ISO/IEC JTC1/SC2/WG2 N 4023

Date: **2011-04-25**

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 4168 (ISO/IEC FCD 10646 3rd Ed., Information Technology – Universal Coded Character Set (UCS))**

Source: **Michel Suignard (project editor)**
Project: JTC1 02.10646.00.00.00.03
Status: For review by WG2
Date: 2011-04-25
Distribution: WG2
Reference: SC2 N4168, N4181
Medium: Paper, PDF file

Comments were received from Armenia, Egypt, Germany, Ireland, Japan, Korea (ROK), U.K, and U.S.A. 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***.**

Armenia (comment not related to a vote)

Technical comments

T1. Currency symbols

It is clear idea to combine all the signs, including the monetary one, in one and the same national block, however it is preferable to place Armenian Dram Symbol into Currency Symbols table. Armenian Dram symbol has two horizontal strokes like the majority of the symbols in that table, and those symbols are grouped there together on the basis of functionality and symbolism.

Propose noted

The preference for moving the Armenian Dram symbol is noted. However a similar argument was made by Armenia in the previous phase and it was not accepted. The rationale was given in the following disposition which still applies (extract from document WG2 N3936, page 2, disposition of comment T1.a) from Armenia):

<<

There is no strict rule concerning the placement of currency symbols in the standard. At this point, many currency symbols are not encoded in the Currency Symbols block, but instead in their script block (for example THAI CURRENCY SYMBOL BAHT in 0BF9, GUJARATI RUPEE SIGN in 0AF1) or in other block. The table 15.1 in the Unicode Standard (section 15.1) shows these locations. The Currency block tends to be reserved for symbols that are used across multiple script contexts and have no style dependencies with their own script. The script blocks are used when the currency has a close relationship with a letter form in their related script block. This is clearly the case here with the letter form being 0534 ARMENIAN CAPITAL LETTER DA. Based on this, the location stays unchanged.

>>

Egypt: Positive with comments

Technical comments

T1. Arabic character names
The existing Arabic character name used for these letters are not the classical naming, please change to the proposed writing, which will facilitate the understanding for this character by all Arabic speaking users:

- replace all (BEH) with (BA')
- replace all (TEH) with (TA')
- replace all (THEH) with (THA')
- replace all (HAH) with (HA')
- replace all (KHAH) with (KHA')
- replace all (REH) with (RA')
- replace all (ZAIN) with (ZAY)
- replace all (ZAH) with (DHA')
- replace all (FEH) with (FA')
- replace all (HEH) with (HA)
- replace all (YEH) with (YA')
- replace (WASLA) with (WASL)

Not accepted
This comment is almost verbatim the same that was done for the CD ballot and was not accepted. The disposition from the CD ballot (document WG2 N3936, page 3) is provided with slight editorial fixes to take into account what was actually done in the FCD text:
<<
There are several reasons to not accept this request:

1) *The current names have been in use for a long time in ISO/IEC character standards, not only in ISO/IEC 10646, but also ISO/IEC 8859-6 and others.*
2) *Per sub-clause 24.2 the character names cannot include an apostrophe.*
3) *Per clause 7, the character names cannot be changed.*

A short note has been added in the Arabic block (0600-06FF) for Arabic letters based on ISO 8859-6:

Arabic letter names follow romanization conventions derived from ISO 8859-6. These differ from the Literary Arabic pronunciation of the letter names. For example, U+0628 ARABIC LETTER BEH has a Literary Arabic pronunciation of ba'.
>>

Germany: Positive with comments

Technical comments

T1. ARABIC LETTER REH WITH LOOP
The name of the proposed character U+08AA RABIC LETTER REH WITH LOOP obviously is mistyped (at the beginning, an "A" is missing).
Proposed change by Germany
U+08AA ARABIC LETTER REH WITH LOOP
Propose not applicable
The typo was present in preliminary charts, but the FCD document is correct.

T2. BATAK SYMBOLS
The following proposed characters:
1BFA BATAK SYMBOL BINDU GODANG
1BFB BAKAK SYMBOL PINARJOLMA
are part of the Batak script, and necessary to represent Batak texts in a typographically appropriate way
Proposed change by Germany
Germany opposes to requests to remove these characters from the FCD, if such are raised.
WG2 discussion
See also comments T2 from US and T1 from Ireland. Additions to the standard have to be agreed by NB before formal inclusion, and when character additions are controversial it has always been the policy of this SC to remove them from final document (documents that have no further technical input) and postpone them for further feedback. Even the nature of these glyphs is controversial. The US consider them to be 'graphic page elements that do not behave at all as characters'. Ireland asserts that these 'have been set in lead type as text elements'.
The prudent approach would be, barring a new consensus during the disposition meeting, to postpone the encoding of these proposed characters.

