L2/05-038

ISO/IEC JTC1/SC2/WG2 N 2926

Date: 2005-01-26

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

Doc. Type:	**Disposition of comments**
Title:	**Disposition of comments on SC2 N 3760 (FPDAM text for Amendment 1 to ISO/IEC 10646:2003)**
Source:	**Michel Suignard (project editor)**
Project:	JTC1 02.10646.00.01
Status:	For review by WG2
Date:	2005-01-26
Distribution:	WG2
Reference:	SC2 N3760, N3779, WG2 N2744, N2840, N2859, N2862, N2863, N2864, N2877, IRG N1095
Medium:	Paper, PDF file

Comments were received from Canada, China, France, Iran, Ireland, Israel (O), Japan, Morocco and USA. The following document is the disposition of those comments. The disposition is organized per country.

As a result of the dispositions there are no remaining negative votes.

Note – The full content of the ballot comments (minus some character glyphs) 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*.

Canada: Yes with comments:

Technical Comments:

Comment 1: Tifinagh name changes
Change the names of the following Tifinagh characters (Rationale is provided in document JTC1/SC2/WG2 N2862 from Morocco, Canada and France – P. Andries).
2D34 TIFINAGH LETTER YAGHH
(ending with YAGHH instead of YAGGH)
2D35 TIFINAGH LETTER BERBER ACADEMY YAJ
(replacing KABYLE in current name with BERBER ACADEMY)
2D3F TIFINAGH LETTER YAKHH
(ending with YAKHH instead of YAKKH)
2D41 TIFINAGH LETTER BERBER ACADEMY YAH
(replacing KABYLE in current name with BERBER ACADEMY)
Accepted
Similar request from France, Ireland (T.6a) and Morocco.

Comment 2a: Tifinagh name annotation
Add annotations for the following Tifinagh character (see N2862):
2D53 TIFINAGH LETTER YU (Tuareg yaw)
Accepted
Similar request from France and Morocco

Comment 2b: Tifinagh name annotation
Add annotations for the following Tifinagh character (see N2862):
2D6F TIFINAGH MODIFIER LETTER LABIALIZATION (tamatart)
Accepted in principle
Similar request from France and Morocco
Comment T.6b from Ireland was requesting a name change for 2D6F (LABIALIZATION changed to TAMATART). The adopted change is as following:
2D6F TIFINAGH MODIFIER LETTER LABIALIZATION MARK (tamatart)

Comment 3: Tifinagh characters to be added to annex E.2
Add the following Tifinagh characters to the list of mirrored characters in Annex E.2 (Rationale is provided in document WG2 N2863 from Canada – P. Andries):
Note: The new names proposed above are used in the list below. 2D4F and 2D50 have left to right symmetrical glyphs in the charts in FPDAM1. Rationale for including them here is also noted.
2D39 TIFINAGH LETTER YADD
2D3A TIFINAGH LETTER YADDH
2D3E TIFINAGH LETTER TUAREG YAK
2D41 TIFINAGH LETTER BERBER ACADEMY YAH
2D49 TIFINAGH LETTER YI
2D4D TIFINAGH LETTER YAL
2D4E TIFINAGH LETTER YAM
2D4F TIFINAGH LETTER YAN
(Note: YAN has a symmetrical glyph in the charts; however, it may be slanted in the Tuareg variants to distinguish two consecutives YAN from a YAL, see section D.9 of WG2 document N2739.)
2D50 TIFINAGH LETTER TUAREG YAGN
(Note: YAGN has a symmetric glyph in the charts – however, it is slanted in some Tuareg variants, see table on page 8 of WG2 document N2739.)
2D5B TIFINAGH LETTER YASH
2D5E TIFINAGH LETTER YACH
2D5F TIFINAGH LETTER YATT
2D62 TIFINAGH LETTER YAY
2D64 TIFINAGH LETTER TAWELLEMET YAZ
Accepted in principle
Also discussed in the context of document WG2 N2877: "Possible Defect in Annex E.2 – Other Mirrored Characters". Because the mirroring behavior of characters in scripts such as Old Italic, Tifinagh is largely dependent on the exact shapes used in a font (i.e. some font designer may make a glyph symmetrical in one style and asymmetrical in another one), it is unwise to provide an exact list.

Therefore it was agreed to modify the sub-clause E.2 by removing the list of characters and rephrasing the current introductory text as follows:
E.2: Ambidirectional Scripts

Many ancient scripts and some scripts in modern use can be written either right-to-left or left-to-right. Often it is customary for one of these scripts to use the appropriately mirrored graphical symbol for any character represented by a graphic symbol that is not symmetric around the vertical axis. In such cases, it is up to the rendering system to display the graphic image appropriate for the writing direction employed. The directionality of the representative graphic symbol shown in the character code charts matches the default writing direction for the script.

Examples of such scripts include, but are not limited to, Old Italic, an ancient script for which the default writing direction in this standard is left-to-right, and Tifinagh, a modern script for which the default writing direction in this standard is right-to-left.

Comment 4: remove Coptic characters

The following Coptic characters are all abbreviations of two or more letters. WG2 has rejected similar proposals for abbreviations in other scripts in the past. Other than stating that these are abbreviations, sufficient evidence that these should be encoded as separate characters has not been presented in WG2's document trail. These characters should be removed from the current FPDAM1 set of Coptic characters till more justification is provided.

2CE6 COPTIC SYMBOL PI RO.

N2744 states "*This symbol is used as a standard abbreviation for 'pros' /to/.*"

2CE7 COPTIC SYMBOL STAUROS.

N2744 states "*This symbol represents the word 'stauros' /cross/ in Bohairic texts.*"

2CE8 COPTIC SYMBOL TAU RO

N2744 states "*This symbol is used as a standard abbreviation for 'taur' /taur/ and "stauros" /cross/.*" The document at http://www.unicode.org/L2/L2004/04053-coptic.pdf claims that there is a distinct but similar looking symbol representing the number 900; the latter could be encoded as such, the abbreviation should not.

