ISO/IEC JTC1/SC2/WG2 N 4454
Date: 2013-06-13

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

Doc. Type: **Disposition of comments**

Title: **Disposition of comments on SC2 4268: ISO/IEC CD 10646 4th edition**

Source: **Michel Suignard (project editor)**
Project: JTC1.02.10646.00.00.00.04
Status: For review by WG2
Date: 2013-06-13
Distribution: WG2
Reference: SC2 N4268 N4275, WG2 N4453
Medium: Paper, PDF file

Comments were received from China, Egypt, Ireland, Japan, and USA. The following document is the disposition of those comments. The disposition is organized per country. Comments were also received from South Korea (ROK) and the project editor through separate documents (WG2 N4421 and WG2 N4437 respectively).

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

Based on these dispositions, all countries have now positive votes.

China: Positive with comments

China is in favor of CD of ISO/IEC 10646 Ed. 4 with comments to code charts and CJKU_SR.txt.

Technical comment

T1. G source of U+3828 (CJK Unified Ideograph in Extension - A)

G source of U+03828 𠜎 should be changed from GHZ-10810.02 to GHZ-10810.03 according to Hanyu Da Zidian (漢語大字典).

Accepted

See also comment T45 from Japan.
Originally the GHZ sources were added in Extension A w/o numeric references. ISO/IEC 10646-2003 had none (just said GHZ). They were added later using data provided by the Unihan database from the Unicode standard using the data field 'kIRGHanyuDaZidian field'. Concerning U+3828, the issue is duplicate GHZ source between that character and 21FE2. Both are RS 46.25, Kangxi 0323.161 and for now GHZ 10810.02. The glyphs are very different. Evidences suggest that original Unihan GHZ data concerning 3828 was in error and should have been 10810.03 instead of 10810.02.

T2. G source of U+400B and U+2A279 (CJK Unified Ideographs in Extension A and B)

Both U+0400B 㚻 and U+2A279 𠛹 have G source GHZ-74611.05, the latter glyph is correct according to Hanyu Da Zidian (漢語大字典). It is suggested to remove G source of U+0400B and keep its T source..

Accepted

See also comment T45 from Japan.
The issue is duplicate GHZ source between these two characters as mentioned above. The GHZ numerical reference for U+400B came from Unihan, the GHZ reference for U+2A279 came from IRG. The characters have different RS (108.16 versus 197.10) and Kangxi (0798.171 versus 1507.311). Therefore the GHZ source for 400B will be removed.

T3. G source of U+3ABF (CJK Unified Ideograph in Extension - A)

The G source of U+03ABF㪿 is recorded GHZ, but it is not found in Hanyu Da Zidian (漢語大字典). It is suggested to remove its G source and kept its T source and J source.

Accepted

See also comment T45 from Japan.
There were never any numerical GHZ references for 3ABF, either from Unihan or IRG.

Egypt: Positive with comments

Technical comment

T1. Range code: 0600-06FF: Arabic, FE70-FEFF: Arabic Presentation Forms-B, FB50-FDFF: Arabic Presentation Forms-A, 1EE00-1EEFF: Arabic Mathematical Alphabetic Symbols, 0750-077F: Arabic Supplement, 08A0-08FE: Arabic Extended-A

The Arabic characters names as per the current citation of the standard document are not the letter names used in the Arabic references and consequently used by the native Arabic speakers.
That is why we suggest here to use the naming as pronounced by the Arabic speakers.
The use of the suggested naming will facilitate the understanding of these characters by all Arabic speaking users.
Note: Another ranges that may need to be changed accordingly like Farsi ranges.

Proposed change by Egypt
Replace all Arabic letter BEH occurrences with BA' ب
Replace all Arabic letter TEH occurrences with low TA' ت
Replace all Arabic letter THEH occurrences with THA' ث
Replace all Arabic letter HAH occurrences with HA' ح
Replace all Arabic letter KHAH occurrences with KHA' خ
Replace all Arabic letter REH occurrences with RA' ر
Replace all Arabic letter ZAIN occurrences with ZAY ز
Replace all Arabic letter TAH occurrences with High TA' ط
Replace all Arabic letter ZAH occurrences with DHA' ظ
Replace all Arabic letter FEH occurrences with FA' ف
Replace all Arabic letter HEH occurrences with HHA' هـ
Replace all Arabic letter YEH occurrences with YA' ي.

Not accepted
The comment was originally specified as 'editorial' but is indeed 'technical' because it refers to character name changes. This comment has been made several times by Egypt and has been answered in a similar fashion every time.
This comment cannot be accepted for two reasons:

1. *The proposed names contain U+0027 APOSTROPHE which is not part of the repertoire allowed for character names (see sub-clause 24.2 Name Formation in ISO/IEC 10646:2012).*
2. *More important, names of all characters cannot be changed once encoded (see clause 7 of ISO/IEC 1064:2012).*

It also be noted that the charts contain the following text in the main Arabic block (0600-06FF) as an introduction to the 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'.

T2. Character additions for Arabic mathematical operators and symbols:

Add more codes for:

1- Arabic mathematical Operators (like Ranges U+2200 : U+22FF)
2- Arabic mathematical symbols (like 1D700:1D7FF) ,because alphabetic symbols (1EEXX) is not enough.

Proposed change by Egypt
Separate Proposal will be prepared and sent to SC later:
Add function (Lim (نهاية) , cos (جتا) ,……..)
Add mathematical symbols (Integration , limited integration, differentiation ,……..)
Noted

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 1061: Row A720: Latin Extended-D.

Ireland reiterates its support for the encoding of the character at A78F and opposes further attempts to delay or prevent the encoding of this character. We note the following facts:

- Andrew West proposed this character in N3567 (2009-01-24, revised 2009-04-04) on the basis that his scientific work in Tangut and ’Phags-Pa requires a letter for transliteration of the letter [ʔ] whose transliteration is represented by a kind of dot, a use which goes back to Sinologists Dragonov in the 1930s and Karlgren in the 1940s and was taken over by Chinese scholars as well. Typography in these sources was not uniform, but a good practice can be established from them for modern use. We recommend the change of the character name from LATIN LETTER MIDDLE DOT to LATIN LETTER GLOTTAL DOT, and addition of an additional informative note to assist font developers and to reduce what the US National Body has suggested might be a measure of confusion about the character:

  A78F LATIN LETTER GLOTTAL DOT
  • used for transliteration for Phags-Pa and for phonetic transcription for Tangut
  • glyph is about 50% larger than the dots of a colon and is centred on the x-height line

  An example can be seen here of what appears to be the clearest practice:

  Tangut: ·üge·ü: ’Phags-pa

Noted
See also US comment TE1 and its disposition.
[The last two paragraphs from the Irish comment were withdrawn and have been removed from this disposition].

The proposed name is changed to LATIN LETTER SINOLOGICAL DOT with a larger and raised dot.

T2. Page 1061: Row A720: Latin Extended-D.

Ireland requests that two characters be moved:

ɑ A7AE LATIN SMALL LETTER INVERTED ALPHA
ꭥ A7AF LATIN LETTER SMALL CAPITAL OMEGA

should be moved to

ɑ AB60 LATIN SMALL LETTER INVERTED ALPHA
ꭥ AB61 LATIN LETTER SMALL CAPITAL OMEGA

Accepted in principle
These characters are added by Amendment 2 which is still under ballot. This is where these two characters are proposed for encoding.
AB64 LATIN SMALL LETTER INVERTED ALPHA
AB65 GREEK LETTER SMALL CAPITAL OMEGA

T3. Page 1222: Row 108E: Hatran.

Ireland requests the deletion of three characters and the movement of one character:

𐣻 108F8 HATRAN NUMBER ONE
𐣼 108F9 HATRAN NUMBER TWO
𐣽 108FA HATRAN NUMBER THREE
𐣾 108FB HATRAN NUMBER FOUR

should be changed to

𐣻 108FB HATRAN NUMBER ONE

Accepted

(Irish comment modified to show range 108F8..108FB and 108FB because 108F8..108FB is the actual proposed range for these four characters and 108FB is the last character in the range; the original Irish comment has the range 108F9..108FC.)

T4. Page 1244: Row 10C8: Hungarian.

With reference to ISO/IEC JTC1/SC2/WG2 N4374R "Old Hungarian/Szekely-Hungarian Rovas Ad-hoc Report", Ireland would like to note that it would not oppose a change of the script name from "Hungarian" to "Szekely-Hungarian" or "Szekler-Hungarian". Ireland notes the following from N4374R:

> In N4197 "Remarks on Old Hungarian and other scripts with regard to N4183", it is noted that "the preferred term in current Hungarian scientific literature is *'székely írás'* i.e. 'Szekler script'." Other terms for the script which have been used are "Hungarian Runic", "Hungarian script", and "Szekler-Hungarian script" (the last of which is similar to "Székely-Hungarian Rovas" promoted by "the Rovas side").