T3. RAISED SYMBOLS (2E33-2E34)
The names of the following proposed characters:
U+2E33 RAISED DOT
U+2E34 RAISED COMMA
which are to be displayed slightly above of the baseline, are misleading, as the term "raised" in the names of Unicode characters until now denotes a position above the x-height line, as for:
U+18DF CANADIAN SYLLABICS FINAL RAISED DOT
U+2E0C LEFT RAISED OMISSION BRACKET
For more details, see document JTC1/SC2/WG2 N3984
Proposed change by Germany
Germany requests not to introduce unnecessary arbitrarity and inconsistency into Unicode character naming.
Germany suggests the following names for the proposed characters:
U+2E33 LIFTED DOT
U+2E34 LIFTED COMMA
For more details, see document JTC1/SC2/WG2 N3984.
WG2 discussion
The editor has no strong preference either way but does not necessarily see why 'raised' is bad. Usage of the various positioning adjectives in names is not formally specified in either Unicode or ISO/IEC 10646 and trying to introduce such formalism after the fact is not necessarily productive, given the many imperfect usages already in place.

T4. LATIN LETTER GLOTTAL DOT (2E33-2E34)
The following proposed character:
U+A78F LATIN LETTER GLOTTAL DOT
is required for the applications shown in document JTC1/SC2/WG2 N3567.
More applications are shown in document JTC1/SC2/WG2 N3984.

Also, it is typographically distinct from U+00B7 MIDDLE DOT.
For more details, see document JTC1/SC2/WG2 N3984.
Proposed change by Germany
Germany strongly opposes to requests to remove this character from the FCD, if such are raised.
Germany suggests the name and annotations to be changed to:
U+A78F LATIN LETTER ELEVATED DOT
= turned full stop (referring to metal type)
• usually displayed at x-height
• used as glottal dot in transliteration for Phags-Pa and phonetic transcription for Tangut
For more details, see document JTC1/SC2/WG2 N3984.

WG2 discussion

See comments T1 from US and T2 from Ireland. This is another controversial proposed character which will have to be postponed unless a consensus is reached during disposition.

T5. Addition to collection 288 MULTILINGUAL LATIN SUBSET

The following proposed character:
U+2E3A TWO-EM DASH
was accepted after the German request which yielded the inclusion of the character collection 288 "Multilingual Latin Subset" in the current FCD. This character collection contains the characters which can be input by means described in the current version of the keyboard standard ISO/IEC 9995-3. However, the current draft of the German keyboard standard DIN 2137 has added U+2E3A TWO-EM DASH to the list of imputable characters by extending that means, and Germany will request to add this character to ISO/IEC 9995-3 on the next due revision.
Proposed change by Germany
Germany requests the addition of U+2E3A TWO-EM DASH to the character collection 288 "Multilingual Latin Subset".

WG2 discussion

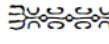
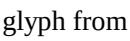
That collection was never part of an amendment and was added directly to the text of 10646 2nd edition per resolution WG2 M55.11. The intent of the addition as reflected in documents N3685 and N3704 has one unsolved aspect, which is either the collection should be 'fixed' or not. N3685 used the following line:
284 MIS-1 see A.4.4 *
[284 and MIS-1 were tentative values which were changed into 288 and MULTILINGUAL LATIN SUBSET]
The '' typically denotes a fixed collection, but the resolution M55.11 (part of N3704) does not clarify that point. The 2nd edition of 10646 does not have '*', but in sub-clause A.5.8 the collection 288 is mentioned as 'fixed'. There are two alternatives:*

1) *Remove the term fixed from sub-clause A.5.8, and add A78F to the collection as requested by Germany,*
2) *Add a '*' in sub-clause A.1 for collection 288, making fully consistent with current A.5.8, and add a new collection (fixed?) 289 made of 288 augmented with A78F.*

Ireland, Negative

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

Technical comments

T1. Page 219, Row 1BC: Batak
Ireland reiterates its support for the characters being balloted at 1BFA and 1BFB. Ireland opposes the removal of these characters from the CD as it has been demonstrated that they have been set in lead type as text elements, regardless of the fact that the glyphs in handwritten manuscripts are often larger and more ornate. We note that the same could be said for illuminated initial Latin letters in the Book of Kells. We would favour reducing the size of the code chart glyph from  to  . The size is merely glyph variation. The characters' semantic is to indicate the beginning of documents, just as similar characters in Tibetan and Sundanese and Javanese do.
WG2 discussion
See also comments T2 from US and T2 from Germany.

