into the CJK UNIFIED IDEOGRAPHS EXTENSION F block, and leave the space at the end of the CJK UNIFIED IDEOGRAPHS block free for more urgent characters.

Moving these characters to the CJK UNIFIED IDEOGRAPHS EXTENSION F block will satisfy the issues relating to 9FD8 through 9FE9 raised in GB comments to PDAM2.1.

Proposed change by UK:

Move 9FD1 through 9FE9 as follows:

9FD1: Insert after 2D070

9FD2: Insert after 2DD4E

9FD3: Insert after 2DD55

9FD4: Insert after 2DD98

9FD5: Insert after 2DDB1

9FD6: Insert after 2CECB

9FD7: Insert after 2CED5

9FD8: Insert after 2CFBD

9FD9: Insert after 2CFCE

9FDA: Insert after 2CFEE

9FDB: Insert after 2D08F

9FDC: Insert after 2D845

9FDD: Insert after 2DA75

9FDE: Insert after 2DAD0

9FDF: Insert after 2DD18

9FE0: Insert after 2DD1A

9FE1: Insert after 2DD39

9FE2: Insert after 2DD3F

9FE3: Insert after 2DD3F

9FE4: Insert after 2DD3F

9FE5: Insert after 2DD4D

9FE6: Insert after 2DD63

9FE7: Insert after 2DD9D

9FE8: Insert after 2DDA2

9FE9: Insert after 2DDB1

Accepted in principle

The characters in this list came in two batches:

1. *UTC L2/12-333 Request to UTC to Propose 226 Characters for inclusion in CJK Extension F, link: <http://www.unicode.org/L2/L2012/12333-cjk-f.pdf>*
2. *UTC L2/13-009 Proposal to Encode Chinese Characters Used for Transliterating Slavonic, link: <http://www.unicode.org/L2/L2013/13009-slavonic-chinese.pdf>*

Inserting characters into an existing block is expensive from a chart production point of view. References tables have to be regenerated, fonts or mapping to fonts have to be redone.

Another consideration is the additional comments from UK concerning Extension F (T16 to T19) which request significant reviews and re-arrangement. Based on this, it seems prudent to postpone CJK Extension F to a future amendment or committee draft.

T10. Clause 31 (32) – Marchen

The glyph shapes of 11CA8 MARCHEN SUBJOINED LETTER -A and 11CB0 MARCHEN VOWEL SIGN AA are identical, which is problematic, as it may result in multiple spellings for the same word.

A full set of subjoined Marchen letters is proposed for encoding following the Tibetan encoding model, but MARCHEN SUBJOINED LETTER -A is not attested in Marchen texts, and the corresponding Tibetan letter (0FB0 TIBETAN SUBJOINED LETTER -A) is exceedingly rare in Tibetan text. As it is unlikely that Subjoined Letter -A is required for Marchen usage, the best solution may be to remove MARCHEN SUBJOINED LETTER -A and leave 11CA8 reserved. If evidence for a distinct Marchen form of Subjoined Letter -A (visibly different to MARCHEN VOWEL SIGN AA) is forthcoming in the future then it can be added at that time.

Proposed change by UK:

Remove 11CA8 MARCHEN SUBJOINED LETTER -A, and leave the code point reserved.

Accepted

T11. Clause 31 (32) – Tangut

The glyph for 176CE (L2008-4148) has an extra stroke which is a mistake in the source (Li Fanwen 2008). Examination of primary sources shows that this character should be 11 strokes rather than 12 strokes.

Evidence for the glyph correction is provided in WG2 N4650.

Proposed change by UK:

Correct the glyph for 176CE as recommended in N4650.

Move 176CE to 176A1, and reorder 176A1..176CD to 176A2..176CE.

Correct the kRSTUnicode value for L2008-4148 to "104.11".

Accepted

See also comment T2 from Ireland.

T12. Clause 31 (32) – Tangut Radicals

There is an empty row between the Tangut and Tangut Radicals blocks (18800..188FF). As it is not known whether a Tangut Extension block will be required in the future, or how large such a block would need to be, it is not useful to reserve this row for future standardization of the Tangut script. The best solution would be to move the Tangut Radicals block up one row to fill this gap.

Proposed change by UK:

Reorder 18900.. 18BF0 to 18800.. 18AF0.

Accepted in principle

See also comments TE3 from US and its disposition.

The block name is changed to Tangut Components.

T13. Clause 31 (32) – Tangut Radicals

The ranges of Tangut ideographs listed for Tangut Radical-017 onwards are out by one. This is because the addition of the new Tangut ideograph at 17132 (under Radical 17) has not been taken into account.

Proposed change by UK:

Recalculate the ranges of Tangut ideographs listed for Tangut radicals. We are willing to provide the project editor with a corrected file.

Accepted in principle

See also US comment TE5 and its disposition.

As a result these ranges are not shown anymore.