The name "Hungarian" on its own for this script is simply not found in the literature, and the name "magyar írás" seems to refer, in Hungarian, to the Latin alphabet as used for the Hungarian language. We note that "Szekler" does not require an accent where "Székely" ought to have one.

Accepted in principle

(The comment submitted by Ireland is a subset of the comment submitted for DAM2 ballot)
The script under question was part of Amendment 2 to the 3rd edition of ISO/IEC 10646 for the DAM2 ballot and the disposition is identical to the one provided for Amendment 2 (see WG2 N4453).
The name change from 'Hungarian' to 'Old Hungarian' is accepted, but the block will be moved to the 4th edition of 10646 to allow further technical review.

T5. Page 1272, Row 1158: Siddham.

Ireland requests the change of two character names.

115C4 SIDDHAM SEPARATOR-1
115C5 SIDDHAM SEPARATOR-2

should be changed to names which describe the shape of the characters:

115C4 SIDDHAM SEPARATOR DOT
115C5 SIDDHAM SEPARATOR VERTICAL BAR.

Accepted in principle

This is again a comment against additions proposed in Amendment 2 (similarly to T2, T4, and E1).

The second name change is slightly modified resulting in the following names:

115C4 SIDDHAM SEPARATOR DOT
115C5 SIDDHAM SEPARATOR BAR.

T6. Page 1298, Row 1248: Early Dynastic Cuneiform.

Ireland requests that the gap at 124D2 be closed up by moving the following characters up one position.

Accepted

T7. Page 1321, Row 1440: Anatolian Hieroglyphs.

Ireland requests the addition of several character annotations:

144A0 ANATOLIAN HIEROGLYPH A134
= syllabic ara/i
≡ 1449F ◌ 145B1

14546 ANATOLIAN HIEROGLYPH A290
= syllabic hara/i
≡ 14548 ◌ 145B1

14562 ANATOLIAN HIEROGLYPH A315
= syllabic kar
≡ 14561 ◌ 145B1

145A4 ANATOLIAN HIEROGLYPH A371A
= iudex+ra/i, tara/i-x
≡ 145A3 ◌ 145B1.

Accepted in principle
Theses proposed canonical decompositions will be modified into cross references. Similarly, the existing canonical decomposition for U+144F0, U+145B9, and U+145F8 will also be modified into cross references.

Editorial comments

E1. Page 1222, Row 108E: Hatran.

Ireland requests a change to the glyph of HATRAN LETTER RESH so that it looks more like HATRAN LETTER DALETH.
Withdrawn
See US comment TE3.

E2. Page 1227, Row 109A: Meroitic Cursive.

Ireland requests that the size of the glyphs in this block be reduced in size so that they fit better in the code chart cells.
Accepted

E3. Page 1259, Row 1118: Sharada.

Ireland [requests] the correction of the encoding error in the informative note to 111CC.
Accepted
See also comment ED1 from US.
This was a production error.

E4. Page 1320, Row 1440: Anatolian Hieroglyphs.

Ireland requests the addition of the dotted circle in the glyph for ◌𔕃 145B1 ANATOLIAN HIEROGLYPH A383 COMBINING RA OR RI.

Not accepted

After further discussion among experts and result of disposing comment T7 from Ireland and discussion of document WG2 N4441 it was decided to not add a dotted circle and change the name as follows:

145B1 ANATOLIAN HIEROGLYPH A383 RA OR RI

E5. Page 1321, Row 1440: Anatolian Hieroglyphs.

Ireland [requests] the correction of the encoding error in the informative notes to this block.

Accepted

See also comment ED1 from US.

This was a production error.

As a result of these dispositions Ireland changed its vote to Yes.

Japan: Positive with comments

General, Technical, and Editorial comments (noted as G, T, or E)

E1. Page 1 before Clause 1

Just before “1 Scope” there is a title of the standard with an extra dash at its end as follows:
Information technology — Universal
Coded Character Set (UCS) —

Proposed change by Japan
Remove the extra dash.
Accepted

T2. Page 5 Sub-clause 4.18 - Note

This NOTE looks strange. Although the NOTE says that DELETE and FORM FEED do not correspond to formal character names, they actually do. DELETE and FORM FEED are formal names given by ISO/IEC 6429. (ESC is not; its official name is ESCAPE.)

To make this NOTE valid, the names following "such as" should only include commonly-used but unofficial ones.

Proposed change by Japan
Change "DELETE, FORM FEED, ESC" to "DEL, FF, ESC"..
Not accepted
DELETE and FORM FEED are formal names in ISO/IEC 6429. In ISO/IEC 10646 control characters have no names (see Sub-clause 6.4). The long names from ISO/IEC 6429 are listed in Note 2 in clause 11 of the standard.

E3. Page 5, Sub-clause 4.21 – default state – Note

The note refers to F.2.2 and F.2.3. However, there are three sub-clauses in F.2 that mention default state. Listing only two of three is not a good practice.

Proposed change by Japan
Replace "See F.2.2 and F.2.3" to "See F.2.2, F.2.3 and F.2.4."
Accepted

E4. Page 5, Sub-clause 4.25 – extended collection

"NF" in the abbreviation "NFC" stands for "normalization form", so saying "normalization form NFC" is redundant. Clause 21 (correctly) defines the format as "normalization form C" or "NFC".

Proposed change by Japan
Replace "normalization form NFC" with "normalization form C".
Accepted in principle
The text sequence "normalization form NFC" will be replaced by "Normalization Form C (NFC)".

E5. Page 6, Sub-clause 4.25 – extended collection – NOTE 2

"NF" in the abbreviation "NFC" stands for "normalization form", so saying "normalization form NFC" is redundant.

Proposed change by Japan
Replace "normalization form NFC" with "normalization form C".
Accepted in principle
The text sequence "normalization form NFC" will be replaced by "Normalization Form C (NFC)".

E6. Page 16, Sub-clause 9.2 UTF-16 – Table 4

In the upper left, upper right, and lower left cells, there are 17 "x", but there should be 16.

Proposed change by Japan

Remove an extra "x" (one per a cell) from the upper left, upper right, and the lower left table cells.

Accepted

T7. Page 17, Sub-clause 10.7 UTF-32

UTF-16 encoding schemes and UTF-32 encoding schemes are specified in a symmetric way. 10.2 UTF-16BE and 10.5 UTF-32BE contain similar words in a same order, and 10.3 UTF-16LE and 10.6 UTF-32BE contain similar words in a same order. However, 10.4 UTF-16 and 10.7 UTF-32 have a big difference; 10.7 has only two paragraphs, while 10.4 has three paragraphs.

The first and last paragraphs of 10.4 and 10.7 match well. The second paragraph of 10.4 specifies the semantics of the signature in the UTF-16 encoding scheme. 10.7 (or anywhere else in the standard) does not contain corresponding information for the UTF-32 encoding scheme.

The UTF-32 version of the second paragraph of 10.4 should be added.

Proposed change by Japan

Add the following paragraph between the first and last paragraphs of 10.7:

In the UTF-32 encoding scheme, the initial signature read as <00 00 FE FF> indicates that the more significant octets precede the less significant octets, and <FF FE 00 00> the reverse. The signature is not part of the textual data.

Accepted

T8. Page 19, Sub-clause 12.2 Identification of a UCS encoding form – 1st paragraph

The standard text says "the identification of a UCS encoding form", but each of the listed designation sequences specifies both an encoding form and an encoding scheme.

Proposed change by Japan

Insert "and a UCS encoding scheme (see 10)" after "the identification of a UCS encoding form (see 9)".

Accepted in principle

The sub-clause 12.2 describes Escape sequences that identify a combination of encoding form and encoding scheme. But because an encoding scheme implies an encoding form it is unnecessary to mention the encoding form in the title and definition. Consequently the title and definition will only say 'encoding scheme'. See disposition of comment T10 for new text.

E9. Page 19, Sub-clause 12.2 Identification of a UCS encoding form – NOTE 1

The control character ESC was missing from some of the listed escape sequences.

The final character 04/11 in the last escape sequence is written as 04/011.

Proposed change by Japan

Replace

"ESC 02/05 02/15 04/00, ESC 02/05 02/15 04/01, ESC 02/05 02/15 04/03, ESC 02/05 02/15 04/04, 02/05 02/15 04/07, 02/05 02/15 04/08, 02/05 02/15 04/10, 02/05 02/15 04/011"

with

"ESC 02/05 02/15 04/00, ESC 02/05 02/15 04/01, ESC 02/05 02/15 04/03, ESC 02/05 02/15 04/04, ESC 02/05 02/15 04/07, ESC 02/05 02/15 04/08, ESC 02/05 02/15 04/10, ESC 02/05 02/15 04/11".