T2. Page 1063, Row A72: Latin Extended-D
Ireland reiterates its support for the character being balloted at A78F, LATIN LETTER GLOTTAL DOT. Ireland opposes the removal of A78F LATIN LETTER GLOTTAL DOT from the FCD. We note that WG2 has changed its name from LATIN LETTER MIDDLE DOT (as it was in the CD) to prevent any confusion between this letter and the punctuation mark 00B7 MIDDLE DOT. The character is required for use as a phonetic letter (not a punctuation mark) in transcriptions of Tangut and Chinese, and our approval of this ballot is in part contingent upon the retention of this letter in the Standard.
WG2 discussion
See also comments T1 from US and T4 from Germany.

T3. Page 1211, Row 109A: Meroitic Cursive
Because of the appearance of a recent article by Jochen Hallof in Beiträge zur Sudanforschung vol. 10 (2009), entitled "Ein meroitisches Zahlenostrakon aus Qasr Ibrim", presenting Meroitic numbers from 1 up to 900,000, Ireland requests the removal of the Meroitic fractions and numbers of the characters 109C0..109F0 from the FCD pending further study. We hope to have this study complete before the Helsinki meeting.
WG2 discussion
These characters should be removed unless satisfactory evidence is provided during the Helsinki meeting.

Editorial comments

E1. Page 124, Row 098: Bengali
Ireland recommends that the rather inconsistent font used for Bengali be replaced with the one given in the table shown below.
Proposed acceptance
However, the table should be reviewed by experts before approval, especially concerning these points:
- *Horizontal position of combining marks,*
- *Significant glyph changes (e.g. 09C2, 09C4)*
- *Relative size of Bengali-specific additions (09F0-09FB) to the rest*

E2. Page 259, Row 20D: Combining Diacritical Marks for Symbols
Ireland recommends that the erroneous dotted circles in 20E4, 20E5, 20E6, 20E7, and 20E8 be corrected.
Proposed acceptance

E3. Page 1080, Row AAE: Meetei Mayek Extensions
Ireland suggests that the chart glyph for AAF6 should be harmonized with that for 1039, 1A60, and 10A3F.
Proposed acceptance

E4. Page 1223, Row 1110: Chakma
Ireland suggests that the chart glyph for 11133 should be harmonized with that for 1039, 1A60, and 10A3F.
Proposed acceptance

E5. Page 1320, Row 1F0A: Playing Cards
Ireland recommends that the font used for Playing Cards be replaced with the one given in the table shown below. The shapes of the suits and the face cards are more distinct in the new font, and are more recognizable.
Proposed acceptance

Japan, Negative

JP.1 (Editorial): Foreword, last paragraph
"ISO/IEC 10646:2010" is wrong.
Proposed change by Japan
"ISO/IEC 10646:2011"
Accepted

JP.2 (Editorial): Sub-clause 4.4, Note
"(see 0)" is wrong.
Proposed change by Japan
"(see 4.23)"
Accepted

JP.3 (Editorial): Sub-clause 9.1, Table 2
On the third row, the first letter "z" is mistakenly capitalized.
Proposed change by Japan
"zzzzyyyyyyxxxxxx"
Accepted

JP.4 (Editorial): Sub-clause 9.2, Table 4
On the first row, the first letter "x" is mistakenly capitalized.
Proposed change by Japan
"xxxxxxxxxxxxxxxx"
Accepted

JP.5 (Editorial): Sub-clause 11, Note 2 (end of)
"ISO/IEC 6492" is wrong.
Proposed change by Japan
"ISO/IEC 6429"
Accepted

JP.6 (Technical): Sub-clause 16.5, Variations selectors and variation sequence, Note 2
This note explicitly cites the "version 2007-12-14" of IVD. However, the UTS#37, normatively referred to by 10646, designates the IVD as "http://www.unicode.org/ivd", with no version indication, that always refers to the latest version of IVD. So the IVSes allowed in 10646 is those listed in the latest version of the IVD.
Proposed change by Japan
Change " version 2007-12-14" to "version 2010-11-14" in the text. (URL is correct.).
Accepted
See also comment T1 from UK.