T14. Clause 31 (32) – Nushu

We have carried out a review of the Nushu repertoire, with particular reference to the comments in N4610. In the comments below, NSDB = *Nüshu Duben*; NSYZBJ = *Nüshu Yongzi Bijiao* (Beijing 2006); Chiang = William W. Chiang, *We Two Know the Script; We Have Become Good Friends* (1995).

General points

A. The preferred glyph form for a given Nüshu character may vary from one Nüshu user to another, and there is no good reason to prioritise the preferences of one user over another user. Therefore, we do not consider that it is acceptable to change the glyph shape of a proposed character simply because the preferred glyph shape for one or more informants for N4610 differs from the representative glyph shape in the Chinese proposal. In cases where there is disagreement over the preferred glyph, we consider that an acceptable solution would be to keep the proposed glyph and define variation sequences for alternative glyph forms.

B. The inability of one or more informants for N4610 to recognise a particular character does not indicate that it is not required for encoding. There is no reason to remove any Nüshu character from this amendment if it is attested in the source NSDB.

C. Where a character unifies several variants with different stroke counts, NSDB orders the character by the stroke count of one particular variant. However, in some cases the glyph used in the code chart is a different variant with a different stroke count, with the result that the character is ordered under the wrong stroke count in the code chart. We have noted all such examples that we have found during our review in "Additional Comments".

D. In some cases the code chart glyph is not ideal, and we have had to refer to NSDB and/or NSYZBJ in order to determine how the character should be correctly drawn (this is case for 1B1DA and 1B1F2 for example). We also note that many characters with a diagonal box are drawn with a gap between the strokes at the top (e.g. 1B159), when no such gap is shown for these characters in NSDB, NSYZBJ, or Chiang. We further note that the rotational angle of some characters (e.g. 1B15F and 1B1B5) differs significantly from the glyph examples given for the corresponding character in NSDB, NSYZBJ and Chiang. We have not individually indicated such minor glyph issues as we suppose that they fall within the acceptable range of glyph variation for the character, but we do think that they may cause problems for font designers.

Response to comments in N4610

1B108 : The representative glyph shape is consistent with NSDB (p. 36), NSYZBJ (p. 27), and Chiang (p. 190). No good reason to change the glyph shape.

1B117 : The character and glyph shape is attested in NSDB (p. 37), NSYZBJ (p. 141), and Chiang (p. 173). Do not remove.

1B11F : We agree that the glyph form for 1B11F in PDAM2.2 is incorrect, and should have a short vertical stroke instead of a circle, as shown in NSDB, NSYZBJ and Chiang (p. 148). On the other hand *i*⁵ (similar to the glyph for 1B11F in PDAM2.2) is already encoded at 1B14D and does not need further consideration. Correct the glyph for 1B11F to reflect the shape shown in NSDB.

1B124 : The character and glyph shape is attested in NSDB (p. 38), NSYZBJ (p. 60), and Chiang (p. 192). Keep the character and representative glyph shape, and consider adding a Variation Sequence (or sequences) for the alternative glyph(s) if necessary in the future.

1B13A : The character and glyph shape is attested in NSDB (p. 39), NSYZBJ (p. 27), and Chiang (p. 187). Keep the character and representative glyph shape, and consider adding a Variation Sequence for the alternative glyph if necessary in the future.

1B13B : The character and glyph shape is attested in NSDB (p. 39), NSYZBJ (p. 28), and Chiang (pp. 189–190). Keep the character and representative glyph shape, and consider adding a Variation Sequence for the alternative glyph if necessary in the future. However, we agree that the stroke count for the representative glyph should be 7 strokes not 5 strokes, so reorder the character under 7 strokes.

1B13F : ku²¹ is encoded at 1B115, and tç'ie⁴⁴ is encoded at 1B13F., so there is no good reason to remove either character. Variation Sequences may be added for alternative glyphs for either of these characters if necessary in the future.

1B147 : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B148 : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B149 : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B14F : The character and glyph shape is attested in NSDB (p. 40) and NSYZBJ (p. 81). Keep the character and representative glyph shape, and consider adding a Variation Sequence for the alternative glyph if necessary in the future.

1B156 : We agree that evidence for the disunification of 1B156 and 1B157 is not compelling, but we see no harm in keeping 1B157. Therefore keep both 1B156 and 1B157.

1B15E : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B15F : The character and glyph shape is attested in NSDB (p. 41), NSYZBJ (p. 120), and Chiang (p. 154). Keep the character and representative glyph shape, and consider adding a Variation Sequence for the alternative glyph if necessary in the future.

1B166 : The character and glyph shape is attested in NSDB (p. 41), NSYZBJ (p. 141), and Chiang (p. 173). Keep the character and representative glyph shape.

1B172 : The character and glyph shape is attested in NSDB (p. 42), NSYZBJ (p. 37), and Chiang (p. 215). Keep the character and representative glyph shape.

1B178 : No good reason to change the glyph to the 10-stroke form, which could be defined as a variation sequence in the future if required. However, correct the glyph to remove the extra dot in the middle (see comment on 1B178 below).

