
ISO/IEC JTC1/SC2/WG2 N 3325

Date: 2007-09-16

**ISO/IEC JTC1/SC2/WG2
Coded Character Set
Secretariat: Japan (JISC)****Doc. Type:** Draft disposition of comments**Title:** Draft disposition of comments on SC2 N 3940 (PDAM text for Amendment 5 to ISO/IEC 10646:2003)**Source:** Michel Suignard (project editor)**Project:** JTC1 02.10646.00.05**Status:** For review by WG2**Date:** 2007-09-16**Distribution:** WG2**Reference:** SC2 N3940, 3959**Medium:** Paper, PDF file

Comments were received from China, Germany, India, Iran, Ireland, Japan, Korea, United Kingdom, and USA. The following document is the draft disposition of those comments. The disposition is organized per country.

Note – 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*.

China: Positive with comments

China votes YES to ISO/IEC JTC1/SC2 N3940 with technical comments..

Technical comments

T1. CJK Extension C, character deletion

The following 63 characters should be removed from PDAM5:

2A702 2A7F5 2A84A 2A8C3 2A8DB 2A938 2A939 2A93F 2A97F 2A9D0
2AA14 2AA17 2AA35 2AA39 2AA6F 2AA76 2AA77 2AACF 2AAD0 2ABBB
2AC0B 2AC49 2AC55 2AC78 2AC98 2ACAD 2AD0B 2AD73 2AD94 2AE47
2AE5D 2AE68 2AE77 2AEE3 2AF0F 2AF11 2AFAB 2B02C 2B0B6 2B1B9
2B1BD 2B224 2B236 2B25C 2B265 2B29E 2B460 2B488 2B497 2B561
2B58F 2B597 2B5D9 2B611 2B68F 2B129 2AA06 2AA83 2ABE6 2AF2F
2B247 2B5CA 2AEDF.

Propose acceptance

Also requested by Japan, Korea. Note that there are four additional CJK Extension C characters to be removed per IRG request in WG2 N3280 and 3281.

T2. Source reference removal

T sources of the following 4 characters should be removed :

2A8BA 2AC83 2AC85 2AC87

For details, read ISO/IEC JTC1/SC2/WG2 N3280 and N3281 please

Proposed acceptance

Endorsed by IRG.

Germany: Negative

Technical comments

Meitei Mayek

We request to retain the two punctuation characters

U+1CCD MEITEI MAYEK DANDA (ceikhan iyek)

U+1CCE MEITEI MAYEK DOUBLE DANDA (ceikei iyek)

in their positions within the Meitei Mayek block, as proposed in N3206.

We acknowledge the concerns raised about these characters and the suggestion that they could be unified with the Devanagari DANDA and DOUBLE DANDA, but, after reconsidering them and taking the rationale into account which is provided in the original proposal, we think that the case for encoding them separately has been made. Although similar characters in the major Indic scripts have been unified in the past, we note that the opposite has been practiced repeatedly more recently with some belonging minority scripts. Once a policy of not unifying Danda's for minority scripts (Ol Chiki, Saurashtra, Lapcha) has been settled, this direction should be sustained in similar cases (as is with Meitei Mayek). We do not see any possible harm arising from encoding the two characters on their proposed positions. And at the same time, we do not see any benefits from unifying them with the Devanagari Danda's and thus possibly causing cultural issues when the script gets implemented and used. In fact, to stress their uniqueness perhaps the original names should be used, f.ex. CEIKHAN IYEK instead of DANDA.

WG2 discussion

There is nothing to 'retain' as the characters in questions are not part of the amendment. However the comment should probably be understood as a request to encode the two additional Dandas. The comment above does not bring any new element to the discussion made in last WG2 meeting (Frankfurt). Changing the name does not affect what are still inherently Dandas and in fact muddle the issues. If they are added they should be named 'Danda' to make their identity clear. See Irish comment T1 for further consideration.

India: Positive with comments

Technical comments

Change spelling of "Meitei Mayek" script as "Meetei/Meitei Mayek" script everywhere in the document.

Propose non acceptance

See comment T1 from Ireland. In addition the suggested name contains a character 'l' which is not allowed in collection, character name, and block, However, it is always possible to create a note in Figure 4 as already done for New Tail Lue and suggested for Lanna.

Iran (Islamic Republic of): Positive with comments

Technical comments

ISIRI wishes to emphasize its support for encoding of all Avestan punctuation in exactly the same way presented in SC2 N3940.

WG2 discussion

See disposition of Irish comment T.3.

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 13, Row 1C8: Meitei Mayek.