JP.7 (General): Sub-clause 23.5 and 31.3, Code charts for CJK Compatibility Ideographs and their format
The new code charts for CJK Compatibility Ideographs have problems: (a) Representative glyphs for KP-source compatibility ideographs seem not from the corresponding KP source standards; (b) Each cell now shows the corresponding unified ideograph let by "≡" sign with a representative glyph, and Japan considers it is not a good idea to show only one representative glyph (especially that from a different source group than the compatibility ideograph is from.)
Proposed change by Japan
Revert the code charts and the explanation of the code chart format, i.e., use the single column format as before. Otherwise, update the code charts appropriately (under the new format.)
WG2 discussion
Concerning item a) glyph differences from KP source standard, the new chart uses the same DPRK source font that was used to represent glyph in the single column format, so if they are errors, exact values should be provided.

Concerning item b) the single column format has the same issue, so reverting to the old format would not improve the situation.
The editor is open to suggestion in how to describe the unified ideograph (at the right of the "≡" sign) if other NBs see the need. It is also possible to create tables to pick another glyph for the 'unified' ideographs, although that would be an issue for compatibility ideographs that have multiple sources.
Overall, most experts see the new format as a significant improvement and because all issues raised by Japan already exist in the previous format, there seems no need to revert back to it.

JP.8 (General): Sub-clause 31.3, Name lists for CJK compatibility ideographs

The new name lists for CJK compatibility ideographs have problems: (a) Some classification (grouping) doesn't make sense and confuse readers rather than help understanding (e.g., J3-763B is classified under "Pronunciation variants from KS X 1001:1998" but it is absolutely not); (b) Many of additional information led by "→" seem inappropriate and don't make sense..
Proposed change by Japan
Revise the list accordingly
WG2 discussion
The grouping (item a) is the same as the single column format, so this is not a new issue. The fact is that the new multicolumn format reveals some limitation of the grouping. The issue mentioned by Japan concerning J3-763B arose because Japan recycled a KS X entry for its own usage, and because J sources are ordered before K sources, the J sources appear first.
The grouping issue could be solved by either segmenting the groups when multiple sources exist or just adding notes explaining that the groups describe major category (such as KSX 1001:1998 for F900-FA0B), but that they may contain characters from other sources not related to these categories.
Concerning b), the additional information pointed to by (such as for F9B8 and FAD4) are appropriate and describes original mapping errors that cannot be fixed because of mapping immutability. The editor is open to suggestion to better describe the issues if current text does not make sense to some NBs.

JP.9 (General): Sub- clause 31.3 Code chart for CJK Ideographs Extension C and D

Japan understood that those code charts are exactly as in their corresponding amendments, but they are revised unexpectedly in the 2nd edition. Japan considers those in the amendments are better. Moreover, [more text missing?].
Proposed change by Japan
Revert the change
Propose non acceptance
There should not be any difference between the 2nd and 3rd edition. Concerning Extension C, the production tool changed significantly between the Amendment that added it (Amd5) and these new editions. This has introduced some errors such as 2AAC9 that have been captured (ref SC2 N4176) and will be fixed in this edition. In addition, Taiwan also went through a revision of its Ext C source characters. All this to say that these blocks are not static and can evolve after an amendment is adopted. Furthermore, it is not possible to revert to the original Amendment 5 production environment.
Concerning ext D, the editor is not aware of any issue but will fix any if made aware of specifics.