1B182 : The character and glyph shape is attested in NSDB (p. 43), NSYZBJ (p. 89), and Chiang (p. 137). Keep the character and representative glyph shape.

1B183 : The character and glyph shape is attested in NSDB (p. 43), NSYZBJ (p. 91), and Chiang (p. 145). Keep the character and representative glyph shape.

1B18A : The character and glyph shape is attested in NSDB (p. 43), NSYZBJ (p. 100), and Chiang (p. 174). Keep the character and representative glyph shape.

1B190 : The character and glyph shape is attested in NSDB (p. 43), NSYZBJ (p. 117), and Chiang (p. 192). Keep the character and representative glyph shape.

1B193 : The character and glyph shape is attested in NSDB (p. 43) and NSYZBJ (p. 122). Keep the character and representative glyph shape.

1B194 : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B19F : The character and glyph shape is attested in NSDB (p. 44), NSYZBJ (p. 148), and Chiang (p. 149). Keep the character and representative glyph shape.

1B1AC : The character and glyph shape is attested in NSDB (p. 45), NSYZBJ (p. 38), and Chiang (p. 139). Keep the character and representative glyph shape.

1B1AD : The character and glyph shape is attested in NSDB (p. 45), NSYZBJ (p. 39), and Chiang (p. 162). Keep the character and representative glyph shape.

1B1AE : The character and glyph shape is attested in NSDB (p. 45) and NSYZBJ (p. 42). Keep the character and representative glyph shape.

1B1AF : The character and glyph shape is attested in NSDB (p. 45) and NSYZBJ (p. 43). Keep the character and representative glyph shape.

1B1B1 : The character and glyph shape is attested in NSDB (p. 45) and NSYZBJ (p. 45). Keep the character and representative glyph shape.

1B1BD : The character and glyph shape is attested in NSDB (p. 45), NSYZBJ (p. 90), and Chiang (p. 139). Keep the character and representative glyph shape.

1B1BE : The character and glyph shape is attested in NSDB (p. 45) and NSYZBJ (p. 97). Keep the character and representative glyph shape.

1B1C4 : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B1CA : The character and glyph shape is attested in NSDB (p. 46) and NSYZBJ (p. 131). Keep the character and representative glyph shape.

1B1D1 : The character and glyph shape is attested in NSDB (p. 46), NSYZBJ (p. 146), and Chiang (p. 214). Keep the character and representative glyph shape.

1B1D5 : The character and glyph shape is attested in NSDB (p. 46), NSYZBJ (p. 158), and Chiang (p. 195). Keep the character and representative glyph shape.

1B1F4 : The character and glyph shape is attested in NSDB (p. 48) and NSYZBJ (p. 98). Keep the character and representative glyph shape.

1B1F8 : The character and glyph shape is attested in NSDB (p. 48) and NSYZBJ (p. 108). Keep the character and representative glyph shape.

1B1FD : The character and glyph shape is attested in NSDB (p. 48), NSYZBJ (p. 117), and Chiang (p. 139). Keep the character and representative glyph shape.

1B213 : The character and glyph shape is attested in NSDB (p. 49), NSYZBJ (p. 158), and Chiang (p. 195). Keep the character and representative glyph shape.

1B219 : The character and glyph shape is attested in NSDB (p. 50) and NSYZBJ. Keep the character and representative glyph shape.

1B21B : The character and glyph shape is attested in NSDB (p. 50), NSYZBJ (p. 36), and Chiang (p. 215). Keep the character and representative glyph shape.

1B21F : No good reason to change the glyph for this character. A Variation Sequence may be added for the alternative glyph if necessary in the future.

1B220 : The character and glyph shape is attested in NSDB (p. 50) and NSYZBJ (p. 54). Keep the character and representative glyph shape.

1B223 : The character and glyph shape is attested in NSDB (p. 50) and NSYZBJ (p. 66). Keep the character and representative glyph shape.

1B22B : The character and glyph shape is attested in NSDB (p. 50), NSYZBJ (p. 100), and Chiang (p. 175). Keep the character and representative glyph shape.

1B233 : The character and glyph shape is attested in NSDB (p. 51), NSYZBJ (p. 118), and Chiang (p. 192). Keep the character and representative glyph shape.

1B23C : The character and glyph shape is attested in NSDB (p. 51), NSYZBJ (p. 146), and Chiang (p. 214). Keep the character and representative glyph shape.

1B23E : The character and glyph shape is attested in NSDB (p. 51) and NSYZBJ (p. 87). Keep the character and representative glyph shape.

1B252 : The character and glyph shape is attested in NSDB (p. 53) and NSYZBJ (p. 25). Keep the character and representative glyph shape.

1B25D : The character and glyph shape is attested in NSDB (p. 53) and NSYZBJ (p. 25). Keep the character, but modify the representative glyph shape as discussed in comment for 1B25D below.

1B25F : The character and glyph shape is attested in NSDB (p. 53) and NSYZBJ (p. 31), but modify the representative glyph shape as discussed in comment for 1B25F below.