All but two of the characters in the proposal document, N3206, were put on the PDAM ballot at the Frankfurt meeting of WG2. The two characters which were not accepted were the punctuation characters DANDA (ceikhan iyek) and DOUBLE DANDA (ceikhei iyek). According to the Unicode Standard, the nine *major* Indic scripts (Devanagari, Bengali, Gurmukhi, Gujarati, Oriya, Tamil, Telugu, Kannada, Malayalam) are all supposed to use DEVANAGARI DANDA and DEVANAGARI DOUBLE DANDA. There are users of those scripts who disagree with this choice, but the UTC has chosen so far not to disunify them for the major scripts of India. UCS practice for other minority scripts in India and for scripts outside of India has not, typically, been subject to this constraint.

- | || Devanagari DANDA and DOUBLE DANDA have a typically thick shape with angled top and bottom.
- ๑| Thai has ANGKHANKHU (like a double danda but with a different shape)
- | || Tibetan has SHAD and NYIS SHAD.
- | || Myanmar has LITTLE SECTION and SECTION.
- / || The four Philippine scripts share a SINGLE PUNCTUATION and a DOUBLE PUNCTUATION (though only Hanunoo really uses them).
- 𑄎 𑄎| Khmer KHAN and BARIYOOSAN have “flags” like the Thai ANGKHANKHU.
- ꦱ ꦱ Balinese has CARIK SIKI and CARIK PAREREN.
- | || Ancient Kharoshthi has a DANDA and a DOUBLE DANDA.
- | || 𑄎 Cham (which is under ballot) has a DANDA, a DOUBLE DANDA, and a TRIPLE DANDA.
- | || 𑄎 Lanna has characters KAAAN and KAANKUU which look like Myanmar LITTLE SECTION and SECTION, and SATKAAN and SATKAANKUU which look like the Khmer KHAN and BARIYOOSAN.
- | Kayah Li has a single SHYA.

All of these scripts above are used outside of India. Some scripts used within India also have script-specific punctuation:


- | || Lepcha has TA-ROL and NYET THYOOM TA-ROL.
- | || OI Chiki has MUCAAD and DOUBLE MUCAAD.
- | || Saurashtra has DANDA and DOUBLE DANDA.

We have not seen any principled reason why the Meitei Mayek script punctuation should be treated differently from the punctuation of Lepcha, OI Chiki, and Saurashtra. The uniqueness of the Meitei Mayek script is intimately tied in with the cultural identity of its user community. Meitei experts Dr Chungkham Yashawanta Singh and Dr L. Sarbajit Singh of the Department of Linguistics, Manipur University, have reiterated their support for the encoding of script specific punctuation for Meitei Mayek (see page 5 below). No cogent argument has been given for the unification of Meitei Mayek punctuation with any particular script (Devanagari? Saurashtra?). We understand that the US National Body has a position of “avoiding the encoding of additional Danda characters” but we do not see how this position provides any advantage to the Meitei user community. Nor do we see how the Meitei user community (or anyone else) could be *disadvantaged* by the encoding of Meitei-specific punctuation. We therefore request the addition of the following characters to the Meitei Mayek block.

- | U+1CCD MEITEI MAYEK DANDA (ceikhan iyek)
- || U+1CCE MEITEI MAYEK DOUBLE DANDA (ceikhei iyek)

We would also find an encoding with their Meiteilon names to be acceptable:

- | U+1CCD MEITEI MAYEK CEIKHAN IYEK
- || U+1CCE MEITEI MAYEK CEIKHEI IYEK

Regarding the name of the script, Ireland reiterates its support of the spelling “Meitei” as this correctly transliterates  *meitei* and is the most common spelling found in English language literature. We are sensitive to the preference of some revivalists to the spelling “Meetei” and would not object to its being added in a similar way to notes which have been previously added for scripts like Lanna.

WG2 discussion

Duplicating the same symbols over and over is not optimal. Keeping duplicating them in minority scripts is not an established policy. Duplicating the same symbol makes implementation more complicated because they have to add more test cases to detect those superfluous Dandas.

If anything, the comment shows that many punctuations that could have been unified in the past were not, and this without clear policy. Punctuation unification does not prevent the representative glyph to be tailored for a specific writing system. Finally, many Indic experts also do not see the need for further Danda dis-unification. Sharing punctuation among various writing systems is a common feature and a benefit of the standard as it avoids duplication of

what are inherently single semantic units which can easily differentiated at presentation time. Before adding any new Dandas a clear policy is needed to established when it is appropriate to create new dandas.

T2. Page 19, Row A6A: Bamum.