JP.10 (General): Sub- clause 31.3 Code chart for CJK Ideographs Extension B

Reflect the review report by IRG
Proposed change by Japan
Update accordingly
Propose acceptance

JP.11 (Editorial): Sub-clause A.4.3, Last sentence

"See 0" is wrong.
Proposed change
"See 23.1"
Accepted

JP.12 (Editorial): Sub-clause A.5.5, First bullet

"See 023.1" is wrong.
Proposed change
"See 23.1"
Accepted

JP.13 (Editorial): Annex M, Glagolitic

Title of ISO 6861 is wrong; it lacks a word "alphabet".
Proposed change
Insert "alphabet" as "Glagolitic alphabet coded"
Accepted

JP.14 (Editorial): Annex M, Latin

Title of ISO 6937 is wrong; a singular word "set" is mistakenly spelled as "sets".
Proposed change
Change "sets" to "set" as "Coded graphic character set"
Accepted

JP.15 (Editorial): Annex I.1.2, Row for "SURROUND FROM BELOW" in Table I.1

The "IDS examples represents" column for the row is inappropriate, because it shows 土 inside while the corresponding IDS has 士 as D2.
Proposed change
Use a correct picture as in the 2003 edition.
Accepted in principle

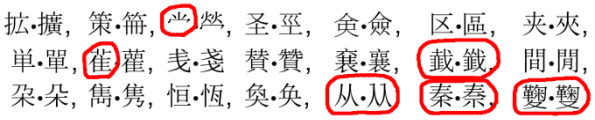
This edition code characters for this table, but the example will use 2067D 𠔼 instead of 51F7 凶. This will be then equivalent to the example shown in the 2003 edition.

JP.16 (General): Clause S.1.4.3, Chart of examples

(The same comment has been submitted as a part of the ballot comment against FDIS 10646 2nd Ed.)

Project editor introduced several issues by his changes to Annex S examples. IRG discussed on many of them and reverted most of the problematic changes. However, changes on S.1.4.3 were not found until recently.

The circled examples in the following chart taken from the current draft have problems because they don't make any sense or they change the intension of examples:



when compared to the corresponding examples in the 2003 edition of ISO/IEC 10646:

The first two circled examples in the current draft don't make any sense, because there is no chance that one thinks two ideographs with the components are subject to unification; while corresponding examples in 2003 edition are

helpful because many experienced users of ideographs may misunderstand two ideographs with the components can be unified.

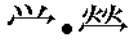
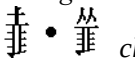
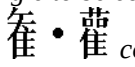
The last four circled examples in the current draft are valid; they show misleading cases. However, the examples in the current draft have less value than those in the 2003 edition, because they appear the cases are only applicable to the specific combinations of components as shown, while in 2003 edition the corresponding examples covered any patterns combined with the shown components.

Proposed change
Revert the examples in S.1.4.3 to those in 2003 edition.

Accepted in principle

The intent of the update was to update the example content from picture based to font based to provide a better quality document. Because sometimes, examples are not actual ideograph but fragment that are not separately encoded, this has proved challenging. As in the sub-clause S.1.5, some of the examples will be reversed back to pictures when no good alternative exists.

The examples from the 2003 edition for which there are no font based solution will be reverted to pictures. This will address all 'circled' cases. Worth noting:

1) *The pictures*  *and*  *clearly represents fragments and should have an accompanying dotted rectangle to be consistent with other examples.*
2) *The picture*  *could be represented using a font, only if the editor could find the coded character for the first glyph (if any). Apparently, it is only encoded as component; in such case a dotted rectangle should be added as well.*

JP.17 (Technical): Sub-clause S.1.6, G-Source

The list of G-source standards is updated from 2nd Ed. by adding two new ones: GB 12052-90 and GB 15564-1995. Japan believes adding them here is wrong, since this particular subclause is primarily a record of what we did when we created the very first CJK UNIFIED IDEOGRAPHS block and we didn't see these two standards.

Proposed change
Remove "GB 12052-90" and "GB 15564-1995".

Propose non acceptance

The problem is that the information collected in the previous text is incorrect. Some of the G sources characters were not in any of the original sources. These 2 sources are required to correctly describe all G sources. A note could be added to explain this.

Korea (ROK): Negative

R.O.Korea will change its vote to "Yes" if the following request is accepted.

Technical comments:

T1 Annotations for U11xx and U31xx

- Rep. of Korea suggests that annotations for U11xx and U31xx be added.
- The list of annotations are in Annexes 1 and 2.

Annex 1. 30 annotations for U11xx characters.