1B27A : The character and glyph shape is attested in NSDB (p. 55), NSYZBJ (p. 132), and Chiang (p. 140). Keep the character and representative glyph shape.

1B27D : The character and glyph shape is attested in NSDB (p. 55), NSYZBJ (p. 154), and Chiang (p. 187). Keep the character and representative glyph shape.

1B288 : The character and glyph shape is attested in NSDB (p. 56) and NSYZBJ (p. 73). Keep the character and representative glyph shape.

1B28B : The character and glyph shape is attested in NSDB (p. 56) and NSYZBJ (p. 54). Keep the character and representative glyph shape.

Additional Comments

1B11E : The character is ordered under 4 strokes, but the glyph is drawn with 5 strokes (the stroke count is ambiguous for the glyph shown in NSDB p. 37, but is clearly 5 strokes in the glyphs shown in NSYZBJ p. 26). Reorder the character under 5 strokes.

1B11F : The character is ordered under 4 strokes, consistent with NSDB (p. 37), NSYZBJ (p. 41), and Chiang (p. 148), but the glyph is drawn with 5 strokes, with a circle instead of a short vertical stroke. Correct the glyph to reflect the 4-stroke shape shown in NSDB.

1B129 : The glyph does not accurately reflect the glyph shape given in NSDB (p. 38), NSYZBJ (p. 83), and Chiang (p. 180), and is confusable with 1B11D. Correct the glyph to reflect the shape shown in NSDB and NSYZBJ.

1B13B : The character is ordered under 5 strokes, but the glyph is drawn with 7 strokes. There are two glyph forms in NSDB (p. 39), one 7 strokes and one 5 strokes. Therefore either change the representative glyph or reorder the character under 7 strokes.

1B178 : The character is ordered under 6 strokes, consistent with NSDB (p. 42) and NSYZBJ (p. 45), but the glyph is drawn with 7 strokes, with an extra dot in the bottom middle. Correct the glyph to reflect the 6-stroke shape shown in NSDB.

1B196 : The character is ordered under 6 strokes, consistent with NSDB (p. 43) and NSYZBJ (p. 127), but the glyph is drawn with 5 strokes, with a missing dot on the left. Correct the glyph to reflect the 6-stroke shape shown in NSDB.

1B19A : The character is ordered under 6 strokes, consistent with NSDB (p. 44), NSYZBJ (p. 137), and Chiang (p. 153), but the glyph is drawn with 7 strokes, with a dot on the left that should be the bottom left limb of an 'x', with the result that it is confusable with 1B1AA. Correct the glyph to reflect the 6-stroke shape shown in NSDB.

1B1B3 : The character is ordered under 7 strokes, but the glyph is drawn with 8 strokes. There are two glyph forms in NSDB (p. 45), one 7 strokes and one 8 strokes. Therefore either change the representative glyph to the 7-stroke form or reorder the character under 8 strokes.

1B1C1 : The character is ordered under 7 strokes, consistent with NSDB (p. 45) and NSYZBJ (p. 112), but the glyph is drawn with 6 strokes, with a missing diagonal stroke on the right. Correct the glyph to reflect the 7-stroke shape shown in NSDB.

1B1CB : The character is ordered under 7 strokes, but the glyph is drawn with 9 strokes. There are two glyph forms in NSDB (p. 46), one 7 strokes and one 9 strokes. Therefore either change the representative glyph to the 7-stroke form or reorder the character under 9 strokes.

1B1CD : The character is ordered under 7 strokes, but the glyph is drawn with 8 strokes. The character has 8 strokes in NSYZBJ (p. 143) and NSDB (p. 46). Therefore reorder the character under 8 strokes.

1B1D1 : The character is ordered under 7 strokes, but the glyph is drawn with 9 strokes. There are two glyph forms in NSDB (p. 46), one 7 strokes and one 9 strokes. Therefore either change the representative glyph to the 7-stroke form or reorder the character under 9 strokes.

1B1D3 : The character is ordered under 7 strokes, but the glyph is drawn with 6 strokes. The examples given in NSDB (p. 46), NSYZBJ (p. 156) and Chiang (p. 194) show either 6 strokes or 7 strokes (this is also the case for 1B1DF, 1B20E and 1B221, which share the same component). Either change the representative glyph to the 7-stroke form or reorder the character under 6 strokes. Taking into consideration 1B1DF, 1B20E and 1B221, we think it would be best to change the glyph for 1B1D3 to be 7 strokes.

1B1DF : The character is ordered under 8 strokes, and the glyph is drawn with 8 strokes, consistent with NSDB (p. 47). However, NSYZBJ (p. 48) shows 8-stroke and 9-stroke forms, and if 1B1D3 is changed to a 7-stroke form then this character should also be changed to a 9-stroke form.

1B1E6 : The character is ordered under 8 strokes, consistent with NSDB (p. 47) and NSYZBJ (p. 79), but the glyph is drawn with 6 strokes, missing a circle in the middle. Correct the glyph to reflect the 8-stroke shape shown in NSDB and NSYXBJ.