2CEA COPTIC SYMBOL SHIMA SIMA.

N2744 states "*This symbol is used as a standard abbreviation for 'côis' /lord/.*"

Not accepted

According to the result of the Ad hoc meeting, enough evidences were produced concerning the characters 2CE6, 2CE7, and 2CE8 to convince the Canadian experts to encode these three characters. However no consensus was reached for 2CEA COPTIC SYMBOL SHIMA SIMA. The contributing editors will consider the suggested glyph change for that character. The full Canadian report on the ad hoc meeting concerning this point follows:

The issue of the proposed encoding of 4 abbreviations (2CE6, 2CE7, 2CE8 and 2CEA) opposed by Canada was discussed in an ad hoc meeting. Contrary to 2CE4, 2CE5 and 2CE9, for which evidence of existence as symbols was present in the proposal, no such evidence was present for 2CE6, 2CE7, 2CE8 and 2CEA. The meeting produced enough evidence for 2CE6, 2CE7 and 2CE8 to be encoded as symbols, but Canada remains unconvinced that 2CEA COPTIC SYMBOL SHIMA SIMA is anything but a normal contraction. Canada would have been convinced that 2CEA was indeed a symbol if evidence had been given of SHIMA SIMA contractions which are systematically not ligated (viz. the SHIMA is not flattened and linked to the SIMA). Should this 2CEA be encoded, the reference glyph should at least be one with a bar above, which is by far the most common form for the */cois/* abbreviation. Evidence to that effect was shown from the set of fonts called the Coptic Font Standard (http://www.copticchurch.net/coptic_fonts/), used for standardized interchange prior to UCS standardization of Coptic; no contrary evidence was supplied. Similar usage is found in most textbooks Canada has seen.

Comment 5: Case-paired characters in Coptic

The amendment proposes to encode case variants (uppercase and lowercase) for the whole Coptic alphabet, despite the fact that the script has only fairly recently acquired casing behaviour and that some Old Coptic characters have never had this behaviour. Encoding case variants of these case-less Old Coptic letters therefore amounts to ISO 10646 inventing some new characters. The justification offered in document N2744 does not warrant this. Part of the justification is "In order to facilitate modern casing operations...", which amounts to nothing since casing operations are based on character property files which list case properties for each character. Another part is a scenario (cast in the conditional by an explicit "would be") of an ALL CAPS title that would somehow use case-less letters. Such a hypothetical scenario does not justify encoding non-existent characters. It looks like what could be solved by proper selection of fonts is being forced into the encoding of these case-less characters as artificially cased characters! We request that the artificial case variants of the actually case-less subset of Coptic be removed.

Not accepted
As mentioned below in the Canadian report on the ad hoc meeting, the existence of casing pairs is justified by the current encoding model used for Coptic. The full Canadian report on the ad hoc meeting concerning this point follows:

The issue of encoding Coptic as a cased script was discussed in an ad hoc meeting. It was argued that the application of casing pairs to Coptic letters is not a question of inventing letters which do not exist; it is an application of systemic logic based on the encoding model selected, which model is not made explicit anywhere but in its results. Canada does not consider such a model sufficient grounds for encoding non-existent characters and maintains its objection to encoding case counterparts to old Coptic characters when they have never existed. This model does not seem suited to the actual characters found in Coptic and should be reviewed.

China: Yes with comments

Technical Comments:

Comment 1: U+23DA (EARTH GROUND)
The shape of U+23DA (EARTH GROUND) in the "Miscellaneous Technical" block is not the original that submitted (H-889A) by Hong Kong SAR. The shape adopted appears to be with one more horizontal line at the bottom. It is requested to keep the original shape like what submitted before. See picture below please.

Accepted in principle
The contributing editors will work with China to come up with a new glyph compatible with both the HKSCS original design and design typically used to represent the EARTH GROUND symbol. For reference, the current glyph is as following:

Comment 2.1: New Tai Lue font
It is strongly required to use fonts provided by China, as agreed at ISO/IEC JTC1/SC2/WG2#45.
Accepted in principle
The font had not been yet provided to the contributing editor. Assuming it satisfies common requirement for chart production purpose, it will be used for the amendment publication.

Comment 2.2: New Tai Lue 19DF NEW TAI LUE SIGN LEW
The name of U+19DF should be LAEV rather than LEW. It is consonant L + Vowel AE (U+19B6) + FINAL V (U+19C1).
Accepted
However to maintain consistency the name for U+19DE will be changed as well, resulting in the following two characters changed from:
19DE NEW TAI LUE SIGN LE
19DF NEW TAI LUE SIGN LEW
To:
19DE NEW TAI LUE SIGN LAE
19DF NEW TAI LUE SIGN LAEV

Comment 2.3: New Tai Lue script name
It is requested to USE Dai rather than Tai.
Not accepted
While 'Dai' is the common pinyin transliteration for the script, the common English transliteration is 'Tai'. English based transliteration is already used for the names for that script as well as for other scripts such as 'Tai Le'. Changing the script name to 'Dai' will bring inconsistency in the standard.

Comment 3: IICORE
The IICORE character list and source information should be updated according to IRG#23 resolutions.
Accepted in principle
A cursory check by the editor seems to indicate that the collection content is unchanged, however this should be confirmed by another source..

France: Positive with comments

Technical Comments:
France supports document SC 2 N 2862 as contribution to our approval of ISO/IEC 10646:2003 FPDAM. 1
Accepted in principle
See disposition of Canadian comments 1, 2a and 2b.

Iran, Islamic Republic of: Positive with comments

Technical Comments:
ISIRI believes that a font consistent with the current font used for the Arabic block (U+0600..U+06FF) should be used for rendering character tables for the new Arabic Supplement block (U+-0750..U+077F). Considering ISIRI's experience with the Arabic script languages, ISIRI volunteers to design such a font for this or a future revision of ISO/IEC 10646 and the Unicode Standard, according to the time requirements of the committees, provided a copy of the font already used for the Arabic block is provided to ISIRI as a reference.
Noted
The contributing editors are welcoming contributions to improve consistency of glyphs used within the Arabic blocks.