1109 ᄉ HANGUL CHOSEONG SIOS
• voiceless lenis alveolar fricative consonant
110A ᄊ HANGUL CHOSEONG SSANGSIOS
• voiceless unaspirated glottalized alveolar fricative consonant
110B ᄋ HANGUL CHOSEONG IEUNG
• zero sound
110C ᄌ HANGUL CHOSEONG CIEUC
• voiceless or voiced lenis alveolar affricate consonant
110D ᄍ HANGUL CHOSEONG SSANGCIEUC
• voiceless unaspirated glottalized alveolar affricate consonant
110E ᄎ HANGUL CHOSEONG CHIEUCH
• voiceless aspirated alveolar affricate consonant
1112 ᄒ HANGUL CHOSEONG HIEUH
• voiceless glottal fricative consonant
113C ᆺ HANGUL CHOSEONG CHITUEUMSIOS
• voiceless lenis dental fricative consonant
113D ᆻ HANGUL CHOSEONG CHITUEUMSSANGSIOS
• voiceless glottalized dental fricative consonant
113E ᆹ HANGUL CHOSEONG CEONGCHIEUMSIOS
• voiceless lenis retroflex fricative consonant
113F ᆾ HANGUL CHOSEONG CEONGCHIEUMSSANGSIOS
• voiceless glottalized retroflex fricative consonant
1140 ᅀ HANGUL CHOSEONG PANSIOS
• voiced alveolar fricative consonant
114C ᅌ HANGUL CHOSEONG YESIEUNG
• velar nasal consonant
114E ᆽ HANGUL CHOSEONG CHITUEUMCIEUC
• voiceless or voiced lenis dental affricate consonant
114F ᆿ HANGUL CHOSEONG CHITUEUMSSANGCIEUC
• voiceless unaspirated glottalized dental affricate consonant
1150 ᇀ HANGUL CHOSEONG CEONGCHIEUMCIEUC
• voiceless or voiced lenis retroflex affricate consonant
1151 ᇁ HANGUL CHOSEONG CEONGCHIEUMSSANGCIEUC
• voiceless unaspirated glottalized retroflex affricate consonant
1154 ᇇ HANGUL CHOSEONG CHITUEUMCHIEUCH
• voiceless aspirated dental affricate consonant
1155 ᇈ HANGUL CHOSEONG CEONGCHIEUMCHIEUCH
• voiceless aspirated retroflex affricate consonant
1159 ᅙ HANGUL CHOSEONG YEORINHIEUH
• glottal stop consonant
119E ᆞ HANGUL JUNGSEONG ARAEA

• rounded open-mid back vowel
11BA ᆺ HANGUL JONGSEONG SIOS
• voiceless lenis alveolar fricative consonant
11BB ᆻ HANGUL JONGSEONG SSANGSIOS
• voiceless unaspirated glottalized alveolar fricative consonant
11BC ᆼ HANGUL JONGSEONG IEUNG
• velar nasal consonant
11BD ᆽ HANGUL JONGSEONG CIEUC
• voiceless or voiced lenis alveolar affricate consonant
11BE ᆾ HANGUL JONGSEONG CHIEUCH
• voiceless aspirated alveolar affricate consonant
11C2 ᇂ HANGUL JONGSEONG HIEUH
• voiceless glottal fricative consonant
11EB ᇫ HANGUL JONGSEONG PANSIOS
• voiced alveolar fricative consonant
11F0 ᇰ HANGUL JONGSEONG YESIEUNG
• velar nasal consonant
11F9 ᇹ HANGUL JONGSEONG YEORINHIEUH
• glottal stop consonant

Annex 2. 6 annotations for U31xx characters.
3147 ㅇ HANGUL LETTER IEUNG
• zero sound as initial or velar nasal consonant as final
314E ㅎ HANGUL LETTER HIEUH
• voiceless glottal fricative consonant
317F ㅿ HANGUL LETTER PANSIOS
• voiced alveolar fricative consonant
3181 ㆁ HANGUL LETTER YESIEUNG
• velar nasal consonant
3186 ㆆ HANGUL LETTER YEORINHIEUH
• glottal stop consonant
318D ㆍ HANGUL LETTER ARAEA
• rounded open-mid back vowel
* Note. Currently U3181 and U3186 has annotations. R.O.Korea suggests that the current annotations be replaced by the ones suggested above.