Ireland is confident that the set of characters in this block is suitable for encoding but is sensitive to a request from the user community that processing of the script be delayed for further study. Without prejudice to a possible future encoding of the Bamum syllabary in the BMP, we request that this script be removed from the PDAM.

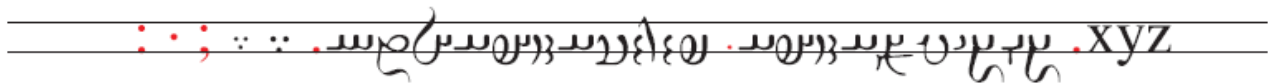
Propose acceptance

Similar comment from US.

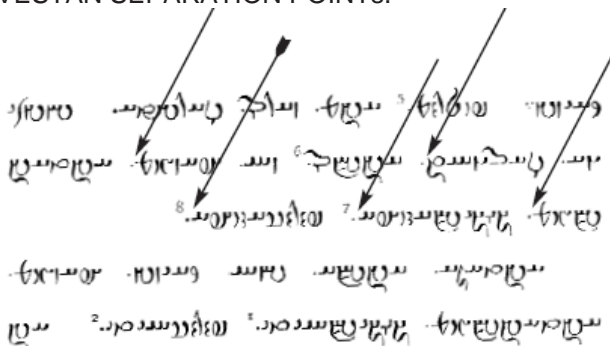
T3. Page 26, Row 10B0: Avestan.

Ireland reiterates its support for the retention of U+10B38 AVESTAN SEPARATION POINT, a character distinct from FULL STOP which is also used in Geldner's edition of the Avesta. We acknowledge that in some Avestan texts, a WORD SEPARATOR MIDDLE DOT is used, and certainly agree that such a character would be useful for some editors. We would agree to its encoding at U+2E31. But in Karl Geldner's very careful Avestan typography (and Geldner's edition is the standard edition of the *Avesta*), the AVESTAN SEPARATION POINT follows on directly at the end of a word and is smaller than than (and distinguished from) the FULL STOP. It is not "yet another middle dot" as it rests on the baseline. And it does not behave as the proposed WORD SEPARATOR MIDDLE DOT does; that character is typically centred between two words, and raised from the baseline.

Here we show COLON, MIDDLE DOT, SEMICOLON compared with TINY TWO DOTS OVER ONE DOT PUNCTUATION, SMALL TWO DOTS OVER ONE DOT PUNCTUATION, FULL STOP at the end of an Avestan word and AVESTAN SEPARATION POINT at the end of another Avestan word, and FULL STOP and three Latin letters to illustrate the baseline.



Below is a passage of Avestan text from Geldner 1880 showing FULL STOP (with fletched arrow) alongside AVESTAN SEPARATION POINTs.



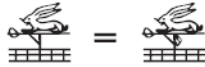
We therefore reiterate our support for the retention of the AVESTAN SEPARATION POINT character in the Avestan block, unless, along with a generic WORD SEPARATOR MIDDLE DOT a generic baseline smaller-than-FULL-STOP word separator dot be added as well.

WG2 discussion

It is not clear whether a thinner variant of the period (full stop) introduced by Geldner should be preserved in the encoding. At this point there are already many periods and various dot punctuations implemented in generic or script-specific blocks, although all such punctuations located in the alphabetic baseline have the same size.


T4. Page 29, Row 1300: Egyptian Hieroglyphs. (delete 130FB)

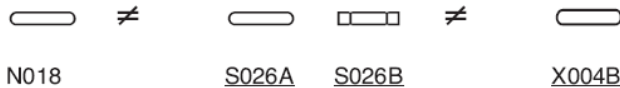
The character at U+130FB EGYPTIAN HIEROGLYPH E035 (shown on the left below) is a duplicate of U+13247 EGYPTIAN HIEROGLYPH NU015 (shown on the right), which refers to the Upper Egyptian Nome of Wn. It should be deleted and the remaining characters moved up one position.



Proposed acceptance

T5. Pages 31 and 32, Row 1300: Egyptian Hieroglyphs.(add one character after 133D2)

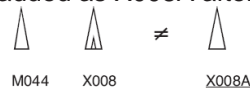
The character U+13201 EGYPTIAN HIEROGLYPH N018 represents “a sandy tract”. Gardiner describes other characters “with a similar outline” which, because they are different characters, should be encoded uniquely. The first character to be added is the “garment” or “roll of cloth” S026A, which Allen also distinguishes as A1999:441:130a; this also has another variant given at G1957:507:(N18) which should likewise be added as S026B. These should be added after U+132F0 EGYPTIAN HIEROGLYPH S026. The third character to be added is the “roll of bread” X004B(compare  X004), which Möller also distinguishes as M1927:50:554. It should be added after U+133D2 EGYPTIAN HIEROGLYPH X004A.