1B1EA : The character is ordered under 8 strokes, but the glyph is drawn with 9 strokes. The examples in NSDB (p. 47) and NSYZBJ (p. 63) are not clear, but appear to be 9 strokes, consistent with the glyph shown in the code chart. Therefore reorder the character under 9 strokes.

1B1F0 : The character is ordered under 8 strokes, but the glyph is drawn with 5 strokes. The examples in NSDB (p. 48) and NSYZBJ (p. 90) are 5 strokes, consistent with the glyph shown in the code chart. Therefore reorder the character under 5 strokes.

1B1FF : The character is ordered under 8 strokes, but the glyph is drawn with 9 strokes. The example in NSDB (p. 48) is 9 strokes but most examples in NSYZBJ (p. 118) are 8 strokes (without the stroke on the top left). Therefore either change the representative glyph to the 8-stroke form or reorder the character under 9 strokes.

1B20E : The character is ordered under 8 strokes, but the glyph is drawn with 9 strokes, consistent with NSDB (p. 49). However, NSYZBJ (p. 48) shows 8-stroke and 9-stroke forms, and if 1B1D3 is changed to a 7-stroke form then this character should be reordered under 9 strokes; otherwise the glyph should be changed to an 8-stroke form.

1B212 : The character is ordered under 8 strokes, consistent with NSDB (p. 49) and NSYZBJ (p. 157), but the glyph is drawn with 7 strokes, missing a stroke in the middle. Correct the glyph to reflect the 8-stroke shape shown in NSDB and NSYXBJ.

1B221 : The character is ordered under 9 strokes, but the glyph is drawn with 8 strokes. However, NSDB (p. 50) and NSYZBJ (p. 59) show 9-stroke forms. Therefore change the representative glyph to the 9-stroke form shown in NSDB and NSYZBJ.

1B22D : The character is ordered under 9 strokes, but the glyph is drawn with 8 (?) strokes. NSDB (p. 51) and NSYZBJ (p. 109) appear to show 8-stroke forms. Therefore reorder the character under 8 strokes.

1B230 : The character is ordered under 9 strokes, but the glyph is drawn with 8 strokes. The glyph seems to be somewhat different from the examples shown in NSDB (p. 51) and NSYZBJ (p. 114), and should be replaced with a glyph that reflects the examples given in NSDB. The character may also need to be reordered under 8 strokes.

1B23B : The character is ordered under 9 strokes, consistent with NSDB (p. 51) and NSYZBJ (p. 145), but the glyph is drawn with 8 strokes, missing the dot on top. Correct the glyph to reflect the 9-stroke shape shown in NSDB and NSYXBJ.

1B23D : The character is ordered under 9 strokes, consistent with NSDB (p. 51) and NSYZBJ (p. 147), but the glyph is drawn with 10 strokes, with a circle instead of a vertical stroke. Correct the glyph to reflect the 9-stroke shape shown in NSDB and NSYXBJ.

1B253 : The character is ordered under 10 strokes, consistent with NSDB (p. 53) and NSYZBJ (p. 114), but the glyph is drawn with 9 strokes. Correct the glyph to reflect the 10-stroke shape shown in NSDB and NSYXBJ.

1B25D : The character is ordered under 11 strokes, consistent with NSDB (p. 53) and NSYZBJ (p. 25), but the glyph is drawn with 10 strokes, missing a dot on the bottom right. Correct the glyph to reflect the 11-stroke shape shown in NSDB and NSYXBJ.

1B25F : The character is ordered under 11 strokes, but the glyph is drawn with 9 (?) strokes. NSDB (p. 53) and NSYZBJ (p. 31) appear to show 10-stroke forms. Therefore change the glyph to reflect the 10-stroke shape shown in NSDB and NSYZBJ, and reorder the character under 10 strokes.

1B260 : The glyph is drawn with two x-shaped components at the bottom, but the examples in NSDB and Chiang (p. 194) show two inverted v-shaped components. Although NSYZBJ (p. 36) shows both glyph forms (but v-shaped more common than x-shaped), as NSDB is the primary source for the character repertoire we would expect that the glyph would correspond to that shown in NSDB.

1B267 : The character is ordered under 11 strokes, but the glyph is drawn with 12 strokes. It is not clear from the examples in NSDB (p. 54) and NSYZBJ (p. 69) whether this character should be 11 strokes or 12 strokes. Check the correct glyph form, and either change the glyph to 11 strokes or reorder the character under 12 strokes as appropriate.

1B26A : The character is ordered under 11 strokes, consistent with NSDB (p. 54) and NSYZBJ (p. 85), but the glyph is drawn with 10 strokes, missing a stroke in the left side component. Correct the glyph to reflect the 11-stroke shape shown in NSDB and NSYXBJ.

1B276 : The character is ordered under 12 strokes, but the glyph is drawn with 11 strokes. NSDB (p. 55) and NSYZBJ (p. 67) appear to show 11-stroke forms, consistent with the glyph shown in the code chart. Therefore reorder the character under 11 strokes.