3181 ㆁ HANGUL LETTER YESIEUNG
• old velar nasal
3186 ㆆ HANGUL LETTER YEORINHIEUH
• old glottal stop

WG2 discussion

Similar requests have been made in previous ballots concerning the 2nd and 3rd edition which were all withdrawn during comment disposition. These requests were trying to add annotation for all 50 Hangul Compatibility Jamos (3131-318E). This is now replaced by annotation requests for:

a) *30 Jamos (in area 1100-11FF), including 20 initial consonants (Choseong), 1 medial vowel (Jungseong), and 9 final consonants (Jongseong) and*
b) *6 Hangul Compatibility Jamos in area 3130-318F)*

Further study shows that the annotations for the 9 final consonants are identical to the related initial consonants with one exception: 110B HANGUL CHOSEONG IEUNG (zero sound) versus 11BC HANGUL JONGSEONG IEUNG (velar nasal consonant). The 6 Hangul Compatibility Jamos are again a subset of the consonant/vowel annotated in the 11xx block. It is not clear why this subset was chosen, and why it is not just sufficient to annotate the 20 initial consonants, the one medial vowel, and the one final consonant which seems to be the exception to the 'identical' annotation rule (namely IEUNG).

The editor is not opposed at adding annotation, but they should be some rationale concerning the chosen subset, furthermore annotation should not repeated as much as possible.

Finally, annotations are all editorial in nature, so this should not considered a technical comment.

United Kingdom: Positive with comments

Editorial comments

E.1. Sub-clause 16.5

"This International Standard incorporates by reference the variation sequences listed in version 2007-12-14 of the Ideographic Variation Database"
The version referenced should be "2010-11-14"
Proposed change:
Change to:
"This International Standard incorporates by reference the variation sequences listed in version 2010-11-14 of the Ideographic Variation Database".
Accepted
See also comment JP6 from Japan.

E.2. Sub-clause 23.1

The G-source "ZhongHua ZiHai" is incorrectly capitalized.
Proposed change:
Capitalize as "Zhonghua Zihai"
Accepted

E.3. Sub-clause 23.2, 23.4

"2nd field: Radical-Stroke index (d{1,3}'.d{1,2}). This informative field contains radical index (one to three digits), optionally followed by an apostrophe for alternate index, followed by a full stop, and ending by one or two digits for the stroke count. NOTE 2 – All ideographs are classified following radical/stroke indexes in various CJK dictionaries. The primary value provided in this field is the most common one, while alternate indexes provide variant values also in use. More information is available in the Unicode Standard UAX#38 Unicode Han Database at http://www.unicode.org/reports/tr38/."
The description of the Radical-Stroke index in 23.2 and 23.4 is confusing. Firstly, the optional apostrophe following the radical index indicates a simplified form of the radical, not an "alternate index". Secondly, in 23.2 Note 2 it is not clear what the "alternate indexes" refer to; it could be read to mean that multiple radical-stroke indexes are provided in this field (a primary index and alternate indexes), although only a single radical-stroke index is given for each character.
Proposed change:
In 23.2 change to:
"2nd field: Radical-Stroke index (d{1,3}'.d{1,2}). This informative field contains a radical index (one to three digits), optionally followed by an apostrophe for simplified radicals, followed by a full stop, and ending with one or two digits for the stroke count. NOTE 2 – All ideographs are classified following radical/stroke indexes in various CJK dictionaries. The value provided in this field is the most common one. More information is available in the Unicode Standard UAX#38 Unicode Han Database at http://www.unicode.org/reports/tr38/."
In 23.4 change to:
"3rd field: Radical-Stroke index (d{1,3}'.d{1,2}). This informative field contains a radical index (one to three digits), optionally followed by an apostrophe for simplified radicals, followed by a full stop, and ending with one or two digits for the stroke count."
Accepted

E.4. Clause 29

"The TIP (plane 03) is used for ancient ideographic scripts that are related but not classified as CJK unified ideographs. No character is encoded in the TIP."
The grammar of these two sentences is awkward.
Proposed change:
Change to:
"The TIP (plane 03) is used for ancient ideographic scripts that are related to but not classified as CJK unified ideographs. No characters are currently encoded in the TIP."
Accepted

E.5. Clause 30

"The SSP (plane 0E) is used for special purpose use graphic characters."
Tag characters in the SSP are format characters not graphic characters.
Proposed change:
Change to:
"The SSP (plane 0E) is used for special purpose use graphic characters and format characters."
Accepted

E.6. Sub-clause S.1.4.3

Some of the examples illustrating different structure of corresponding ideographic components are incorrect or suboptimal compared with the examples given in the 2003 edition of the standard. In particular the 3rd and 9th examples show pairs of unrelated components that could not be mistaken as being subject to unification.
Proposed change:
Revert to the examples given in ISO/IEC 10646:2003 S.1.4.3.
Accepted in principle
See disposition of comment JP16 from Japan.