Proposed acceptance







T6. Page 32, Row 1300: Egyptian Hieroglyphs. (add one character after 133D7)

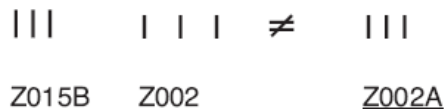
The character U+131EF EGYPTIAN HIEROGLYPH M044 represents a thorn. The conical loaf U+133D7 EGYPTIAN HIEROGLYPH X008 has a Hieratic variant without the inner markings. This character should be added as X008A after U+133D7; Möller distinguishes the two as M1927:50:553 and M1927:51:569.



Proposed acceptance

T7. Page 32, Row 1300: Egyptian Hieroglyphs. Gardiner gives the example  ddt.

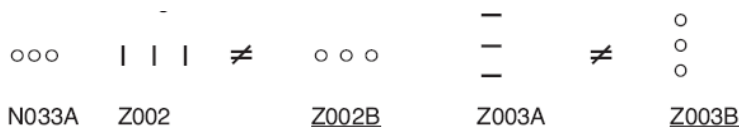
These three strokes are an alternative form of Z002 and should also be distinguished from numeral three  (Z015B). Allen (A1999:447) notes  as a variant of . Faulkner gives many examples such as  *imyw mw* (F1962:18:aquatic animals) and  *wd3w* (F1962:75:amulets) where  is used as a ligature in the expression of plurality. The character should be added as Z002A after U+133E2.










WG2 discussion

T8. Page 32, Row 1300: Egyptian Hieroglyphs. (add two characters after 133E2 and 113E4)

The character U+13201 EGYPTIAN HIEROGLYPH N033A represents three grains of sand. Gardiner notes that the plural strokes are sometimes represented by a similar character; two other characters “with a similar outline” which, because they are different characters, should be encoded uniquely. The first character should be added as Z002B after U+133E2; the second should be added as Z003B after U+133E4. Möller distinguishes the second as M1909:54:563.



WG2 discussion

T9. **Page 32, Row 1300: Egyptian Hieroglyphs.** Gardiner gives the example  *nfrw* ('beautiful', m. plural). Allen (A1999:447) notes  and  as variants of . Faulkner gives  *nfrwy* (F1962:126:citizens, townsmen) and other examples. Faulkner also gives five uses of the  version – example  *wh'rw* (fowling). Both should be added as Z002C and Z002D respectively after U+133E2.



WG2 discussion

Editorial comments

E1. Page 9, Row 110: Hangul Jamo.

With reference to ISO/IEC JTC1/SC2/WG2 N3172 “Add annotations for existing 5 Hangul Jamo names”, Ireland requests that the following annotations be added:

- 11EC HANGUL JONGSEONG IEUNG-KIYEOK (yesieung-kiyeok)
- 11ED HANGUL JONGSEONG IEUNG-SSANGKIYEOK (yesieung-ssangkiyeok)
- 11EE HANGUL JONGSEONG SSANGIEUNG (ssangyesieung)
- 11EF HANGUL JONGSEONG IEUNG-KHIEUKH (yesieung-khieukh)

Accepted

This was part of resolution M50.34.

E2. Row 313: Hangul Compatibility Jamo.

With reference to ISO/IEC JTC1/SC2/WG2 N3172 “Add annotations for existing 5 Hangul Jamo names”, Ireland requests that the following annotation be added:

- 3180 HANGUL LETTER SSANGIEUNG (ssangyesieung)

Accepted

See E.1.

E3. Page 16, Row 2CB: Coptic.

In response to feedback received from Copticists, Ireland requests that the shapes of U+2CC6 and U+2CC7 be angled clockwise away from the vertical. The new shapes are shown below.



Proposed acceptance

E4. Page 28, Row 1300: Egyptian Hieroglyphs.

The glyphs for U+13007 EGYPTIAN HIEROGLYPH A006A and U+13008 EGYPTIAN HIEROGLYPH A006B should be reversed.



Proposed acceptance

E5. Page 30, Row 1300: Egyptian Hieroglyphs.

The glyphs for U+13263 EGYPTIAN HIEROGLYPH O010A is incorrect. The “door” (the small square) should be in the bottom right of the square, not the top right.



Proposed acceptance

Japan, Negative

Japan votes against SC2 N3940 (ISO/IEC 10646:2003 PDAM5) with the following comments. Japan will change its vote if they are accepted accordingly.