Accepted

T10. Page 19, Sub-clause 12.2 Identification of a UCS encoding form – NOTE 2

The current content of the NOTE 2 appears ambiguous and problematic.

This note is the only reference to ESC 02/05 04/07 in this standard, and this is a NOTE. That means this particular escape sequence is not a part of 10646 normative specifications, i.e., no conforming devices nor conforming interchanges are allowed to use this escape sequence.

If this interpretation is correct, the statement in this NOTE 2 is *false*, because it says ESC 02/05 04/07 *may be used*, while the standard prohibits it.

There is another interpretation, however.

When this NOTE was first introduced into 10646, i.e., in Amendment 2 to 10646-1:1993, published in 1996, the ISO/IEC Directives allowed normative notes, i.e., a NOTE was allowed to contain *normative requirements*. It is possible that the NOTE was intended to be a part of normative specification when first appeared, then the Directives changed, but we WG 2 failed to update this particular part of the standard to align with the new rules.

Based on the second interpretation, Japan proposes to make the contents of NOTE 2 a usual specification text.

An additional modification to the NOTE to 12.5 is required if we take this, because ESC 02/05 04/07 doesn't include octet 02/15, i.e., ESC 02/05 04/07 is a designation of a coding system *with standard return*.

Proposed change by Japan

Remove NOTE 2 and [move] its contents a usual standard text (at the same place.)
Rename NOTE 1 to NOTE.
In the last sentence of NOTE to 12.5, add "(except for ESC 02/05 04/07)" after "identification of UCS".

Accepted in principle

*The second interpretation is correct. However, the proposed changes by Japan do not totally clarify the situation, especially concerning when the padding is required. It is useful to mention that the UTF-8 encoding does not require padding which makes the existing sentence in Note 2: "*The escape sequence used for a return to the coding system of ISO/IEC 2022 is not padded (see 12.5)*" unnecessary. The last paragraph of sub-clause 12.5 is modified to read: "*If such an escape sequence appears within a code unit sequence conforming to this International Standard, it shall be padded in accordance with clause 11 when the identified encoding form is either UTF-16 or UTF-32. No padding in necessary when the identified encoding form is UTF-8.*"*
The modified sub-clauses 12.2and 12.5 are shown below in totality to facilitate comprehension:

12.2 Identification of a UCS encoding scheme

When the escape sequences from ISO/IEC 2022 are used, the identification of a UCS encoding scheme (see Clause 10) specified by this International Standard shall be by a designation sequence chosen from the following list:

ESC 02/05 02/15 04/09
UTF-8 encoding form; UTF-8 encoding scheme

ESC 02/05 02/15 04/12
UTF-16 encoding form; UTF-16BE encoding scheme

ESC 02/05 02/15 04/06
UTF-32 encoding form; UTF-32BE encoding scheme

NOTE – The following designation sequences: ESC 02/05 02/15 04/00, ESC 02/05 02/15 04/01, ESC 02/05 02/15 04/03, ESC 02/05 02/15 04/04, ESC 02/05 02/15 04/07, ESC 02/05 02/15 04/08, ESC 02/05 02/15 04/10, ESC 02/05 02/15 04/11 used in previous versions of this standard to identify implementation levels 1 and 2 are deprecated. The remaining designation sequences correspond to the former level 3 which is now the only supported content definition for code unit sequences.

ESC 02/05 04/07
UTF-8 encoding form; UTF-8 encoding scheme

If such an escape sequence appears within a code unit sequence conforming to ISO/IEC 2022, it shall consist only of the sequences of bit combinations as shown above.

If such an escape sequence appears within a code unit sequence conforming to this International Standard, it shall be padded in accordance with Clause 11 when the identified encoding form is either UTF-16 or UTF-32. No padding in necessary when the identified encoding form is UTF-8.

12.5 Identification of the coding system of ISO/IEC 2022

When the escape sequences from ISO/IEC 2022 are used, the identification of a return, or transfer, from UCS to the coding system of ISO/IEC 2022 shall be by the escape sequence ESC 02/05 04/00. If such an escape sequence appears within a code unit sequence conforming to this International Standard, it shall be padded in accordance with Clause 11.

If such an escape sequence appears within a code unit sequence conforming to ISO/IEC 2022, it shall consist only of the sequence of bit combinations as shown above.

> NOTE – Escape sequence ESC 02/05 04/00 is normally used for return to the restored state of ISO/IEC 2022. The escape sequence ESC 02/05 04/00 specified here is sometimes not exactly as specified in ISO/IEC 2022 due to the presence of padding octets. For this reason the escape sequences in clause 0 for the identification of UCS (except for ESC 02/05 04/07) include the octet 02/15 to indicate that the return does not always conform to that standard.

G11. Page 23, Sub-clause 16.5 Variation selector sequences

The organization of 16.5 (Variation selectors and variation sequences) appears confusing.

- It begins with defining variation selector and variation sequence (this is fine),
- The second paragraph essentially says "nothing else" (this is also fine),
- The third paragraph essentially says "IVSes are registered in IVD" without introducing the term IVS or Ideographic Variation Sequence,
- Then, the standard defines the format of the "UCSVariants.txt", whose purpose ("that specifies standardized variation sequence") is hidden in the detailed format specification, (it also uses a term Standardized Variants without giving its definition,)
- Then the categories of Standardized Variation Sequences are discussed, without explaining what is the Standardized Variation Sequence is at all.

Japan proposes to re-organize 16.5 as follows:

16.5.1 General

- Introduction of variation selector and variation sequence (the current paragraph #1),
- Introduction of standardized variation sequence, standardized variant, and ideographic variation sequence,
- The "nothing else" paragraph (the current paragraph #2),

16.5.2 Standardized variation sequences

- Specify that the standardized variation sequences are defined by the attached text file,
- Definition of the "UCSVariants.txt" file format,
- Discussion of the categories of standardized variation sequences,

16.5.3 Ideographic variation sequences

- Definition of ideographic variation sequence,
- IVSes are registered in IVD (the current paragraph #3.)

Proposed change by Japan

Insert the following new sub-clause heading after the heading for 16.5:

16.5.1 General

Insert the following paragraph after NOTE 1:

A variation sequence whose variation selector is from VARIATION SELECTORS block is called a standardized variation sequence. The variant form of a graphic symbol specified by a standardized variation sequence is called a standardized variant. A variation sequence whose base character is a CJK unified ideograph and whose variation selector is from VARIATION SELECTORS SUPPLEMENT block is called an ideographic variation sequence.

Keep the second paragraph as it currently is. Postpone the third paragraph (that begins with "Variations sequences composed of a unified ideograph...") along with NOTE 2. Insert the following sub-clause heading and a new paragraph before the fourth paragraph (that begins with "The content linked to is ..."):

16.5.2 Standardized variation sequences

Standardized variation sequences are defined by a machine-readable format that is accessible as a link.

Remaining parts of the current 16.5 comes here.

Insert the following sub-clause heading, the postponed third paragraph and NOTE 2 at the end of 16.5:

"**16.5.3 Ideographic variation sequences**"

Update the numbers of NOTEs accordingly throughout 16.5.

Accepted in principle

The reorganization makes a lot of sense. However, the paragraph defining the standardized variation sequences needs to be refined. It is not correct as stated. A variation selector from the VARIATION SELECTORS SUPPLEMENT block could be part of a standardized variation sequence (as long as it is not associated with a CJK unified ideograph). In addition, there are variation selectors outside the two variation selectors blocks, for example the MONGOLIAN FREE SELECTOR characters. The paragraph proposed by Japan for 16.5.1 is replaced by the following 2 new paragraphs:

Variations selectors are made of all coded characters included in the blocks VARIATION SELECTORS and VARIATION SELECTORS SUPPLEMENT and the three Mongolian Free Variation Selectors (FVS1 to FVS3).

A variation sequence whose base character is a CJK unified ideograph and whose variation selector is from VARIATION SELECTORS SUPPLEMENT block is called an ideographic variation sequence. All other variation sequences are called standardized variation sequences. The variant form of a graphic symbol specified by a standardized variation sequence is called a standardized variant.

E12. Page 23, Sub-clause 16.5 – Variation selector sequences

The second sentences of the third list item (for Phags-pa variation sequences) and the fifth list item (for CJK Unified Ideographs variation sequences) include a phrase "variation selector sequences". It should be "variation sequences" (without "selector").

Proposed change by Japan

Replace "variation selector sequences" with "variation sequences" (removing "selector".)

Accepted

(Same as comment E3 from Japan for DAM2)

T13. Page 23, Sub-clause 16.5 – CJK Compatibility Ideographs variation sequences

The current standard says that the newly introduced standardized variation sequences for CJK Unified Ideographs are equivalent to CJK Compatibility Ideographs and that they are preferred representation (over CJK Compatibility Ideographs), but such statements are misleading.