1B279 : The character is ordered under 12 strokes, but the glyph is drawn with 11 strokes. NSDB (p. 55) and NSYZBJ (p. 122) show somewhat different 11-stroke glyphs (the left component may be 1B1E9 in some examples). Check the correct glyph shape, and reorder the character under 11 strokes.

Proposed change by UK:

1. Do not remove any individual characters from the Nushu repertoire without consensus from interested experts.

2. Correct the glyph and/or reorder the following characters, as discussed in "Additional Comments":

1B11E	1B178	1B1C1	1B1D3	1B1F0	1B221	1B23D	1B260
1B11F	1B196	1B1CB	1B1DF	1B1FF	1B22D	1B253	1B267
1B129	1B19A	1B1CD	1B1E6	1B20E	1B230	1B25D	1B26A
1B13B	1B1B3	1B1D1	1B1EA	1B212	1B23B	1B25F	1B276

1B279

3. If it is determined that a face-to-face meeting is required to resolve the Japanese and GB comments we will not object to the entire Nushu repertoire being removed from Amd. 2.

Accepted in principle

See also comments TE1 from US and its disposition.

The Nushu repertoire is removed from Amendment 2 and will be part of the next amendment or committee draft.

T15. Clause 31 (32) – Miscellaneous Symbols and Pictographs

We note the addition of five Emoji modifier characters at 1F3FB..1F3FF.

Proposed change by UK:

No change required.

Noted

See also comments T1 and T2 from Germany and T3 from Ireland.

T16. Clause 31 (32) – CJK Unified Ideographs Extension F

With respect to the request to remove USAT source references from 50 characters in CJK Unified Ideographs Extension F (See IRG N2041), and consequently remove from CJK Unified Ideographs Extension F those 49 characters that only had a USAT source reference:

A. We agree to the removal of 48 characters with USAT source references;

B. We request that 2D30C (USAT-00856) not be removed as it is a character attested in use in Bernard Karlgren's highly influential *Grammata Serica* (1940) and *Grammata Serica Recensa* (1957). This character and other unencoded characters in Karlgren's *Grammata Serica Recensa* were proposed as part of the UTC Extension F submission (see IRG N1888), but the entire submission was summarily rejected by IRG on procedural grounds, which we consider to have been extremely unfortunate. As this character has been proposed by the UTC we request that it be kept in CJK Unified Ideographs Extension F with the source reference changed to UTC-01155.

C. We note that four of the fifty USAT characters that are requested to be removed are identical to existing encoded characters:

U+2D6AC (USAT-01869) = 2266C 惝

U+2D9AE (USAT-02066) = 2AC87 柚

U+2D9DE (USAT-03431) = 234C3 桤

U+2DBD2 (USAT-01739) = 6FD3 濂

These four characters have simple IDS sequences, and we would have expected that automated IDS checking of CJK-F would have identified these characters as duplicates. That these duplicates were not detected prior to submission to WG2 indicates a failure in the IRG Quality Assurance process, which we consider to be very worrying.

Proposed change by UK:

Remove the following 48 characters from CJK Unified Ideographs Extension F:

2CED4 (USAT-04335)

2D06B (USAT-00332)

2D29A (USAT-01778)

2CEE7 (USAT-04345)

2D16F (USAT-05701)

2D402 (USAT-05302)

2CF42 (USAT-02160)

2D18A (USAT-60046)

2D442 (USAT-03388)

2CF6D (USAT-60012)

2D1BF (USAT-05796)

2D4BD (USAT-04810)

2CF75 (USAT-03966)

2D20A (USAT-01366)

2D4E0 (USAT-90141)

2D531 (USAT-02486)	2D810 (USAT-02889)	2DA24 (USAT-03899)
2D588 (USAT-03840)	2D82D (USAT-00313)	2DA3E (USAT-01373)
2D59C (USAT-03564)	2D82E (USAT-01556)	2DAD2 (USAT-02590)
2D5D0 (USAT-03076)	2D83F (USAT-00055)	2DB04 (USAT-60177)
2D5D4 (USAT-02802)	2D86C (USAT-04803)	2DB41 (USAT-03269)
2D623 (USAT-60095)	2D912 (USAT-03855)	2DBD2 (USAT-01739)
2D6AC (USAT-01869)	2D987 (USAT-00998)	2DC09 (USAT-00268)
2D732 (USAT-60123)	2D98C (USAT-02114)	2DC6D (USAT-00966)
2D744 (USAT-02396)	2D9AE (USAT-02066)	2DCEB (USAT-04194)
2D74E (USAT-03777)	2D9DE (USAT-03431)	2DCF9 (USAT-03483)
2D78E (USAT-03807)	2D9E8 (USAT-04809)	2DDBB (USAT-00307)

Accepted in principle

See also comments T1 from China.

See also comments T18 and G19 from UK which are resulting in the removal of CJK Ext F from this amendment.