Ireland: Negative

General Comments
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. In addition, it should be noted that Ireland confirms its strong support for the [New] Tai Lue character set as it appears in the ballot, and we do not favour its alteration or deletion.
Noted
The New Tai Lue character set was not substantially changed by the resolutions which only resulted in a couple of name changes (see disposition of Chinese comment 2.2).

Technical comments

T.1. Page 1358, Annex B.1 List of all combining characters. With reference to ISO/IEC JTC1/SC2/WG2 N2840 "Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS", Ireland, which requested the addition (to PDAM-2) of HEBREW POINT HOLAM HASER FOR VAV at position U+05BA, requests that HEBREW POINT QAMATS QATAN here be moved from U+05BA to position U+05C7.
Accepted
Similar to US comment T.11.

T.2. Page 1393, Annex P, Additional information on characters. As with T1, 05BA here should be changed to 05C7, with an appropriate change to the first sentence in italics here.
Accepted
Similar to US comment T.11.

T.3. Table 6 - Rows 01-02: Latin Extended-B. With reference to ISO/IEC JTC1/SC2/WG2 N2860 "Revised Proposal to Encode Orthographic Glottal Stops in the UCS", Ireland requests the addition at 0242 of the character x-height LATIN SMALL LETTER GLOTTAL STOP as the casing partner of 0241 LATIN CAPITAL LETTER GLOTTAL STOP. It is clear that the capital and small glottal stop act in casing relation

ship to one another in a natural orthography, which may itself coincide in plain text with a phonetic representation of Athabaskan languages which makes use of the neutral 0294 LATIN LETTER GLOTTAL STOP. If this letter is not added, we request the removal of the capital letter pending further study.
Not accepted
Neither the removal of 0241 LATIN CAPITAL LETTER GLOTTAL STOP or the addition of 0242LATIN SMALL LETTER GLOTTAL STOP were accepted. There is prevailing view that, should a smaller form of glottal stop be required, the existing 0294 LATIN LETTER GLOTTAL STOP *would suffice. The issue may still be entertained in the context of the 2nd amendment.*

T.4. Hebrew Characters. With reference to ISO/IEC JTC1/SC2/WG2 N2840 "Proposal to add HEBREW POINT HOLAM HASER FOR VAV to the BMP of the UCS", Ireland has requested the addition of HEBREW POINT HOLAM HASER FOR VAV at position U+05BA, which would entail moving HEBREW POINT QAMATS QATAN (under ballot in here) to position U+05C7.
Withdrawn

T.5. Table 22 - Row 09: Bengali. With reference to the examples given at the end of this document, Ireland requests the addition of the characters BENGALI DANDA at position U+09E4 and BENGALI DOUBLE DANDA at position U+09E5. Addition of these characters to PDAM-2 would satisfy us.
Withdrawn

T.6.a Table 101 - Row 2D: Tifinagh. Ireland requests the following name changes: 2D34 change from "YAGGH" to "YAGHH", 2D35 change from "KABYLE" to "BERBER ACADEMY", 2D3F change from "YAKKH" to "YAKHH", 2D41 change from "KABYLE" to "BERBER ACADEMY"….
Accepted
Similar to comment 1 from Canada and comments from France and Morocco

T.6.b Table 101 - Row 2D: Tifinagh. …Ireland also requests the following name change: 2D6F change from "LABIALIZATION" to "TAMATART" on the grounds that "labialization" is a function and "TAMATART" is an actual name.
Accepted in principle
See disposition of Canadian comment 2.b.

T.6.c Table 101 - Row 2D: Tifinagh. …Further, Ireland requests the deletion of the parenthetical notes "(Tuareg four-points)" and "(Tuareg staggered five-points)" on several grounds. The phraseology "four-points" and "five-points" is not an English nominal construction. The term is a graphic description of the character, which could as well be applied to many of the other characters; 2D42 also has four dots in it. "Staggered" does not accurately describe the dot positioning in 2D58. We do not favour other name changes to this set.
Accepted

T.7. Table 115 - Row 31: CJK Basic Strokes. With reference to ISO/IEC JTC1/SC2/WG2 N2864 "Proposal to add a block of CJK Basic Strokes to the UCS", Ireland requests the complete replacement of Table 115 in this FPDAM with the table on the last page of N2864. If necessary this table should be deleted from the FPDAM and added to PDAM-2 if it is considered that further evaluation be considered for this character set. Ireland believes that the table as it is should not be accepted as part of the standard given the strong arguments in N2864.
Withdrawn

Editorial comments

E.1 Page 21, Sub-clause 27.1. Source references for CJK Unified Ideographs. Where "11-lines header" and "12-lines header" appear, check to see that these aren't supposed to be "11-line header" and "12-line header".
Accepted in principle
This notation is used everywhere a linked text file is pointed to in the standard and in the amendments. Today the vast majority use the 'xx-lines' notation instead of 'xx-line', the only exception being Annex R with a '5-line' mention in the Note page 1396. This will be fixed in the next edition.

E.2 Page 21, Sub-clause 27.1. Source references for CJK Unified Ideographs. Under Note 3 where it says "those indexes are converted" read "those indexes have been converted".
Accepted

E.3 Page 22, Sub-clause 27.2. Source references for BMP Unified Ideographs. Is this meant to be "CJK Unified Ideographs"? Check please. Also, change "When non empty" to "When non-empty", and at the end of the paragraph for "using a 'H-' prefix." read "using the prefix 'H-'."
Accepted
'BMP Unified Ideographs' should read 'BMP CJK Unified Ideographs', 'CJK' was missing. Others accepted as suggested.