Technical comments

[JP1] On the new subclause 27.1 Source references for CJK Unified Ideographs,

The format for the G source entries include "(G_KXdddd)". However, five digits is insufficient to identify a head character in the Kangxi Dictionary. Japan wants to change the format for G_KX ideographs.

One possibility is to use a format "(G_KXdddd.ddd)" where dddd.ddd is the kangxi index as used in IRG works. Any other format is fine as long as it can unambiguously identify a character in Kangxi dictionary. The actual CJKU_SR.txt file should also be updated to use the new format.

Proposed acceptance

[JP2] On the new text for subclause 34.2 Character names list,

It is unclear that which item in the name list is normative and which is informative. Current text in the clause (which itself is normative) uses the wording like: "some" or "such as". The text should look like something like:

The name list contains both normative and informative information. The code position and the associated character name shall be normative. Other information shall be informative. Informative items contain:

Proposed acceptance in principle

The paragraph will be rewritten to make clear which parts of the name list description are normative and which parts are informative. Clearly the name and code position are normative. The issue is a bit more complicated for information related to normalization, such as the mapping preceded by '≈' where the normative definition is insured by the normative reference to the Unicode UAX#15.

[JP3] On CJK UNIFIED IDEOGRAPHS EXTENSION C,

Japan supports inputs from IRG. The draft should be updated to reflect the changes proposed in WG2N3280 and N3281.

Proposed acceptance

[JP4] On some Japanese glyphs for CJK UNIFIED IDEOGRAPHS EXTENSION C

During the PDAM review period, Japan made some extensive review and improved design of some example glyphs that are originally submitted from Japan. Japan wants to replace those example glyphs with new ones. (See attachment for details.)

Proposed acceptance in principle

This means that the relevant characters in WG2 N3279 need to have their glyph replaced accordingly.

[JP5] On Names list

Since we are changing the format of Names list (included in the clause 34), we need to update all names lists to follow the new format, before the final publication of the Amendment 5. For the moment, Japan wants to see some comments (such as Editor's note) that says all names lists will be updated before final publication in the draft.

Proposed acceptance in principle

This implies publication of all chart pages which means that it makes sense to publish a new consolidated edition of ISO/IEC 10646 instead of just the amendment.

[JP6] Additional ideographs

Japan wants to have six ideographs proposed in WG2N3318 be in the amendment 5, assuming WG2 agrees on it.

WG2 discussion

Korea (Republic of): Negative

XX DISAPPROVAL OF THE DRAFT FOR REASONS ON THE ATTACHED

XX Acceptance of these reasons and appropriate changes in the text will change our vote to approval

Technical comments:

T1. p. 1, right column,

- Replace "K5H" with "K5"

- reason: To be consistent with other sources (K0, K1, K2, K3, and K4), we would like to replace "K5H" with "K5".

Proposed acceptance in principle

This requires the chart produced by IRG to be updated accordingly.

T2. p. 2, left column,

- Replace "K5Hdddd" with "K5-dddd"

- reason: To be consistent with other sources (K0-hhhh, K1-hhhh, K2-hhhh, K3-hhhh, and K4-hhhh), we would like to replace "K5Hdddd" with "K5-dddd". (Note that five digits are replaced with four digits. The leftmost 'd' in "K5Hdddd" format is always '0' and will be removed in "K5-dddd" format.)

- Note: K5 uses a decimal notation in contrast with a hexadecimal notation in K0-hhhh, K1-hhhh, K2-hhhh, K3-hhhh, and K4-hhhh. We will check later if it will be better to use a hexadecimal notation in K5 too.

Proposed acceptance in principle

As suggested, the change should go all the way to move from decimal to hexadecimal notation. This can be easily achieved algorithmically for the CJKU_SR.txt. This is indeed similar to what was done for the K4 source. Note that this would impact the chart (ref comment T.1)

T3. p. 3, left column, 34.2 Character names list

- We suggest that the formats of character names be discussed at the Hangzhou meeting. This issue is relevant with many other scripts.

Proposed acceptance in principle

This should be however constrained within the scope of existing contributions, including the comments received for this ballot. Also refer to minutes concerning agenda item 11.2 in document WG2 N3253.

T4. p. 10, left column,

- delete "Korean combining alphabet".

- reason:

1) Hangul Jamos are not combining (Note 1, Clause 26.1 is shown below).

NOTE 1 _ Hangul Jamo are not combining characters.

2) We would like to keep the title "Hangul Jamo".