T17. Clause 31 (32) – CJK Unified Ideographs Extension F

IRG N2042 identifies 11 further characters in CJK Unified Ideographs Extension F that should be removed as duplicates or unifiable with existing characters:

- 2D127 (USAT-01722) : unifiable with 2057D 𠂇
- 2D3AD (GCY-0697.00) : unifiable with 2144F 塙
- 2D5A5 (USAT-03456) : unifiable with 536E 卮
- 2D666 (USAT-04922) : unifiable with 224BF 𠂇
- 2D6B9 (USAT-01338) : unifiable with 22758 惶
- 2D754 (KC-01326) : unifiable with 2F8B1 𠂇
- 2D834 (USAT-04653) : identical to 2ABBE 攥
- 2D9A4 (USAT-01096) : unifiable with 6752 𠂇
- 2DAE2 (KC-01963) : identical to 27BF8 𠂇
- 2DD0B (JMJ-059937) : unifiable with 20924 眞
- 2DD82 (JMJ-058841) : unifiable with 488B 𠂇

That these unifiable and duplicate characters were only identified after CJK-F was submitted to WG2 further indicates a failure in the IRG Quality Assurance process, and suggests that CJK-F may have been prematurely submitted to WG2.

Proposed change by UK:

Remove the following 11 characters from CJK Unified Ideographs Extension F:

- 2D127 (USAT-01722)
- 2D3AD (GCY-0697.00)
- 2D5A5 (USAT-03456)
- 2D666 (USAT-04922)
- 2D6B9 (USAT-01338)
- 2D754 (KC-01326)
- 2D834 (USAT-04653)
- 2D9A4 (USAT-01096)
- 2DAE2 (KC-01963)
- 2DD0B (JMJ-059937)
- 2DD82 (JMJ-058841).

Accepted in principle

See also comments T1 from China.

See also comments T18 and G19 from UK which are resulting in the removal of CJK Ext F from this amendment.

T18 Clause 31 (32) – CJK Unified Ideographs Extension F

We have carried out a partial review of CJK-F, focusing primarily on the USAT source characters, and note the following issues.

2CEF3 (JMJ-056849) has a round dot above which is not a stroke used in the Han script. Is this really a distinct character? Or is it simply 20000 𠄎 with an editorial dot, in which case it can be represented as 20000 𠄎 plus 0307 combining dot above.

2D13F (USAT-00061) is unifiable with 20991 𠄎.

2D260 (JMJ-059428) should be 30.10 strokes not 30.11.

2D459 (USAT-60078) 𠄎破女 is actually 5619 𠄎. This is evident from the 𠄎 element which is large and not aligned with 波 (cf. the size and position of 𠄎 in 2D446 USAT-00947).

2DB74 (USAT-05567) may be unifiable with 6EDB 𠄎.

2DD0F (JMJ-058197) should be radical 86.13 not 112.12.

Proposed change by UK:

Remove 2CEF3 (JMJ-056849) for further study.

Remove 2D13F (USAT-00061).

Reorder 2D260 (JMJ-059428) as appropriate.

Remove 2D459 (USAT-60078).

Remove 2DB74 (USAT-05567) for further study.

Reorder 2DD0F (JMJ-058197) as appropriate.

Accepted in principle

Based on this, IRG needs to review these comments and issue a new repertoire. As a result, CJK Ext F is removed from this amendment and postponed to the Committee Draft of the 5th edition.

G19 Clause 31 (32) – CJK Unified Ideographs Extension F

We note that a very large number of the JMJ source characters do not appear to be suitable candidates for encoding.

A very large proportion of the JMJ source characters appear to be idiosyncratic, calligraphic or semi-cursive variants of the same character (e.g. 2D004 through 2D007 and 2D009). Encoding these variants seems to us to go against the spirit of Annex S, and we believe that such variants would be best dealt with as IVS sequences.

Some of the JMJ source characters are weird squiggles, which look nothing like CJK characters (e.g. 2CEF8 through 2CEFD and 2CEFF), and we wonder whether they really are distinct CJK characters.

Proposed change by UK:

Consider removing all JMJ source characters from CJK-F for further consideration by IRG, in order to determine which of these characters are appropriate for encoding according to the Character-Glyph model, and which would be better dealt with as IVS sequences.

Accepted in principle

Based on this, IRG needs to review these comments and issue a new repertoire. As a result, CJK Ext F is removed from this amendment and postponed to the Committee Draft of the 5th edition.

E20 file CJKSrc.txt

File is incorrectly named "CJKSrc.txt.txt".

Proposed change by UK:

Ensure file is correctly named "CJKSrc.txt".