E.4 Page 30-1348, Clause 33. Code Tables and list of character names. Change "1st" to "first" please, and ensure that there are no other Word-induced superscripts of this type in the document.
Partially accepted
The requested change is consistent with other occurrence of 'first edition'. However using superscript for numeric notation is common editing practice and has nothing to do with Microsoft Word behavior and is in fact used frequently in the current standard without objection by other parties.

E.5 In the text at the bottom of page 5, first column, for "insert H Hanzi content" read "insert H-Hanzi content".
Accepted in principle
Change 'H Hanzi content' to 'Hanzi H source content' and 'J3A Kanji content' to 'Kanji J3A source content'.

E.6 Page 1374, Annex F Alternate format characters. For "and are thus represented, like the space characters, in the character code tables by dotted boxes" read "and, like the space characters, are represented in the character code tables by dotted boxes"
Accepted

E.7 Page 1374, Annex F.1.1 Zero-width boundary indicators. In the note, for "one of the two case" read "one of the two cases".
Accepted in principle

E.8 Page 1376, Annex F.2 Script-specific format characters. The character does not actually "suppress an inherent vowel", and such suppression is not an example of formatting in any case. Replace this sentence with:

> KHAROSHTHI VIRAMA (10A3F): This character, which indicates the suppression of an inherent vowel, when followed by a consonant, causes a combined form consisting of two or more consonants. When not followed by another consonant, it causes the consonant which precedes it to be written as subscript to the left of the letter before it and is not displayed as a visible stroke or dot as VIRAMAs are in other scripts.

Accepted

E.9 Table 6 - Rows 01-02: Latin Extended-B. Adjust the glyph of 0241 so that it has metrics which harmonize with, for instance, the O and Y on the same page. Ireland is prepared to provide the editor with such a glyph.
Accepted

E.10 Table 10 - Row 03: Greek and Coptic. The glyph at 03F9 should be the same as the glyph at 03FE but without the dot.
Accepted

E.11 Table 19 - Row 07: Arabic Supplement. The glyphs at 075B and 076A must have their outlines corrected so that the overlap error does not occur. Ireland is prepared to provide the editor with corrected glyphs.
Accepted in principle
See comment from Iran which is also proposing to provide new glyphs for the whole new block.

E.12 Table 39 - Row 12: Ethiopic. The glyphs at 125A-125D and 1260-1261 must be shown.
Accepted
Already noted page 3 of the amendment

E.13 Table 69 - Row 20: Superscripts and Subscripts. The glyph at 2094 should be the same as that of 2091 turned.
Accepted

As a result of this disposition Ireland changes its vote to YES.

Israel, (O)

Technical Comments

The SII has studied the proposals currently discussed to add various new characters to the Hebrew script, and has the following comments:

1) The current issues and others which may be foreseen today should be consolidated in one comprehensive and coherent proposal. Change is always disruptive and the number of times such disruption is imposed on users should be minimized.
2) Hebrew is an ancient and stable script, and a vast amount of Hebrew data exist, which conforms to standard ISO/IEC 10646 as published until now. Any change to the standard relative to the Hebrew script must not invalidate this mass of existing data. In other words, existing data must remain conformant and software implementing the updated standard must be able to give the same results as before when applied to the existing data.
3) Most, if not all, current proposals concerning the Hebrew script are meant to answer the needs of scholars of Biblical Hebrew and other ancient texts. Without denying the validity of such needs, SII is mainly concerned with the needs of users of modern Hebrew, which are satisfied by the current version of the standard. Users of modern Hebrew constitute undoubtedly a much larger population than the scholars interested in ancient Hebrew text, and as an average probably a less sophisticated one. Whatever additions are made to the Hebrew script must not entail any change in the working habits of users of modern Hebrew or require from them linguistic abilities higher than as of today.
4) All interested parties must be given the opportunity to express their views and to suggest whatever addenda they see fit. The intention is to solve all current and foreseeable issues, so as to ensure the long term stability which is so essential for any standard.
5) SII is very interested in this subject and volunteers to host a meeting of WG2 in which all current issues concerning the Hebrew script could be discussed, and hopefully resolved.

Therefore, we cannot accept the Hebrew related proposals and request that they be deleted from the present draft.
Not accepted

- *Although the Hebrew additions were submitted in several documents, the bulk of the proposal is in Amendment 1 with a minor addition in Amendment 2. So the changes are processed in a manner which is already as less disruptive as possible.*
- *The additions are not invalidating existing data.*
- *The Hebrew block is not just covering the need of the Modern Hebrew community, but also the need of the Biblical Hebrew community which has expressed a strong interest in getting these additions and is eagerly waiting the adoption of these characters. Because the additions do not invalidate Modern Hebrew there is no need to postpone them.*
- *The ISO amendment process is giving all interest parties sufficient time to provide feedback; therefore it does not seem necessary to postpone further the additions of these characters.*

Japan, Negative

Technical Comments

J1. Align graphics for J3A Kanji with referred JIS.

Regarding "Page 30-1348 Clause 33, Code Tables and...", on seven sub-tables shown after the text "In the CJK Unified Ideographs code tables, insert J3A Kanji content for the following entries:", replace the J column graphics in the sub-tables with the following graphics:

Rationale: This change makes the J column graphics exactly identical to the ones in the referred source standard (JIS X 0213:2004), making the reference as accurate as possible. The graphics taken from the standard is shown here for easy reference. Japan is willing to provide high quality outline data of the above graphics for IS publication if this comment is accepted.

Accepted

The picture graphics provided above will be used for the amendment publication as

J2. IICORE collection should be specified only by code points.

Regarding "Page 1356, Annex A.5 Fixed collections...",

(a) The new complementary text file IICORE.txt contains unnecessary information other than UCS code points. The file should be replaced with the one provided by IRG as N1095.