Proposed acceptance in principle

The proposed change is to move from 'Korean combining alphabet' to 'Korean Conjoining alphabet'. It should also be noted that the title for these pages has always been 'Hangul Jamo'.

T5. p. 10, left column,

- delete "initial consonants"

- reason: We suggest that the formats of character names be discussed at the Hangzhou meeting.

- Note: The term "syllable-initial (-final) character", instead of "initial (final) consonants", is used in ISO/IEC 10646.

Proposed acceptance in principle

The two terminologies should be aliased (like in the Unicode Standard) to allow both uses and therefore remove the need for deletion.

T6. p. 11, left column,

- delete "Old Hangul initial consonants"

- reason:

- 1) We suggest that the formats of character names be discussed at the Hangzhou meeting.
- 2) Furthermore, the title will give a wrong impression that characters under the subtitle "initial consonants" are NOT Old Hangul initial consonants. However, there are many Old Hangul initial consonants under the subtitle "initial consonants".

- Note: The term "syllable-initial (-final) character", instead of "initial (final) consonants", is used in ISO/IEC 10646.

Proposed acceptance in principle

A small note can be added to Old Hangul initial consonants to convey that Old Hangul initial consonants can be from either sub-block. Concerning terminology see disposition of comment T5 .

T7. p. 11, left column,

- delete "Medial vowels"

- reason:

- 1) We suggest that the formats of character names be discussed at the Hangzhou meeting.

- Note: The term "syllable-peak character", instead of "medial vowels", is used in ISO/IEC 10646.

Proposed acceptance in principle

See disposition of comment T5 .

T8. p. 11, right column,

- delete "Old Hangul medial vowels"

- 1) We suggest that the formats of character names be discussed at the Hangzhou meeting.

2) Furthermore, the title will give a wrong impression that characters under the subtitle "medial vowels" are NOT Old Hangul medial vowels. However, there are many Old Hangul syllable-peak characters under the subtitle "medial vowels".

- Note: The term "syllable-peak character", instead of "medial vowels", is used in ISO/IEC 10646.

Proposed acceptance in principle

A small note can be added to Old Hangul medial vowels to convey that Old Hangul initial consonants can be from either sub-block. Concerning terminology see disposition of comment T5 .

T9. p. 11, right column,

- delete "final consonants"

- reason: see comment 5 above.

- Note: The term "syllable-final character", instead of "final consonants", is used in ISO/IEC 10646.

Proposed acceptance in principle

See disposition of comment T5 .

T10. p. 12, right column,

- delete "Old Hangul final consonants"

- reason: see comment 6 above.

Proposed acceptance in principle

See disposition above.

T11. p. 21, right column,

- delete "Korean combining alphabet".

- reason: see comment 4 above.

Proposed acceptance in principle

See disposition above.

12. p. 21, right column,

- delete "Old Hangul initial consonants"

- reason: see comment 6 above.

Proposed acceptance in principle

See disposition above.

13. p. 25, left column,

- delete "Korean combining alphabet".
- reason: see comment 4 above.

Proposed acceptance in principle

See disposition above .

14. p. 25, left column,

- delete "Old Hangul medial vowels"
- reason: see comment 8 above.

Proposed acceptance in principle

See disposition above .

15. p. 25, left column,

- delete "Old Hangul final consonants"
- reason: see comment 6 above.

Proposed acceptance in principle

See disposition above .

16. Missing annotations

According to Resolution M50.34 (Hangul Jamo additions), WG2 accepted to add five annotations proposed in document WG2 N3172 (2006-09-27). However, those five annotations are missing in the PDAM5.

Therefore we request that five annotations be added as shown below:

16.1 p. 12, add four annotations as shown below:

U+11EC HANGUL JONGSEONG IEUNG-KIYEOK

=> HANGUL JONGSEONG IEUNG-KIYEOK (yesieung-kiyeok)

U+11ED HANGUL JONGSEONG IEUNG-SSANGKIYEOK

=> HANGUL JONGSEONG IEUNG-SSANGKIYEOK (yesieungssangkiyeok)

U+11EE HANGUL JONGSEONG SSANGIEUNG

=> HANGUL JONGSEONG SSANGIEUNG (ssangyesieung)

U+11EF HANGUL JONGSEONG IEUNG-KHIEUKH

=> HANGUL JONGSEONG IEUNG-KHIEUKH (yesieung-khieukh)

16.2 The relevant page (page 227 in ISO/IEC 10646:2003) is not included in PDAM5 (one annotation)

- add one annotation as shown below:

U+3180 HANGUL LETTER SSANGIEUNG

=> HANGUL LETTER SSANGIEUNG (ssangyesieung)

Accepted

17. pp. 44 ~ 175, K5 source references

we request that source references for 405 CJK characters whose source reference is K5 be corrected. An example will be given:

- p. 44, for U2A710, the source reference is given as "K5H0041". The correct one is "K5-0418" (a digit '8' to the right of "0041" is missing).