Accepted

E21 file CJKSrc.txt

The following source tags are capitalised incorrectly in the initial comment:

```
#      Hanzi G source: kIRG_Gsource
#      Hanzi T source: kIRG_Tsource
#      Kanji J source: kIRG_Jsource
#      Hanja K source: kIRG_Ksource
#      Hanja KP source: kIRG_KPsource
#      ChuNom V source: kIRG_Vsource
#      Hanzi H source: kIRG_Hsource
#      Hanzi M source: kIRG_Msource
#      Unicode U source: kIRG_Usource.
```

Proposed change by UK:

Correct the capitalisation as below:

```
#      Hanzi G source: kIRG_GSource
#      Hanzi T source: kIRG_TSource
#      Kanji J source: kIRG_JSource
#      Hanja K source: kIRG_KSource
#      Hanja KP source: kIRG_KPSource
#      ChuNom V source: kIRG_VSource
#      Hanzi H source: kIRG_HSource
#      Hanzi M source: kIRG_MSource
#      Unicode U source: kIRG_USource
```

Accepted

USA: Negative

Due to a number of technical issues in the PDAM 2.2 ballot (detailed below), the US votes NO. If te.1 – te. 8 are accommodated, the US will change its vote to YES

Technical comment:

TE.1. Nushu

Based on the questions raised in document WG2 N4610 from Suzuki Toshiya, the US considers Nushu not yet mature enough to progress to an enquiry ballot, and requests that the Nushu block be removed from the ballot until the issues are resolved.

Proposed change by US:

Remove Nushu from the ballot.

Accepted

See also comment T7 from Japan and T14 from UK.

TE.2. CJK Unified Ideographs Extension F

The US requests the removal of 49 characters originating from the SAT project, as documented in N4580SATWithdrawnCharacters.xls. This request is supported by a decision of IRG #42.

Proposed change by US:

Remove the characters.

Partially accepted

See also comment T1 from China and T16 from UK.

Only 48 characters are removed. Note that per disposition of comments T18 and G19, CJK Ext F is removed from this amendment.

TE.3. Tangut Radicals

The US requests that the block be renamed “Tangut Components”, since these characters constitute a set of components. Only a subset of them are used as radicals.

Proposed change by US:

Change the name of the block.

Accepted

This also results in character names for the block to use TANGUT COMPONENT-xxx.

TE.4. Tangut Radicals

The US requests that the character names be algorithmically derived, as was done for Tangut Ideographs.

Proposed change by US:

Change the names of the characters as described.

Accepted in principle

The names currently use a naming convention TANGUT RADICAL-xxx which is algorithmically derived with xxx being their sequence order starting at 001 and ending at 753. It is not clear how the other alternative, using their code point, would be an improvement. Furthermore, the current index number is also used as a component/radical number in the code chart for Tangut characters. Note that per disposition of comment TE.3, the names become TANGUT COMPONENT-xxx.

TE.5. Tangut Radicals

The US asks that the annotations which start “used for...” be removed from the names list. Any change in these annotations would require hand-editing by the Editor, and could introduce errors. This material is more appropriate for a separate data file (or a Unicode Technical Note).

Proposed change by US:

Remove the annotations.

Accepted

See also comment T13 from UK.

TE.6. Annex A.5 Moji-Joho-Kiban collection

The US requests the removal of the extended collection “Moji-joho-kiban ideographs”, until after Extension F (“F1”) and Extension G (“F2”) are processed. WG2 N4625 mentions that “2,000 ideographs [in the proposed extended collection] are on the process of CJK Unified Ideographs extensions F1 and F2 standardization.” Adding a collection that includes characters not yet on a ballot is not in accordance with the standardization process.

Proposed change by US:

Remove the extended collection from this Amendment and consider introducing it in a later Amendment or Edition.

Accepted

See also comment E9 from Japan and G19 from UK.

Strictly speaking, the current collection does not include characters not yet on a ballot. It is in fact part of the issue because CJK Extension F (and a new repertoire being worked now by IRG) contains many characters that should be part of that collection. As such, the collection should be added at the same time these proposed characters are encoded, or at least in the pipeline.

It may be difficult to add the collection when the code points are unstable, but it would be feasible to add the collection during the enquiry ballot corresponding to the addition of the last batch of these proposed characters. Finally, the comment G19 from UK put in question the encoding principle for many of the new JMJ characters in CJK Extension F and this may need to be solved before this collection is created.

TE.7. Glagolitic Supplement

The character U+1E00B has an error in its name, “COMBINIGN GLAGOLITIC LETTER I”. Correct the spelling to “COMBINING”.

Proposed change by US:

Correct the spelling as noted.

Accepted

TE.8. Glagolitic Supplement

The spelling of two characters differs between the current Glagolitic block and the Glagolitic Supplement:

(current names in Glagolitic Supplement in PDAM)

1E001 COMBINING GLAGOLITIC LETTER BUKI

1E00E COMBINING GLAGOLITIC LETTER LJUDIE

versus (Glagolitic block)

2C01 GLAGOLITIC CAPITAL LETTER BUKY

2C0E GLAGOLITIC CAPITAL LETTER LJUDIJE

In order to bring the names in line with those in the Glagolitic block, the names should be changed to:

1E001 COMBINING GLAGOLITIC LETTER BUKY

1E00E COMBINING GLAGOLITIC LETTER LJUDIJE..

Proposed change by US:

Change the names of two characters U+1E001 and U+1E00E in the Glagolitic Supplement as noted.

Accepted

---end