Accepted in principle

After further discussion, the two following points were made clear:

1. *The additional information (source data) is not source reference information but rather usefulness index for a given East Asian market.*
2. *As documented by IRG N1094, the IRG IICORE ad hoc group in its notes 2 and 3 has declared that the source data is necessary and that it would maintain it.*

Given this, it was deemed unnecessary to create an additional step where the data is stripped of that information. So in a reversal of discussion entertained during the previous WG2 meeting (M45 in Markham), it was agreed to maintain the source data information in the amendment with the following modifications in A.5.2 370 IICORE:

1. *Replace the term 'source' by 'usage' in all field descriptions (2nd to 8th)*
2. *Add the following Note:*

> *NOTE – The usage information provided in this sub-clause describes the usage and priority level of individual IICORE characters in the context of each source (G, T, J, H, K, M, and KP). This should not be confused with the source references for CJK Ideographs in clause 27 which establish the identity of all CJK Ideographs.*

J2. (b) In the second paragraph in new clause A.5.2, a phrase "Given its large size (9811 characters)" should be changed to "Given its large size (9810 character)".

Accepted

Clearly a mistake, the character count is 9810, not 9811.

J2. (c) In the third paragraph in new clause A.5.2, any references to the 2nd or later fields should be removed.

Rationale: Per the consensus in the WG2 Markham meeting and IRG Jeju meeting, any information other than UCS code point should not be in the normative part of the standard. Also, the IICORE agreed in IRG contains 9810 characters but 9811.

Accepted in principle

See disposition of comments J2.

J3. The word "Basic" should be removed from block, collection and character names of CJK Basic Strokes.

(a) Regarding "2. New tables.", the entry "Table 115 - Row 31: CJK Basic Strokes" should be changed to "Table 115 - Row 31: CJK Strokes".

(b) Regarding "Page 1349, Annex A.1 Collections of...", the entry "124 CJK BASIC STROKES 31C0-31EF" should be changed to "124 CJK STROKES 31C0-31EF".

(c) Regarding "Page 1352, Annex A.2.1 Blocks in the BMP", the entry "CJK BASIC STROKES 31C0-31EF" should be changed to "CJK STROKES 31C0-31EF".
(d) Regarding "Table 115 - Row 31: CJK Basic Strokes", the title of the table should be changed to "Table 115 - Row 31: CJK Strokes".
(e) In the same table as (d) above, change the names of the characters as follows:
"31C0 CJK BASIC STROKE T" to "31C0 CJK STROKE T",
"31C1 CJK BASIC STROKE WG" to "31C1 CJK STROKE WG",
"31C2 CJK BASIC STROKE XG" to "31C2 CJK STROKE XG",
"31C3 CJK BASIC STROKE XG-2" to "31C3 CJK STROKE BXG",
"31C4 CJK BASIC STROKE SZ" to "31C4 CJK STROKE SZ",
"31C5 CJK BASIC STROKE HZZ" to "31C5 CJK STROKE HZZ",
"31C6 CJK BASIC STROKE HZG" to "31C6 CJK STROKE HZG",
"31C7 CJK BASIC STROKE HP" to "31C7 CJK STROKE HP",
"31C8 CJK BASIC STROKE HZWG" to "31C8 CJK STROKE HZWG",
"31C9 CJK BASIC STROKE SZZG" to "31C9 CJK STROKE SZWG",
"31CA CJK BASIC STROKE HZT" to "31CA CJK STROKE HZT",
"31CB CJK BASIC STROKE HZZP" to "31CB CJK STROKE HZZP",
"31CC CJK BASIC STROKE HPWG" to "31CC CJK STROKE HGWG",
"31CD CJK BASIC STROKE HZW" to "31CD CJK STROKE HZW",
"31CE CJK BASIC STROKE HZZZ" to "31CE CJK STROKE HZZZ" and
"31CF CJK BASIC STROKE N" to "31CF CJK STROKE N".
(Also see the comment J4 below for 31C3, 31C9 and 31CC.)
Rationale: The word "basic" in the current names is misleading. Although it came from the inherent definition of CJK strokes being the "basic, primitive elements that constitute CJK Ideographs", the word has a false impression that there were non-basic (extended?) strokes as well, since that is the way UCS uses the word "basic" in block names. It is not the intention to classify CJK strokes into two sets: basic strokes and non-basic strokes. Simply changing the names removes this possible confusion.
Accepted
This results on the block, collection and character names to be changed from "CJK BASIC STROKE" to "CJK STROKE".

J4. Change names of some CJK stroke characters to more appropriate ones.
In "Table 115 - Row 31: CJK Strokes" (after changing the title of the tables as in the comment J3 above), the names of three characters should be changed as follows:
"31C3 CJK BASIC STROKE XG-2" to "31C3 CJK STROKE BXG",
"31C9 CJK BASIC STROKE SZZG" to "31C9 CJK STROKE SZWG" and
"31CC CJK BASIC STROKE HPWG" to "31CC CJK STROKE HGWG".
(Also see the comment J3 above.)
Rationale: For the character 31C3, if we follow character naming convention introduced in IRG N987 (Stroke types in CDL), it would make more sense to use combination of one letter abbreviated type names. Note that the new abbreviation BXG stands for Biǎn-Xié-Gōu (扁-斜-钩).
For the character 31C9, the shape of the last segment for this stroke looks more like WG (弯-钩) rather than ZG (折-钩), so the name should reflect it.
For the character 31CC, the shape of the upper segment for this stroke looks more like HG (横-钩) rather than HP (横-撇), so the name should reflect it.
Partially accepted
Japan asked the following changes:
"31C3 ㇀CJK BASIC STROKE XG-2" to "31C3 CJK STROKE BXG",
"31C9 ㇆CJK BASIC STROKE SZZG" to "31C9 CJK STROKE SZWG", and
"31CC 𠃌 CJK BASIC STROKE HPWG" to "31CC CJK STROKE HGWG".
And the US in its comment T.8 asked the following changes:
"31C3 ㇀CJK BASIC STROKE XG-2" to "31C3 CJK BASIC STROKE WOG", and
"31C4 ㇂ CJK BASIC STROKE SZ" to "31C9 CJK BASIC STROKE SW".
The consensus was the following:
"31C3 ㇀CJK BASIC STROKE XG-2" to "31C3 CJK STROKE BXG",
"31C4 ㇂ CJK BASIC STROKE SZ" to "31C9 CJK STROKE SW", and
"31C9 ㇆CJK BASIC STROKE SZZG" to "31C9 CJK STROKE SZWG".
Rationale: For 31C3 ㇀, the preference was to go for Biǎn-Xié-Gōu (扁-斜-钩) *instead of Wo-Gou* (臥鉤) *based on feedback from the HKSAR representative (the current CJK STROKE repertoire is based on input*