The intention of this list item appears that "the visual appearances specified by these variation sequences are that of CJK compatibility ideographs" and that "if an application needs to normalize the text data, and it needs to distinguish compatibility ideographs and corresponding unified ideographs after the normalization, then use of the standardized variation sequences for CJK Unified Ideographs may help."

It is better to say the point simply. Note that the second sentence is just a hint to the users and not a requirement, and it appears better to be written as a part of the NOTE.

Proposed change by Japan
Replace the list item with the following:

- CJK Unified Ideographs. Each of these variation sequences corresponds to a CJK compatibility ideograph. Its specified appearance is that of the corresponding CJK compatibility ideograph.

Replace the NOTE 7 to the list item with the following:

> NOTE 7 – If an application normalizes text data containing CJK compatibility ideographs, the CJK compatibility ideographs are replaced with the corresponding CJK unified ideographs, and the distinction between the two is lost. It makes lossless two-way code conversion impossible. On the other hand, variation sequences are unchanged by normalization process. If an application needs normalization, and it needs to distinguish appearances of CJK compatibility ideographs and corresponding CJK unified ideographs, use of the standardized variation sequences for CJK Unified Ideographs in place of CJK compatibility ideographs may be a solution. No equivalence between these variation sequences and the corresponding compatibility ideographs are defined. Conversion considerations are out of scope of this International Standard.

Partially accepted
(Same as comment T4 from Japan for DAM2)
The list item replacement is accepted as it is. However the proposed note needs to be altered to show that the use of normalization is more prevalent than suggested by Japan and is often beyond the control of applications. The new note reads as follows:

> NOTE 7 – All normalization forms replace CJK compatibility ideographs with the corresponding CJK unified ideographs, but leave the variation sequences unchanged (see 21). In contexts where normalization forms are used and the distinction between the CJK compatibility ideographs and CJK unified ideographs is desired, the usage of variation sequences is a mechanism to maintain that distinction. No equivalence between these variation sequences and the corresponding compatibility ideographs are defined. Conversion considerations are out of scope of this International Standard.

T14. Page 24, Clause 18 – Compatibility characters – Note 3

Normalization and compatibility ideographs are, in a sense, incompatible in both ways. Stating this fact from one side will mislead users.

Also, the current sentence uses a vague phrase "the distinct identity of compatibility characters". Variation sequences are neither compatibility characters nor compatibility ideographs. As the standard says, variation sequences only specify appearance.

There are some other problems in the current sentences: the NOTE 3 uses a phrase "compatibility characters" although the message strictly aims to users of compatibility ideographs as opposed to general compatibility characters.

Proposed change by Japan
Replace the NOTE 3 with the following:

> NOTE 3 - Because compatibility ideographs are not preserved through any normalization forms, use of standardized variation sequences for CJK Unified Ideographs (See 16.5) may be better if the application needs to perform

normalization and the distinction between CJK compatibility ideographs and the corresponding CJK Unified ideographs needs to be preserved. Another alternative is to avoid normalization at all.

Partially accepted

(Same as comment T5 from Japan for DAM2)

While normalization forms and compatibility ideographs are in a sense incompatible as stated by Japan it is not true that it is only stated from one side. Both the compatibility clause (18) and the normalization form clause (21) mention that situation. If there is bias toward normalization, it is because it is now prevalent in many contexts. And it is also why many experts are reluctant to encode more compatibility ideographs. Furthermore, variation sequences with definitions such as '7DF4 FE00; CJK COMPATIBILITY IDEOGRAPH-F996' might not be compatibility ideographs but they are clearly specified to preserve the concept of compatibility ideographs through context where normalizations forms are used. While variation sequences are clearly intended to specify appearance there is nothing that prevents them to be used to create a distinction between a regular character and its compatibility 'equivalent'. Variation sequences may not be the perfect vehicle to preserve the compatibility concept (including round-tripping where normalization forms are prevalent) but it was felt that using variations sequences avoided the introduction of a whole new mechanism to preserve the separate identity of compatibility ideographs. A new Note 3 is proposed as follows:

> NOTE 3 - Because compatibility ideographs are not preserved through any normalization forms, use of standardized variation sequences for CJK Unified Ideographs (see 16.5) may be preferred in contexts where normalization forms are used and the distinction between CJK compatibility ideographs and the corresponding CJK Unified ideographs needs to be preserved. In context where compatibility ideographs should be preserved normalization forms cannot be used.

T15. Page 27, Clause 21 – Normalization forms – Note 4

The NOTE begins with "Because normalization forms preserve the variation selectors", assuming the reader knows it and the reader also understand normalization replaces some compatibility characters, specifically CJK compatibility ideographs, with the corresponding characters, although it is not always the case. 10646 doesn't explain normalization procedure and does refer to the Unicode Standard, so this NOTE is better to explain more on the point.

Also, this NOTE tells the user only one side of the issue. Doing so is misleading.

Proposed change by Japan

Replace the NOTE 4 with the following:

> NOTE 4 - In all of the four normalization forms, CJK Compatibility Ideographs are replaced with the corresponding CJK Unified Ideographs. Normalization, however, doesn't alter variation selectors, and variation sequences are preserved. Because of this, it may be better to use standardized variation sequences for CJK Unified Ideographs than to use CJK Compatibility Ideographs, in the context of normalization (See 16.5). In other words, if an application needs to use CJK Compatibility ideographs and the distinction between the corresponding CJK Unified Ideographs need to be preserved, use of normalization should be avoided.

Partially accepted

(Same as comment T6 from Japan for DAM2)

Explaining in better terms the situation between normalization forms and variations selectors/sequences is a good thing. However presenting this is a one-sided presentation is in itself misleading. Stating that an option is that normalization should be avoided is unrealistic. In many contexts the benefit of normalization forms are such that they are prevalent and applications have no control on the data set they are served.

Furthermore the proposed sentence: <<In other words, if an application needs to use CJK Compatibility ideographs and the distinction between the corresponding CJK Unified Ideographs need to be preserved, use of normalization should be avoided. *>> is not accurate. The whole idea of the new CJK unified ideographs variation sequences is to allow maintaining the distinction between CJK compatibility ideograph and CJK unified ideograph without using CJK compatibility code points.*

A new Note 4 is proposed as follows:

NOTE 4 - In all of four normalization forms, CJK Compatibility Ideographs are replaced with the corresponding CJK Unified Ideographs. Normalization, however, doesn't alter variation selectors, and variation sequences are preserved. Because of this, the use of standardized variation sequences for CJK Unified Ideographs over the CJK Compatibility Ideographs is preferred in the context of normalization (see 16.5).

E16. Page 29, Sub-clause 22.4 – Source references for pictographic symbols – 2nd list

In each item on the list, a regular expression simply follows a text describing the field content. It is not clear that the regular expression specifies the format of the field.

Proposed change by Japan
Add "in the following format" between the describing text and the regular expression, e.g.,

1st field: UCS code point or sequence, in the format `(hhhh | hhhhh ) (<space> (hhhh | hhhhh))*`
Accepted

T17. Page 29, Sub-clause 22.4 – Source references for pictographic symbols – format definition

The current specification says the 'h' in the format definition is a decimal unit, but it is not. In the "EmojiSrc.txt" file, the UCS code points and various Shift-JIS codes are in hexadecimal notation. A 'h' in a format definition should be a hexadecimal unit.

Proposed change by Japan
Change "a decimal unit" to "a hexadecimal unit".
Accepted

T18. Page 29, Sub-clause 22.4 – Source references for pictographic symbols – 2nd list

The regular expression for the 1st field uses an asterisk to indicate "0, 1, or more iteration". Such use of an asterisk in a regular expression may be common, but the 10646 text has no specification of it.

Proposed change by Japan
Add the following sentence at the end of the paragraph:

An ASTERISK indicates zero, one, or more iteration of the preceding pattern.
Accepted

E19. Page 29, Sub-clause 23.1 – List of source references – text for GCYY source

The Chinese name for the GCYY source appears wrong. Better to be verified by Chinese national body.

Proposed change by Japan
Change
"中国测绘科学院用字"
to
"中国测绘科学研究院用字".
Accepted
This was verified by going to their web site at http://casm.ac.cn.

E20. Page 33, Sub-clause 23.3.1 – Source references presentation for CJK UNIFIED IDEOGRAPH block – 1st paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure." See 6.6.7.4 of ISO/IEC Directives, Part 2, 2011. (NOTE that, although the specification in the Directives says "for example", a Japanese expert on Directives, Part 2 believes that 6.6.7.4 requests that all references to Figures and Tables are in a form Figure X or Table X, where X is the number of the figure/table.).

Proposed change by Japan

Change "The following figure" to "Figure 2".