- The same comment applies to the other 404 CJK characters whose source reference is K5.

31. In file "Am5names.txt", we request that comment 16 be reflected.

41. In file CJKU_SR.txt, we request that comment 2 be reflected.

Accepted

United Kingdom: Negative

The UK votes to DISAPPROVE the amendment, with the following technical and editorial comments. If our comments are satisfactorily resolved we will change our vote to APPROVAL.

Technical comments:

T.1 CJK Extension C Removals

The UK is in favour of removing the 67 characters specified in N3280. In addition, the UK now believes that U+2AEEF should be unified with U+24814, and therefore requests that it be removed from Amd.5. We further request that the code chart be reordered so that there are no gaps in it.

Propose acceptance

T.2 CJK Extension C Glyph Changes

The UK is in favour of the glyph changes for 14 characters specified in N3281. In addition the UK has noticed that the following Japanese source glyphs for CJK-C do not exactly correspond to the glyph for the character given in the supporting evidence (IRG N1225):

U+2A761 [JK-65028] . Cf. IRG N1225 part 2 page 41.

U+2ACCC [JK-65156]. Cf. IRG N1225 part 3 page 97.

U+2AE9E [JK-65370]. Cf. IRG N1225 part 3 page 36.

U+2AFA7 [JK-65424]. Cf. IRG N1225 part 2 page 46.

U+2B057 [JK-65465]. Cf. IRG N1225 part 3 page 101.

U+2B318 [JK-65704]. Cf. IRG N1225 part 2 page 50.

U+2B340 [JK-65723]. Cf. IRG N1225 part 2 page 51.

We request that the correct glyphs for these characters be confirmed, and any necessary glyph changes made.

Note that U+2B340 also has a T source (TD-566D), and so if the J-source glyph is incorrect the T and J source characters may need to be disunified.

Proposed acceptance on principle

Only U+2AFA7 is common with Japanese comment JP4!

T.3 Page 2 : 27.3 Source reference presentation for CJK Unified Ideographs Extension C

There are 9 CJK-C characters with KP source references, but there is no KP column, and the code table does not show any of the nine KP sources. The code table should be revised to include KP sources.

WG2 discussion

This may require a new CJK Extension chart format

T.4 Page 13 : Meitei Mayek

The UK does not believe that it is appropriate to unify the Meitei Mayek dandas with any existing dandas, and therefore requests the addition of the following two characters, as proposed in N3206:

1CCD MEITEI MAYEK DANDA (ceikhan iyek)

1CCE MEITEI MAYEK DOUBLE DANDA (ceikhei iyek)

WG2 discussion

See Irish comment T1

T.5 Page 3 : 34.2 Character names list

> Aliases, either preceded by "=" or "{U+203B}" indicate alternate names for characters.

There is no explanation of the difference between "=" and {U+203B} (REFERENCE MARK). In Unicode the former indicates an informal alias which may be changed or removed at will, whereas the latter indicates a formal alias for a defective character name, and once added may not be changed or removed. It would seem important to make this distinction clear in the text.

This also raises the wider question of whether ISO/IEC 10646 should also define formal aliases for defective character names, mirroring Unicode.

WG2 discussion

Editorial comments

E.1 Pages 1-2 : Source references for CJK Unified Ideographs

> The current full set of CJK Unified Ideographs is represented by the collection 384 CJK UNIFIED IDEOGRAPHS-2007 (See annex A.1).

"384 CJK UNIFIED IDEOGRAPHS-2007" should be "385 CJK UNIFIED IDEOGRAPHS-2008"

> 2nd field: Hanzi G sources(G0-hhhh),

> (G1-hhhh), (G3-hhhh), (G5-hhhh),

> (G7-hhhh), (GS-hhhh), (G8-hhhh),

> (G9-hhhh), (GE-hhhh), (G_KX),
> (G_KXdddd), (G_HZ), (G_HZdddd),
> (G_CY), (G_CH), (G_CHdddd), (G_HC),
> (G_HCdddd), (G_BK), (G_BKdddd),
> (G_FZ), (G_FZdddd), (G_4K),
> (G_GHdddd), (G_GJZdddd),
> (G_XCdddd), (G_CYYdddd),
> (G_ZFYdddd), or (G_ZJWdddd).
"(G_KXdddd)" should be "(G_KXdddddd)"
"(G_GHdddd)" should be "(G_GHdddddd)"
"(G_XCdddd)" should be "(G_XCdddddd)"
> In the linked content file (CJKU_SR.txt) make the following
> modifications:
> Remove the V0-417A source from 04443 entry; Remove the T4-6E3B source
> from 04695 entry; Remove the V2-8D4D source from 06F58 entry; Remove
> the TF-3862 source from 0FA23 entry; In the 24319 entry, replace
> "G_FZ_BK" with "G_FZ".