from the HKSCS set). The changes for 31C4 ㄴ and 31C9 ㄹ were not controversial. The change for 31CC ㄼ was not accepted, based on comments by Richard Cook:
"Note that CDL distinguishes "HP" and "HG" strokes, and we call this other one "HPWG" rather than "HGWG". I think that the use of P vs. G hinges on the fact that the G 'hook' segment is always a final flourish (always stroke final) and is not an intermediary segment. The P strokes can be final, and can have various degrees of curvature, ranging from near zero to more curve to very curved. HPWG can have a little curve or none at all in the P segment, so in this case the distinction is not distinctive. Note that there is some unavoidable (or traditional) inconsistency, I think, in some of the stroke naming. For example, CDL follows the tradition and distinguishes "shu gou" SG (vertical segment with hook rising to the left) and "shu ti" ST (vertical segment hook rising to the right), using "T" for the final flourish of the latter stroke, perhaps simply because the traditional name SG was already taken. In other cases a right (or left) rising hook is called G. Note that in the case of HZT vs. HZG the G hook goes to the left, while T always goes to the right. Again: if the name were to be changed to HGWG, this would imply not one stroke, but two, HG + WG, since G is always stroke terminal."

As a result of this disposition Japan changes its vote to YES.

Morocco, Yes with comments

Technical Comments

See Document N2862 Link : http://anubis.dkung.dk/JTC1/SC2/WG2
Accepted in principle
See disposition of Canadian comments 1, 2a and 2b.

USA: Yes with comments:

Technical comments:

T.1 Clause 4 Clause 4 Terms and definitions
The sub-clause "4.14 Composite sequence" needs to be updated to allow both ZERO WIDTH JOINER and ZERO WIDTH NON-JOINER to maintain synchronization with the Unicode Standard Version 4.01 and above.
As a result the definition of the Composite sequence should become:
A sequence of graphic characters consisting of a non-combining character followed by one or more combining characters, ZERO WIDTH JOINER, or ZERO WIDTH NON-JOINER (see also 4.xx).
Accepted

T.2 Clause 18 Block Names
Currently ISO/IEC 10646 does not provide guidelines for the naming of blocks. The US is requesting to adopt the same guidelines as for characters (defined in Annex L of the standard) plus Latin lowercase letters a to z. All currently specified block names comply with these new guidelines.
Withdrawn
The comment is withdrawn in the context of Amendment 1; however it remains an issue for Amendment 2 (because of issues surrounding the name of the N'Ko script) and will be discussed in that context. See document WG2 N2919.

T.3 Sub-clause 20.4 Variation selectors
The last part of Note 4 is ambiguous as it puts restriction on sequence content without explicitly restricting the scope to sequences containing variation selectors, which is the intended meaning. This should be clarified. Furthermore, the two last sentences of this note should be made normative if their intended meaning is preserved.
Accepted
The two last sentences of the Note 4 mentioned above will be removed from the note. They will be replaced by a new paragraph in the same section following the paragraph starting by "No sequences using characters…" and rephrased as follows:
All the allowed sequences using variation selectors are defined in this clause; all other such sequences are undefined. Furthermore, no sequences containing variation selectors and a mix of combining characters or composite characters will be defined.

T.4 Usage of text element or similar terms versus CC-data element
The terms 'text element' and 'sequence of characters' should be replaced by 'CC-data-element' which is the formal definition of the concept in this standard. This applies to occurrences in the Note 3 of Clause 24, the introductory text of Annex F and sub-clause F.1.1.
Accepted

T.5 Character sequence using UCS Sequences Identifiers (USI)
Following discussions held in meeting WG2 M44, consideration in document WG2 N2589 and resolution M44.16, the US is in favor of adding a list of USI names covering characters sequences resulting from (but not necessarily limited to) the HKSCS set (resolution M45.8).
These names should be listed in a similar way to the regular character names (linked file) in the same annex or a new annex. These names must not collide with regular characters names and must be built following the same guidelines (Annex L).
Withdrawn
For information, this has been resubmitted as document WG2 N2921 and will be considered for inclusion in amendment 2.

T.6 Latin Extended-B
The US is supporting the glyph change requested by document SC2[/WG2]N2859 concerning the character 01B3 Ƴ LATIN CAPITAL LETTER Y WITH HOOK which would result on the hook moving from the left to the right. This would make the glyph consistent with the glyph used for the small letter form 01B4 ƴ SMALL LETTER Y WITH HOOK. The glyph with the hook on the right is also much more prevalent on African writing system where it used. The US is not in favor of the encoding of another capital character Y with the hook on the right.
Accepted

T.7 Miscellaneous Symbols and Arrows
The following characters:
2B00 NORTH EAST WHITE ARROW
2B01 NORTH WEST WHITE ARROW
2B08 NORTH EAST BLACK ARROW
2B09 NORTH WEST BLACK ARROW
have either incorrect names or incorrect glyphs. Once standardized, names cannot be changed. At the same time the pattern used for 2B00-2B01 and 2B08-2B09 is already used for 2196-2199 and 21D6-21D9. However based on the immutability of character names the US is in favor of swapping the glyphs within 2B00-2B01 and 2B08-2B09 respectively.
Accepted

T.8 CJK Basic strokes
Comparing the document WG2 N2817 (Proposal to add a block of CJK Unified Basic Strokes to the UCS) and WG2 N2808R (HKSCS and GB 18030 PUA…) it looks like the character proposed for 31C4 ㇟ CJK BASIC STROKE SZ was not correctly named. Given that the current repertoire is based on HKSCS input, it is important to preserve the glyph as proposed (not the other way around). Therefore the US requests to change the name as follows:
31C4 ㇟ CJK BASIC STROKE SW
This would make the text amendment compatible with WG2 N 2817.
In addition the name for 31C3 should be changed to:
31C3 ㇔ CJK BASIC STROKE WOG
'WOG' is short for 臥鉤 wogou (the Chinese name for this stroke type).
Partially accepted
See dispositions of Japanese comment J4.