Accepted in principle

The first paragraph two sentences are changed as follows:

> For the presentation of the CJK UNIFIED IDEOGRAPH block, the graphic representations for the Hanzi G, H, and T sources, the Kanji J source, the Hanja K source, and the ChuNom V source are shown in that order when present.

E21. Page 33, Sub-clause 23.3.1 – Source references presentation for CJK UNIFIED IDEOGRAPH block – 2nd paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan

Change "The following figure" to "Figure 2".

Accepted

E22. Page 33, Sub-clause 23.3.2 – Source references presentation for CJK UNIFIED IDEOGRAPH EXTENSION A – 1st paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan

Change "The following figure" to "Figure 3".

Accepted in principle

The first paragraph two sentences are changed as follows:

> For the presentation of the CJK UNIFIED IDEOGRAPH EXTENSION A block, up to three sources per characters are represented in a single row.

E23. Page 33, Sub-clause 23.3.2 – Source references presentation for CJK UNIFIED IDEOGRAPH EXTENSION A – 2nd paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan

Change "The following figure" to "Figure 3".

Accepted

E24. Page 34, Sub-clause 23.3.3 – Source references presentation for CJK UNIFIED IDEOGRAPH EXTENSION B – 1st paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan

Change "The following figure" to "Figure 4".

Accepted in principle

The first paragraph two sentences are changed as follows:

> For the presentation of the CJK UNIFIED IDEOGRAPH EXTENSION B block, the first graphic symbol shows the glyph used for the first and second edition of this International Standard (2003 and 2011 respectively) referenced by a ‘UCS2003’ notation.

E25. Page 34, Sub-clause 23.3.3 – Source references presentation for CJK UNIFIED IDEOGRAPH EXTENSION B – 2nd paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan

Change "The following figure" to "Figure 4".
Accepted

E26. Page 34, Sub-clause 23.3.4 – Source references presentation for CJK UNIFIED IDEOGRAPH EXTENSION C, D and E – 1st paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan
Change "The following figure" to "Figure 5".
Accepted in principle
The first paragraph two sentences are changed as follows:

> For the presentation of the CJK UNIFIED IDEOGRAPH EXTENSION C, D, and E block, up to two sources per characters are represented in a single row.

E27. Page 34, Sub-clause 23.3.4 – Source references presentation for CJK UNIFIED IDEOGRAPH EXTENSION C, D and E – 2nd paragraph

In the International Standard, references to a figure should not use a phrase like "the following figure."

Proposed change by Japan
Change "The following figure" to "Figure 5".
Accepted

E28. Page 35, Sub-clause 23.5 – Source references presentation for CJK Compatibility Ideographs - Note

CJK COMPATIBILITY block contains no CJK unified ideograph.

Proposed change by Japan
Change "CJK COMPATIBILITY block" to "CJK COMPATIBILITY IDEOGRAPHS block".
Accepted
There is apparently a typo in the comment (the CJK COMPATIBILITY [IDEOGRAPHS] contains CJK unified ideographs). However the propose change by Japan is correct.

E29. Page 35, Sub-clause 23.5 – Source references presentation for CJK Compatibility Ideographs

In the International Standard, references to a figure should not use a phrase like "the following figure.".

Proposed change by Japan
Change "The following figure" to "Figure 6".
Accepted

T30. Page 37, Sub-clause 24.3 – Single name

24.3 says "Each entity named in this standard shall be given only one name." However, Japan believes there is an exception to this rule: a (normative) character name alias.

When the standard gives a (normative) character name alias to an existing character, the official (non-alias) character name doesn't change, and it is still considered as a normative name of the character. So, the character has two normative entity names.

This sub-clause should be reworded to cover the cases.

Proposed change by Japan
Add the following to the end of the sentence:

... with an exception that a character may be given two or more names; one character name and one or more character name aliases.

Not accepted

The character name alias is not a character name, it is an alias. Furthermore, single character name is a strong tenet within 10646 (see clause 7). However, the description of the single name can be clarified as follows:

> Each entity named in this standard shall be given only one name. However, one or more character name aliases may also be associated with a character.

T31. Page 37, Sub-clause 24.5.4 – Determining uniqueness

The first three lines in 24.5.4 define one of essential rules on the names, using a term "medial HYPHEN-MINUS." The definition of the term "medial HYPHEN-MINUS" does not present anywhere in the 10646's normative text, but only in the NOTE 1 of 24.5.4, which is an informative text. Because the meaning of the term is not trivial, the standard should define the term explicitly under a normative context.

Proposed change by Japan

Remove NOTE 1 and either

Put the exact sentence currently in NOTE 1 as the last sentence of the first paragraph of 24.5.4.

or

Create an entry for the term "medial HYPHEN-MINUS" in 4 Terms and Definitions, giving the phrase after the word "is" of the sentence currently in NOTE 1 as its definition.

Accepted

The first option is preferred (move the text of the note into the first paragraph) because that term is not used elsewhere in the standard.

E32. Page 38, Sub-clause 24.7 – Character names for Hangul syllables

The first level of a list should use a "lower case letter" not a number. See 5.2.5 of ISO/IEC Directives, Part 2, 2011.

Proposed change by Japan

Replace "1)", "2)", "3)", ... for the list to "a)", "b)", "c)", ... Also update references to the items appropriately, e.g., "7) Carry out steps 1 to 4 as described above" should be changed to "g) Carry out steps a to d as described above".

Accepted

E33. Page 39, Sub-clause 24.7 – Character names for Hangul syllables – paragraph just before 7)

The paragraph just before "7)" is currently indented as if it is a part of the list item "6)". However, the content of the paragraph is not a part of the list item "6)".

Proposed change by Japan

Begin the paragraph ("For each Hangul syllable character ...") at the normal left margin (un-indented.).

Accepted in principle

A better solution is to move the paragraph just after the first paragraph of the sub-clause and also make it un-indented as suggested by Japan.

T34. Page 39, Sub-clause 24.7 – Character names for Hangul syllables – 8) EXAMPLE

In the Last sentence of the EXAMPLE below list item 8), the term "additional information" is used. It is called "annotation" in some other places, e.g., a paragraph before 7), and the term "annotation" better matches the title of clause 24.

Proposed change by Japan

Change "additional information" to "annotation".

Accepted

T35. Page 39, Sub-clause 24.7 – Character names for Hangul syllables – Table 5

Title of the Table 5 uses the term "additional information". It is called "annotation" in some other places, including a table heading of the table.

Proposed change by Japan
Change "additional information" in the table title to "annotation".
Accepted

E36. Page 40, Clause 25 – Named UCS Sequence Identifiers – 2nd paragraph

"NF" in the abbreviation "NFC" stands for "normalization form", so saying "normalization form NFC" is redundant.

Proposed change by Japan
Replace "normalization form NFC" with "normalization form C".
Accepted in principle
The text sequence "normalization form NFC" will be replaced by "Normalization Form C (NFC)".

G37. Page 41-45, Clause 26-30 – Structure of planes – Figure 7-12

The figures show block names, but some names are abbreviated. The fact should be noted explicitly to avoid users' confusion.

Proposed change by Japan
Add the following NOTE to each of Figure 7-12:
NOTE Block names in the figure may be abbreviated due to the space limitations. See A.2 for unabbreviated names.
Accepted in principle
The note will be added for Figures 7-11, figure 12 does not need it.

E38. Page 45, Clause 28 – Structure of the Supplementary Ideographic Plane (SIP) – 2nd paragraph

The phrase "compatibility CJK ideographs" appears a mistake.

Proposed change by Japan
Change "compatibility CJK ideographs" to "CJK compatibility ideographs".
Accepted

T39. Page 46, Clause 31 – Code charts and lists of character names – 2nd paragraph

The current text says "Each code chart is followed by a corresponding character names list, except the CJK UNIFIED IDEOGRAPHS blocks and the HANGUL SYLLABLES block." However, code charts for CJK compatibility ideographs are not followed by character names list. They should also be excepted.

Proposed change by Japan
Change
"except the CJK UNIFIED IDEOGRAPHS blocks and the HANGUL SYLLABLES block"
to
"except blocks for CJK ideographs and hangul syllables."
Accepted in principle
New text is: "except blocks for CJK ideographs and Hangul syllables." *(note 'Hangul' instead of 'hangul')*

T40. Page 46, Clause 31.1 – Code chart

31.1 specifies the format of code chart, but the specification is not applicable to those for CJK ideographs.

Proposed change by Japan
Add the following sentences as a second paragraph in 31.1:
Code charts for CJK ideographs have different formats. See Clause 23.

Accepted

E41. Page 46, Clause 31.1 – Code chart – NOTE

See 6.6.7.3.1 of ISO/IEC Directives, Part 2, 2011. (The wording of the directive is ambiguous, but this clause is generally understood that we need to say as "Clause X" when referring to a clause X.)