USA: Negative with comments

Technical comments:

T.1. Latin Extended-D

We reiterate that this character is unnecessary and is a damaging duplication for the standard and should be removed from the amendment. The change of name to “GLOTTAL DOT” only introduced additional problems, for it specifically mentions the function of the character, and opens the door for potential future requests for "non-glottal" middle dot letters as well.

Justification for the request to remove this character is contained in N3678 (L2/09‐278). A viable alternative to encoding a separate letter middle dot, for the purposes cited by the original proposal, would be to use the already encoded modifier letter, U+02D1 MODIFIER LETTER HALF TRIANGULAR COLON.

Proposed change:

The U.S. objects to the renaming of LATIN LETTER MIDDLE DOT from the original amendment to U+A78F LATIN LETTER GLOTTAL DOT, as it is a middle dot, and requests the removal of this character.

WG2 discussion

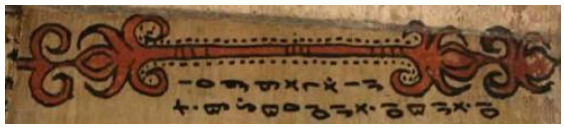
See comments T4 from Germany and T2 from Ireland

T.2. Batak

The evidence provided in N3320 does not demonstrate these marks to be characters, but rather to be graphic page elements that do not behave at all as characters.

Compare the following two figures from N3320.

U+ 1BFA BATAK SYMBOL BINDU GODANG in figure 11:



U+1BFB BATAK SYMBOL BINDU PINARJOLM in figure 13 :

In the above examples, it is clear that the symbols are large ornaments, with text wrapping underneath them. In our assessment such elements should be represented by images, textual markup, or other mechanisms rather than being encoded as single characters.

Proposed change:

The U.S. requests the removal of the following two characters:

1BFA BATAK SYMBOL BINDU GODANG

1BFB BATAK SYMBOL BINDU PINARJOLM

WG2 discussion

See also comments T1 from Ireland and T2 from Germany.

T.3. Optical Character Recognition

The names for U+2448 and U+2449 are Magnetic Ink Character Recognition (MICR) symbols used by banks on checks. The two characters U+2448 and U+2449 originally encoded in 10646-1:1993 were inadvertently mixed up; as a result their current names are misleading about their identity. The formal aliases provide the correct names. Additional corroboration is provided on the following websites: http://www.barcodesoft.com/e13bmapping.htm http://en.wikipedia.org/wiki/Magnetic_ink_character_recognition http://www.printerm.com/fonts2C.htm http://mindprod.com/jgloss/micr.html

Proposed change:

The U.S. asks two formal name aliases be added as listed below:

a. U+2448 OCR DASH

Add the following formal alias: 2448 MICR ON US SYMBOL
b. U+2449 OCR CUSTOMER ACCOUNT NUMBER
Add the following formal alias: 2449 MICR DASH SYMBOL
Propose acceptance

T.4. NUSI.txt

Adding named sequences for 2 symbols from ISO/IEC 9995-7 is premature, pending a complete analysis of the encoding requirements for the repertoire of symbols from that keyboard standard.
Proposed change:
The U.S. requests the removal of the following two named sequences, as requested in N3897:
U+21F3 U+20E2 KEYBOARD SYMBOL SCROLLING
U+2139 U+20E2 KEYBOARD SYMBOL HELP
Propose acceptance

Editorial comments:

E.1. CJK Fonts

The U.S. is aware of a large number of glyph errors in Extension B fonts that have been discovered in the course of IRG and national body review. These glyph errors must be fixed before publication of the third edition.
WG2 discussion
See comment JP10 from Japan

E.2. Arrows

The proposed glyph changes, as proposed in N3987, need further review, as part of the full analysis of the encoding requirements for the repertoire of symbols from the ISO 9995-7 keyboard standard.
Proposed change:
The U.S. requests the glyph changes for U+21E6 - U+21F0 and U+21F3 in the Arrows block, as included in the current ballot, be reverted. While the U.S. is open to improvements to glyphs, the new glyphs are out of sync with the set of arrows at U+2B00-U+2B04 in the Miscellaneous Symbols and Arrows block.
Propose acceptance