T.9 CJK Unified Ideographs H sources
According to resolution M45.8 (PUA of HKSCS-2001 and GB18030) and related document WG2 N2808, 42 new characters part of http://www.info.gov.hk/digital21/eng/hkscs/download/newchar.pdf should have been added as H-source reference. The US is asking their inclusion in the amendment according to the resolution.
Accepted

This was the result of an oversight from the project editor. It should be noted that since the comment was created, the document linked to above has grown by another 18 characters all already included in the standard. In addition, HKSAR is preparing a new version of HKSCS: 2004 which will include a total of 123 new characters from the version defined in 2001, including the 62 (42+18+2) already referenced above. This does not result in character addition to ISO/IEC 10646 beyond the ones already included in this amendment. The complete source reference for these 123 characters will be provided by HKSAR to the project editor before FDAM processing, no latest than end of February 2005.

T.10 CJK Unified Ideographs J sources

There are still many JIS X 213 ideographs not already identified as J0, J1 or JA source in the J data source file for BMP CJK Unified Ideographs. Examples of such characters are 69E9, 69EA which are both J3 level-3 characters. The US is asking for their inclusion in the amendment referenced as J3 (JIS X 0213:2000 level-3), J3A (JIS X 0213:2004 level-3) or J4 (JIS X 0213:2000 level-4) and possibly J4A (JIS X 0213:2004: level-4). This is assuming that all JIS X 213:2000 level-1 and level-2 are already referenced.

In addition, because the original J3 and J4 source as documented in JIS X 0213:2000 do not contain accurate ISO 10646 references for characters that were added later to the CJK Compatibility Ideograph block and to the CJK Unified Ideograph Extension B, the standard should mention that these references were provided in another JIS document (ISBN4-542-20129-5) published in 2002.

Lastly, the character U+9B1D which had a J4 source information: 2-93-27 or J4-7D3B (although not documented in ISO 10646:2003) has lost that source in JIS X 213:2004. However, following the principle mentioned in the newly introduced Note 2 of clause 27.1 (page 2 of the amendment), the original J4 source should be preserved for that ideograph.

Accepted

The source information from JIS X 0213:2004 will be used by the project editor to generate the source data for inclusion in the amendment. To further clarify the role of source reference information in the standard, the following paragraph will be added in clause 27 Source References for CJK Ideographs:

The source reference information establishes the character identity for CJK Ideographs. A source reference is established by associating a CJK Ideograph code position with one or several values in the source standards listed in clause 27.1 and 27.3. Such a source standard originates from the following categoriest:

- *Hanzi G sources,*
- *Hanzi H sources,*
- *Hanzi T sources,*
- *Kanji J sources,*
- *Hanja K sources,*
- *Hanja KP sources,*
- *ChuNom V sources, and*
- *Unicode U sources.*

For a given code position, only one source reference can be created for each of the source standard category (G, H, T, J, K, KP, V, and U). In order to provide a comprehensive coverage for a source standard category, when a source standard is referenced, all its unique associations with existing CJK Ideographs are documented.

T.11 Hebrew script

The US is in favor of moving the newly proposed Hebrew Qamats Qatan from

05BA HEBREW POINT QAMATS QATAN

to

05C7A HEBREW POINT QAMATS QATAN

Accepted

Similar to Irish comments T.1, and T.2.

T.12 Annex A.1

With the new additions of CJK Ideographs (Unified and Compatibility) two new collections should be added such as CJK UNIFIED IDEOGRAPHS-2005 adding 9FA6-9FBB to the content of collection 380 and CJK COMPATIBILITY IDEOGRAPHS-2005 adding FA70-FAD9 to the content of collection 382.

Accepted

T.13 Supplemental punctuation

The chart names of newly proposed U+2E1C and U+2E1D do not correctly reflect the WG2 resolution M45.6. They should read as:

2E1C LEFT LOW PARAPHRASE BRACKET

2E1D RIGHT LOW PARAPHRASE BRACKET

Note that the list of the new mirrored characters in bidirectional context (page 10-13 of the amendment) has already the correct names for these two code positions. The linked file Am1names.txt is also correct.
Accepted
Already mentioned in note page 3 of the amendment.

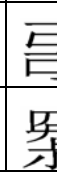
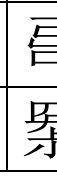
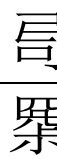
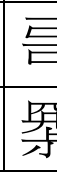
T.14 Annex B List of Combining Characters
The list of combining characters should also contain the Variation selector-17 to 256 located in plane 0E.
Accepted

T.15 Usage of sequences in Annex C (UTF-16) and Annex F
The usage and description of sequences in Annex C is not consistent with the formal definition in sub-clause 6.6 UCS Sequence Identifiers (USI). When USI are used in Annex C they should be formatted as specified in 6.6. In particular the usage of '[' , ']', '<' and '>' should not conflict with that formal specification. Annex F has many occurrences of the '<' '>' notation to bracket single short identifiers (UID) which is disallowed by sub-clause 6.6. Either that sub-clause needs to be amended or a new bracketing symbol pairs need to be used for referencing single UID within a character sequence.
Accepted in principle
Because the incorrect '<>' notation only appears in two locations (Annex C and F), a new bracketing symbol pairs "{}" will be used in these two occurrences.