Proposed change by Japan
Change "See 13." to "See Clause 13."
Accepted

T42. Page 46, Clause 31.2 – Character names list

The second item of the second list says "Subheads grouping various subsets of a given block." The word subsets may be misleading, because 10646 usually use the term subset to refer to that specified in Clause 8. Use of other wording will be better.

Proposed change by Japan
Change "Subheads grouping various subsets" to "Subheads grouping various parts".
Accepted

E43. Page 46, Sub-clause 31.2 – Characters name list – the last list item (for variation sequences)

The text says a TILDE precedes a variation sequence in the name list. However, in the actual name list, a SWUNG DASH does. The definition text and the actual name list should use the same character.

Proposed change by Japan
Change the TILDE sign that appears in "Variation sequences preceded by '~'," to a SWUNG DASH sign.
Accepted
(same as comment E8 from Japan for DAM2 ballot)

E44. Page 46, Sub-clause 31.2 – Characters name list – Example

The example lacks use of a "~" sign for variation sequences.

Proposed change by Japan
Add an appropriate part from the name list to show the use of "~" signs, e.g., a name list entry for 1820 (MONGOLIAN LETTER A), in the EXAMPLE.
Accepted in principle
Based on convenience, another example than 1820 may be chosen.

T45. Page 47, Sub-clause 31.3 – Pointers to code charts and lists of character names – Code Chart for CJK UNIFIED IDEOGRAPHS EXTENSION A

IRG resolved to report to WG 2 for two deletions and a modification of G source. Excerpts from IRG N 1896 that the IRG resolution M39.2 refers to follows:

3.1 Errors in CJKU_SR.txt
The editorial group reviewed IRGN1884 and agreed that China would report to WG2

- G source of U+03828 should be changed from GHZ-10810.02 to GHZ-10810.03 according to Hanyu Da Zidian (漢語大字典)
- G sources of U+0400B and U+03ABF should be deleted because the real sources are not found so far.

The changes are better to be included in the 4th ed.
(Note that both 3ABF and 400B have other source references than G-source, so each of the code points will not be an orphan after removing the G-source reference.)
Proposed change by Japan
Change the G-source reference "GHZ-10810.02" for 3828 to "GHZ-10810.03".

Remove the G-source reference "GHZ" for 3ABF.
Remove the G-source reference "GHZ-74611.05" for 400B.
Also update "CJK_SR.txt" accordingly.

Accepted

See also comments T1, T2, and T3 from China.

E46. Page 47, Sub-clause 31.3 – Pointers to code charts and lists of character names – Code Chart for CJK UNIFIED IDEOGRAPHS

The K2-6F42 glyph is missing for 9A3A.

Proposed change by Japan

Add an appropriate K2-6F42 glyph for 9A3A.

Accepted

G47. Page 2398, Annex F – Format characters

The layout of the texts for F.1, F.2, F.3, F.6, F.7 and F.8 is confusing. They consist of per-character explanation, sometimes preceded by general introductory paragraph(s) for a group of characters. The problem is that it is unclear which paragraph is an introductory and which is a part of explanation of a particular character.

For example, F.1.1 currently has the following structure:

F.1.1 Zero-width boundary indicators

The following characters are ...

SOFT HYPHEN: ...

The inserted graphic symbol, ...

When encoding text that includes ...

When a SOFT HYPHEN is inserted ...

ZERO WIDTH SPACE: ...

WORD JOINER and ZERO WIDTH...

The following characters are ...

ZERO WIDTH NON-JOINER: ...

ZERO WIDTH JOINER: ...

BRAHMI NUMBER JOINER: ...

The three paragraphs between the paragraph beginning with "SOFT HYPHEN" and the paragraph beginning with "ZERO WIDTH SPACE" are parts of the explanation of SOFT HYPHEN, but a paragraph between the paragraph beginning with "WORD JOINER" and "ZERO WIDTH NON-JOINER" is not a part of the explanation of WORD JOINER. It is not easy for a reader to know that.

The "The following characters" on the first paragraph refers to the first four characters, while the same phrase on the 8th paragraph refers to remaining three. Such grouping is better expressed as a sub-clause and/or list structure.
Proposed change by Japan
Organize Annex F in a suitable structure reflecting the grouping of the text. For example, F.1.1 can be organized as follows:

F.1.1 Zero-width boundary indicators
The following characters are ...
a) SOFT-HYPHEN: ...
The inserted graphic symbol, ...
When encoding text that includes ...
When a SOFT HYPHEN is inserted ...
b) ZERO WIDTH SPACE: ...
c) WORD JOINER and ZERO WIDTH...
The following characters are ...
a) ZERO WIDTH NON-JOINER: ...
b) ZERO WIDTH JOINER: ...
c) BRHMI NUMBER JOINER: ....

Partially accepted

Currently the format is nicely balanced among all sub-clauses of that annex and the suggestion from Japan would destroy that. Furthermore the same 'issue' only exists in sub-clause F.1.3 which also contains sub-groups. These sub-clauses could be split to only contain one logical group, or the term 'following characters' better targeted, or a combination of both as suggested below. The other sub-clauses only contain one group of format characters and therefore should not be confusing:
To clarify:

a) *Remove first paragraph of sub-clause F.1.1 (made superfluous by the split below).*
b) *Replace F.1.1 Zero Width boundary indicators by:*
   F.1.1 Hyphen boundary indicator (includes SOFT-HYPHEN)
   F.1.2 Word boundary indicators (includes ZERO WIDTH SPACE, WORD JOINER, and ZERO WIDTH NO-BREAK SPACE)
   F.1.3 Cursive joiners (includes ZERO WIDTH NON-JOINER and ZERO WIDTH JOINER)
c) *replace 'NOTE 2 and NOTE 3' by 'NOTE'.*
d) *Renumber following sub-clauses.*
e) *In current sub-clause F.1.3 (now F.1.5), replace first sentence of first paragraph with 'The characters described in this clause are used in formatting bidirectional text'.*
f) *In the same sub-clause F.1.3, replace in first sentence of third paragraph 'following' with 'following three'.*
g) *In the same sub-clause F.1.3, replace in first sentence of the paragraph following RIGHT-TO-LEFT MARK 'following' by 'following five'.*
h) *In the same sub-clause F.1.3, replace in first sentence of the paragraph following POP DIRECTIONAL FORMATTING 'following' by 'following three'.*

E48. Page 2398, Annex F.1.1 – Zero-width boundary indicators – NOTE 2 and NOTE 3

In F.1.1, there are NOTE 2 and NOTE 3 but no NOTE 1.

Proposed change by Japan
Change "NOTE 2" to "NOTE 1", and "NOTE 3" to "NOTE 2".

Accepted in principle

Per proposed resolution of comment G47, both notes become 'NOTE' (without number).

E49. Page 2400, Annex F.2.1 – Khmer Vowel Inherent characters

The wording "discouraged" is inappropriate for International Standards. (See Annex H of ISO/IEC Directives, Part 2.).

Proposed change by Japan
Change
"The use of these characters is discouraged."
to
"These characters should not be used."

Accepted in principle

Because these characters are no more format characters, they should be removed from the Annex, making the comment moot.

T50. Page 2402, Annex F.5 – Supertending format characters

The current F.5 looks strange.

Firstly, the text says the character explained in F.5 is used to subtend, just in a same way as in F.4, but the clause title says "Supertending format character" (contrary to F.4) It will be better to say that 0605 ARABIC NUMBER MARK ABOVE is one of subtending format characters, especially because 070F SYRIAC ABBREVIATION MARK, which has a very similar function as 0605, has been called a subtending format character in 10646 for a long time, causing no problem.

Secondly, It is also questionable that 10646 explicitly defines the scope of 0605, although the standard says vaguely "as defined in the Unicode Standard for ARABIC END OF AYAH" for other subtending characters. There is no need to specify more details on 0605 than other subtending format characters..

Proposed change by Japan
Move "0605 ARABIC NUMBER MARK ABOVE" to the list in F.4.
Remove F.5, updating the following clause numbers accordingly.

Accepted in principle

There is a typo in the text of F.5. It should say 'is used to supertend' instead of 'is used to subtend'. At the same time, the term 'supertend' is not common. The term 'subtend' although originally related to be 'underneath' is also now used to mean: 'to form or mark the outline or boundary of.' (Random House dictionary). Furthermore, as noted by Japan, the list of subtending characters already contains 070F SYRIAC ABREVIATION MARK. Based on this, the resolution is to accept the changes proposed by Japan, but also to rewrite the last paragraph of F.4 as follows:

> The scope of these characters and more details about their usage can be found in the Unicode Standard (see Annex M for referencing information).

E51. Page 2402, Annex F.6 and F.7 – Shorthand format characters

Both F.6 and F.7 explains format characters dedicated for the shorthand writing systems. It appears better to merge F.7 into F.6. We can break the (new) F.6 into two subclauses, F.6.1 and F.6.2, if sub-grouping is necessary.