These modifications have already been introduced in Amd.4, and so should be removed. In addition, for consistency with the other ChuNom V sources, the "V04" source should be changed to "V4" throughout the text and in CJKU_SR.txt.

WG2 discussion

These modifications were in PDAM4 but are not in FPDAM4 because CJK Ext C was moved to the latter amendment. However the change from V04 to V4 makes sense, similar to US comment E2.

USA: Negative

Technical comments:

The US National Body is voting No with comments on the following SC2 ballot. Satisfying technical comment T.5 would change the vote into a Yes.

Technical Comments:

T.1 Code tables versus code charts (page 1)

Because, in essence, this amendment has replaced code tables by code charts, and to better align with the Unicode Standard terminology, the 'code chart' terminology should be added to 10646. This can be achieved by amending the 'code table' entry in the clause 4 (Terms and definitions) by aliasing code table and code chart, and replacing all occurrence of code table by code charts in the standard text. By not being table anymore, we can remove the need of numbering these entities which has been a major and tedious work item in preparing the past amendments.

Propose acceptance

T.2 New collection numbers (page 4)

In page 4 of the amendment, the collections number for the proposed scripts are incorrect because they re-use number already allocated to scripts added by amendment 4. So the text after '29 HANGUL JAMO' should read:

In the list of collection numbers and names, after

143 CHAM

insert new entries as follows:

144 MEITEI MAYEK	1C80-1CCF
145 BAMUM	A6A0-A6FF
146 HANGUL JAMO EXTENDED-A	A960-A97F
147 TAI VIET	AA80-AADF
148 HANGUL JAMO EXTENDED-B	D7B0-D7FF

These collection numbers are again used in the next page (ref Note 3 about keywords), and therefore they should be updated there as well. See further comment about Bamum (T.6), which, if accepted, would affect the list above.

Propose acceptance

T.3 CJK Collection Extension C content (page 5)


The fixed collection is determined by the range ‘2A700-2B77A’ which is incorrect because there are unassigned code positions in that range.

Propose acceptance

This would however be fixed by removing all holes from the collection

T.4 CJK Extension C content (page 5)

When CJK Extension C was part of amendment 4, it was discovered that the proposed CJK ideographs 2B151 had an incorrect U source reference which was removed as a result. Further investigation has determined that the U source should UTC00029 (instead of UTC00022) for that ideograph. In addition, the US would also like to get the character referenced by UTC00022 to be added in CJK Extension C or in a future extension (such as extension D). The

representative glyph is  .

WG2 discussion

T.5 CJK Extension C content (page 5)

In its latest meeting (#28), IRG requested to remove an additional 67 Unified ideographs WG2N3280. The US is in favor of these removals. However the remaining characters need to be re-arranged so that they are contiguous. In addition, the US is in favor of the 14 glyph updates as indicated in WG2N3281.

Propose acceptance

T.6 Bamum

Based on user community feedback as expressed by document WG2 Nxxx(N/A at this moment), the US NB is asking for removal of the Bamum script from this amendment, pending further study from the stakeholder community. This is not a rejection of the encoding proposal but a request to give more time to the user community to come up with an agreeable proposal.

Propose acceptance

T.7 Meitei Mayek

The US NB wants to reiterate its position at avoiding the encoding of additional Danda characters and would not be in favor of encoding Dandas in the Meitei Mayek block.

WG2 discussion

T.8 Avestan

The US NB is opposed to the encoding of yet another middle dot punctuation at position 10B38 (AVESTAN SEPARATION POINT).

WG2 discussion

Editorial Comments:

E.1 CJK Collection renumbering (page 1)

The current full set for CJK Unified Ideographs is now ‘385 CJK UNIFIED IDEOGRAPHS-2008’, not 384.

Propose acceptance

E.2 ChuNom V sources (page 1)

For consistency with other V sources, the V04 source should be renamed ‘V4’.

Propose acceptance

E.3 Source references for CJK Extension C (page 2)

Remove the superfluous ‘.’ at end of paragraph.

Propose acceptance

----end--