T.16 Note concerning usage of SOFT-HYPHEN in Annex F
The note needs to be clarified to better use terminology from the standard. The suggested new text still uses the '<' and '>' notation that may be updated in response to comment T15.
Old text:
When a SOFT HYPHEN is used to represent a possible hyphenation point, the character representation is that of the text sequence without hyphenation (for example: "tug<00AD>gumi"). When encoding text that includes hard line breaks, including actual hyphenations, the character representation of the text sequence must reflect the changes due to hyphenation (for example: "tugg<2010>" / "gumi").
New text:
When a SOFT HYPHEN is inserted into a CC-data-element to encode a possible hyphenation point (for example: "tug<00AD>gumi"), the character representation remains otherwise unchanged. When encoding a CC-data-element that includes characters encoding hard line breaks, including actual hyphenations, the character representation of the text sequence must reflect any changes due to hyphenation (for example: "tugg<2010>" / "gumi", where / represents the line break).
Accepted

T.17 Ancient Greek Numbers
Concerning the following proposed characters:
1015A GREEK ACROPHONIC HERMIONE ONE
10162 GREEK ACROPHONIC HERMIONIAN TEN
10168 GREEK ACROPHONIC HERMIONIAN FIFTY
The pattern of naming for the rest of the dialectal Greek acrophonic names would suggest that "HERMIONIAN" was actually intended for U+1015A. Its name should be changed accordingly in the chart and names list.
Accepted

T.18 CJK Extension B characters
The document IRG1017 dated 2003.12.24 shows a list of 299 CJK Extension B Ideographs that should be updated in the chart pages. Of those 299 characters, 12 are mentioned to be more significant. The following table shows for these 12 code positions the following information: UCS code position, 10646-2:2001 glyph, 10646:2003 glyph, current G source glyph, current T source glyph and comment as appropriate.

UCS value	10646-2:2001	10646:2003	G source	T source	Comment
20BF6					GT sources, old glyph closer to T source
21BA7					GT sources, old glyph closer to T source

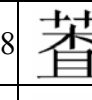
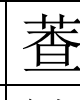
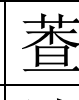
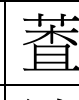
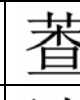
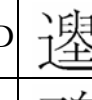
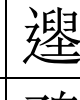
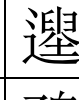
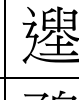
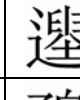
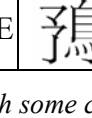
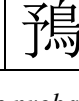
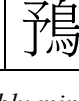
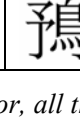
23031					T source, new glyph closer to both
230D4					GTKP sources, can't see the difference with old glyph
25962					GT sources, old glyph closer to T source
25ACD					GT sources, can't see difference
26165					T source, can't see difference
2630B					G source, old glyph closer to T source
264AB					G source, new glyph closer to both
26CD8					GT sources, new glyph closer to G source
285ED					GT sources, old glyph closer to T source
29FCE					J source, new glyph closer to J source

Assuming the current G source is used for the updated unified representation; the proposed changes seem acceptable and should be incorporated in the amendment. Some differences are very subtle (230D4, 25ACD and 26165) and may not need to be incorporated.

Accepted

The IRG referencing in the US comment is incorrect. The information originates from a file called "ModifiedCJKBFont.pdf" prepared by the contributing editor for future publication of the CJK Extension B block, based on feedback from IRG. The following table shows, in addition to the previous table, a new column containing the changes marked as significant by IRG experts.

UCS value	10646-2: 2001	10646: 2003	G source	T source	New Glyph	Comment
20BF6						GT sources, old glyph closer to T source, New glyph closer to G source
21BA7						GT sources, old glyph closer to T source, New glyph closer to G source
23031						T source, new glyph closer to both
230D4						GTKP sources, can't see the difference with old glyph
25962						GT sources, old glyph closer to T source, New glyph closer to G source
25ACD						GT sources, can't see difference
26165						T source, can't see difference
2630B						G source, old glyph closer to T source, New glyph closer to G source

264AB						G source, new glyph closer to both
26CD8						GT sources, new glyph closer to G source
285ED						GT sources, old glyph closer to T source, New glyph closer to G source
29FCE						J source, new glyph closer to G source

Although some changes are probably minor, all these 12 glyph changes will be documented as shown in Amd 1.

T.19 Status of Annex L Character naming guidelines
Given that the so-called guidelines are used normatively within the other clauses of the standard, that annex status should be changed to normative and the term ‘guidelines’ replaced with ‘rules’.
Withdrawn
The comment is withdrawn in the context of Amendment 1; however it remains an issue for Amendment 2 (because of issues surrounding the name of the N’Ko script) and will be discussed in that context. See document WG2 N2919.

Editorial Comments:

E.1 Clause Terms and definitions page 1
Add “4” after “Clause”.
Accepted

E.2 Clause 22 Compatibility characters
In the Note 1 remove the second sentence (See the definition…) as it refers incorrectly to a collection which does not include these twelve code positions.
In the fourth paragraph starting by “The CJK compatibility…” remove the two parenthetical notations “(characters that are part…)”, as they only include a snapshot of these ideographs.
Accepted

E.3 New Note for Annex F1.1. concerning COMBINING GRAPHEME JOINER page 13
The term “DIARESIS” needs to be replaced by “DIAERESIS” (two occurrences). In the last line ‘should represent’ should be replaced by ‘should be used to represent’ to be consistent with the rest of the sentence. Remove the extraneous period at the end of the note.
Accepted

E.4 Clause 24 and 25 out of sequence
Clause 25 introduces issues such as alternate representation and order of combining mark which are addressed by Clause 24 (Normalization). Reversing the two clauses would create a better logical sequence in the standard.
Accepted