Proposed change by Japan
Add a new clause title "F.6 Shorthand format characters", changing the current F.6 and F.7 as F.6.1 and F.6.2.

Accepted in principle

Merging the two sub-clauses as suggested by Japan is accepted. However there is no need to create sub-groups.

T52. Page 2402, Annex F.8 – Contiguity operators

The current text reads "where punctuation or SPACE may be omitted". "SPACE" is written in all capital, and it is interpreted as a character name. It is suspicious the statement is valid, then.

We have no definition for punctuation here. It is used as a general or vague word. It is strange that the paired word SPACE is so strict that only U+0020 is allowed. It may be better to replace the word "SPACE" with "space", an ordinary English word but a character name, as in 16.1's sense.

Proposed change by Japan

[implied by comment]
Accepted in principle
A new title and introduction inspired from the Unicode Standard 6.2 section 15.6 are suggested as follows:

> **F.8 Invisible Mathematical operators**
> In mathematics, some operators and punctuation are often implied but not displayed. Special format control characters known as invisible operators can be used to make such operators explicit for use in machine interpretation of mathematical expressions.

E53. Page 2410, Annex I – Table I.1

There is a NOTE below the Table I.1 that uses an asterisk "*" to point to a figure for IDC-OVL. It appears such style is not following ISO/IEC Directives.

In the directives, a NOTE can't point to a particular part of the standard through a mark such as "*". We need to make it a footnote if we use "*" mark. (See 6.5.2 of ISO/IEC Directives, Part 2, 2011.) Footnotes have different forms if used against a table, however, and we can't use "*" in a table but need to use a super script a, b, c, ... for footnoting. (See 6.6.6.7.)

The better way is to make it a NOTE to a table. (See 6.6.6.6.) A NOTE to a table, like a NOTE to a text, can't use "*" or similar marks to identify which part of the table is under discussion, but we can always write some wording to make the subject of a note clear.

Also note that table NOTEs should be enclosed in the table border. (See an EXAMPLE to 6.6.6.6 of ISO/IEC Directives, Part 2, 2011.).

Proposed change by Japan
Remove an asterisk "*" from the "Relative positions of DCs" column of the bottom raw.
Remove an asterisk before the word "NOTE" of the NOTE.
Insert "In IDC-OVL, " at the beginning of the NOTE.
Enclose the NOTE in a table border.
Accepted

T54. Page 2408, Annex I.1 – Syntax of an ideographic description sequence

The CD updated the definition of IDS by allowing private use characters as its DCs. Although Japan understands a requirement to allow something unencoded in UCS as a DC, it is afraid of opening up an unrestricted distribution of data containing private use characters.

Yes, IRG did use some private use characters as DCs in its own use of IDC-look-alikes, it already caused some problems even in IRG works; many IRG editors misunderstood what shapes those particular private use characters were meant, because their PC showed a different private use characters in place. In practice, it is not easy to detect a given text data contained any private use characters.

Japan considers it was a mistake that we used private use characters in IRG works. Japan worries about the issues IRG experienced may confuse world-wide UCS users.

As an alternative to private use characters, Japan would like to propose use of REPLACEMENT CHARACTER to represent a DC that is not encoded in UCS. REPLACEMENT CHARACTER is better than private use characters in the following ways: REPLACEMENT CHARACTER is expected to appear as its own glyph, that is very unlikely to be mistakenly recognized as an intended component of an ideograph by a receiving person. On the other hand, a private use code point may, by accident, have some ideograph-like character assigned by the receiver-side PC, and the receiving person may not be aware of the use of private character in the IDC, while he/she sees totally different shape than the sender's.

Proposed change by Japan

Replace the following list item to be inserted

"a private use character (as long as the interchanging parties have agreed that the particular private use character represents a particular ideograph or component of an ideograph)"

with the following:

"FFFD REPLACEMENT CHARACTER"

Partially accepted

(Same comment as T10 from Japan for DAM2)
The concern about IRG editors not being able to communicate effectively the information using private use characters is valid. However, the purpose of the new formulation is to make the use of Private Use characters conformant in that context, not to encourage their usage. This is not different from usages of Private Use characters for other purposes. It is up to the IRG group to determine its own policy concerning the use or not of Private Use characters for their own context.
At the same time, it is useful to indicate an un-encoded DC (Description Component) by a special character as suggested by Japan. However the character U+FF1F ? FULLWIDTH QUESTION MARK is preferred. To that effect the following item will be added in sub-clause I.1in the first list (after 'a coded radical…'):

- the character FF1F FULLWIDTH QUESTION MARK to represent an otherwise un-described Description Component.

E55. Page 2413, Annex L – Character naming guidelines – several places

Annex L contains a phrase "Character names and named UCS Sequence Identifiers" several times. This appears inappropriate.

"character names" is fine. "named UCS Sequence Identifiers" is problematic, because it is not a name. A "UCS Sequence Identifier" or USI for short, is a notation of the form "<UID1, UID2, ..., UIDn>" as specified in 6.6, and a "named UCS Sequence Identifier" or NUSI for short, is "a USI associated to a name". So, NUSI is a special kind of USI and is not a name that is associated to a USI. Japan believes distinction between an object and its name is important. We should not confuse NUSI itself and its name.

We have no good wording to refer to a name that is associated with a USI (or NUSI), so we need to use some verbose phrase.

Proposed change by Japan
Replace any occurrences of a phrase
"character names and named UCS Sequence Identifiers"
with
"character names and named UCS Sequence Identifier names"

Accepted in principle

Instead of using 'named UCS Sequence Identifier names' we could just use 'NUSI names' with 'Named UCS Sequence Identifier (NUSI) names' appearing at the first occurrence to re-establish the meaning of NUSI in this context.

E56. Page 2414, Annex L – Character naming guidelines – Guideline 4 – NOTE 1

The Guideline 4 has only one NOTE.
Proposed change by Japan
Change "NOTE 1" to "NOTE".

Accepted

E57. Page 2414, Annex L – Character naming guidelines – Guideline 4 – EXAMPLES

The last line of the EXAMPLES lacks the character's example glyph.

Proposed change by Japan
Add an appropriate glyph at the left most column for the last item (LATIN CAPITAL LETTER U WITH OGONEK AND ACUTE).

Accepted in principle

The last item has the appropriate glyph, it is just obscured by the glyph of the character above. The interline space will be increased to make it more visible.

G58. Page 2414, Annex L – Character naming guidelines – Structure

Organization of Annex L is unordinary. It is better to be in an ordinary structure defined by the Directives.

Proposed change by Japan
Break the current Annex L into the series of clauses as follows:
L.1 General
(The first two paragraphs of Annex L.)
L.2 Guideline 1
L.3 Guideline 2
...
L.12 Guideline 11

Not accepted

This is unnecessary. And having lines such as 'L.3 Guideline 2' where the sub-clause and the guideline number are off by one seems awkward.

E59. Page 2416, Annex M – Source of characters – 1st paragraph

The current sentence "National and international standards are listed first for each category, followed by relevant publications references." is unclear on the point that the entire Annex M is grouped by the category, first.

Proposed change by Japan
Add the following sentence before the sentence "National and ..."
The sources are grouped by their categories.

Accepted

E60. Page 2420, Annex M – Source of characters – Egyptian Hieroglyphic

One of the headings of Annex M currently reads "Egyptian Hieroglyphic". It appears strange. The term Hieroglyphic in Egyptology refers to a particular style of Hieroglyph (as in, e.g., "Egyptian Hieroglyph has three major styles of writing: Hieroglyphics, Hieratic, and Demotic.") This heading is for the script, not a particular style of writing.

Proposed change by Japan
Change a heading "Egyptian Hieroglyphic" to "Egyptian Hieroglyph".

Accepted in principle

The change will be to 'Egyptian Hieroglyphs'.

E61. Page 2420, Annex M – Source of characters – Glagolitic

In the 3rd reference under the heading "Glagolitic", the word "Prosveshchenie" is hyphenated as Prosvesh-chenie, but it is inappropriate. This is a Latin transliteration of a Russian word "просвещение", and "shch" corresponds to a single letter щ. A hyphen should not break in a single letter.

Proposed change by Japan
Change the hyphenation of "Prosvesh-chenie" to "Prosve-shchenie".

Accepted in principle

The hyphenation is automatically generated and clearly has limits when it applies to transliterated words. The best solution is to prevent hyphenation on the word.

T62. Page 2436, Annex P – Additional information on CJK Unified

Contribution WG 2 N4173 (aka IRG N1838) lists 25 CJK-B code points that contained errors, and proposes to add to the standard some information on those errors.

In the current CD 10646, two issues discussed in N4173 have been solved in other ways (i.e., T5-4C6E source reference has been removed from U+21F12 and moved to U+21F2C, and a complaint on U+29450 GKX glyph appears to be covered by the NOTE 3 in 23.1.) Other 23 code points still require some clarification.

Japan prefers to add appropriate information in Annex P.

Note that WG 2 N4173 contains a typo regarding 27B1F. The contribution text says the problematic source glyph is of GHZ-65018.09, but it actually is GHZ-64018.09 as in the code chart. The suggested addition on the right [below in the format of the disposition] contains the correction.

Proposed change by Japan

Add the following 23 entries in Annex P. (Note that the following list is arranged in a same order as N4173 for easy verification. They should be re-arranged in ascending sequence of their code points before actual addition, as required by the Annex P compilation policy.)

2382C
UCS2003 glyph for this code point was mistakenly designed.

23EE4
T7-243F source glyph was mistakenly unified to this code point.

24369
UCS2003 glyph for this code point was mistakenly designed.

27555
UCS2003 glyph for this code point was mistakenly designed.

27B1F
GHZ-64018.09 source glyph was mistakenly unified to this code point.

27D41
UCS2003 glyph for this code point was mistakenly designed.

28B75
Shape of TF-686D source glyph in TCA CNS standard has been changed after the publication of CJK UNIFIED IDEOGRAPHS EXTENSION B in ISO/IEC 10646-2. For consistency with TCA CNS standard, TF-686D glyph needs to be as in this International Standard, although the glyph is not usually unified with UCS2003 glyph of this code point.

293FB
Shape of T5-7C22 source glyph in TCA CNS standard has been changed after the publication of CJK UNIFIED IDEOGRAPHS EXTENSION B in ISO/IEC 10646-2. For consistency with TCA CNS standard, T5-7C22 glyph needs to be as in this International Standard, although the glyph is not usually unifed with GHZ-74512.13 glyph and/or UCS2003 glyph of this code point.

29C52

Shape of T7-5666 source glyph in TCA CNS standard has been changed after the publication of CJK UNIFIED IDEOGRAPHS EXTENSION B in ISO/IEC 10646-2. For consistency with TCA CNS standard, T7-5666 glyph needs to be as in this International Standard, although the glyph is not usually unified with UCS2003 glyph of this code point.

2A0B8
Shape of T7-523a source glyph in TCA CNS standard has been changed after the publication of CJK UNIFIED IDEOGRAPHS EXTENSION B in ISO/IEC 10646-2. For consistency with TCA CNS standard, T7-523A glyph needs to be as in this International Standard, although the glyph is not usually unified with GKX-1494.15 glyph and/or UCS2003 glyph of this code point.

299FB
Shape of GCH glyph for this code point has been changed after the publication of CJK UNIFIED IDEOGRAPHS EXTENSION B in ISO/IEC 10646-2. The GCH glyph needs to be as in this International Standard, although the glyph is not usually unified with UCS2003 glyph of this code point.

235F1
UCS2003 glyph for this code point was mistakenly designed.

28599
V4-5565 source glyph was mistakenly unified to this code point.

20885
T5-3669 source glyph was mistakenly unified to this code point.

24A8A
UCS2003 glyph for this code point was mistakenly designed.

24F15
UCS2003 glyph for this code point was mistakenly designed.

25089
UCS2003 glyph for this code point was mistakenly designed.

2A6C0
GKX-1538.20 source glyph was mistakenly unified to this code point.

22936
T5-6777 source glyph was mistakenly unified to this code point.

28321
T6-632A source glyph was mistakenly unified to this code point.

22BA3
GKX-0440.17 source glyph was mistakenly unified to this code point.

23023
T5-6C34 source glyph was mistakenly unified to this code point.

24229
GKX-0672.02 source glyph was mistakenly unified to this code point.

Accepted in principle

Given the size of the proposed additions, the editor will explore various formats, including tabular presentation. This will probably include glyph representation to facilitate the reading.

T63. Page 2436, Annex P – Additional information on CJK Unified

In the code chart for CJK UNIFIED IDEOGRAPHS EXTENSION B, the code point 25B88 has a same issue as in 2382C (in the previous comment.) Same words are required in Annex P.

WG 2 N4173 doesn't contain a note on 25B88, but it **is** included in the revised version of IRG N1838. The current WG 2 N4173 is a copy of the previous version of IRG N 1838 that lacked 25B88 as an editorial mistake.

Proposed change by Japan
Add the following entry in Annex P:

25B88
UCS2003 glyph for this code point was mistakenly designed.
Accepted in principle
See disposition of comment T62.

T64. Page 2436, Annex P – Additional information on CJK Unified

In the code chart for CJK UNIFIED IDEOGRAPHS EXTENSION B, the T7-2F4B has mistakenly dis-unified from 24381 and allocated to a separate code point 243BE; **and** the glyph for 243BE was wrong in 10646-2:2001. The IRG recognized that this was an error, but the consensus in its Chongqing meeting was to keep 243BE's source reference to T7-2F4B.

Since this appears an error to be corrected, we need some explanation in Annex P.

Proposed change by Japan
Add the following entry in Annex P:

243BE
The source glyph for T7-2F4B should have been unified with 24381 but was allocated here by a mistake. The UCS2003 glyph for this code point should have been based on T7-2F4B but showed different shape by a mistake. For consistency with TCA CNS standards, 24381's source reference to T7-2F4B is kept as in this International Standard.
Accepted in principle
See disposition of comment T62.

T65. Page 2438, Annex R – Names of Hangul syllables

Annex R uses the term "additional information" (twice). It is called "annotation" in some other places, including the title of 24.

Proposed change by Japan
Change two occurrences of "additional information" to "annotation".
Accepted

T66. Page 2441, Annex S.1.5 d) – Differences of actual shapes

In the second pair for "Differences in protrusion at the folded corner of strokes", the right-hand figure was changed in a wrong way, and the intended difference in actual shapes that was clear in the older editions has disappeared in the recent edition, making this example useless. (It appeared broken in the 3rd Ed.)

The wrong figure in this CD follows:

The correct figure that was in 2nd Ed (2011) follows:

Proposed change by Japan
Change two occurrences of "additional information" to "annotation".
Accepted

E67. Page 2442, Annex S.1.6 – Source separation rule – NOTEs

S.1.6 currently has two NOTEs.
Proposed change by Japan
The first NOTE should be "NOTE 1" and the second "NOTE 2".
Accepted

E68. Page 2443, Annex S.3 – Source code separation examples – clause title

S.3 is titled as "Source code separation examples", but the "source code separation" is an old wording. We now call it "source separation."

Proposed change by Japan
Change the title to "Source separation examples".
Accepted

E69. Page 2443, Annex S.3 – Source code separation examples – NOTE

The last sentence of the NOTE reads "The source groups that correspond to these letters are identified at the beginning of this annex." However, the exact location the letters are identified is in S.1.6.

Proposed change by Japan
Change "identified at the beginning of this annex" to "identified in S.1.6."
Accepted

T70. Page 2443, Annex S.3 and S.4 – Source code separation examples and Non-unification examples – Example glyph pairs (and triples)

S.3 and S.4 are sorts of records to show what the group did when it first created CJK UNIFIED IDEOGRAPHS. The current example pairs and triples in S.3 and S.4 are, however, typeset with fonts that are different from that used when CJK UNIFIED IDEOGRAPHS was first created, making the list less useful.

For an example, the following pair taken from S.3 is for a source separation caused by T source.

T

5B0E 5B14

However, this is strange; CJK Unified Ideographs are arranged in their radical-stroke order. Why a character with an extra dot has a smaller code point than that without, then?

Because, the current figure is wrong. The following figure is taken from 2003 edition, that preserved the original intention from the early days:

5B0E 5B14

A shape with a dot is on 5B14, and that without is on 5B0E. Also note the significant difference of the vertical position of the dot (between the 5B0E shape of the current CD and the 5B14 shape of the 2003 edition.) This difference is not important in this particular pair, but might have been if occurred in other pairs.

S.3 and S.4 examples are frequently referred to clarify the borders between unification and dis-unification, we should avoid changes that obscure the purposes of these examples as much as possible.

Proposed change by Japan
Revert all figures used in S.3 and S.4 to those used in earlier editions, i.e., 2003 edition.

Partially accepted

There are several considerations:

- *Many editions have been published since 2003 and no one has objected to the update of these two sub-clauses until now,*
- *The other sub-clauses of Annex S have already been reversed to pictures where it mattered,*
- *Unlike the other parts of Annex S, sub-clauses S.3 and S.4 contains code points and readers may be surprised that the shapes shown there do not correspond to the IRG source glyphs,*
- *Glyph outlines look much better than pictures,*
- *The modifications to S.3 and S.4 were done before the multi-column format for CJK was done and before the various IRG sources fonts were available to the editor. Commercial fonts were used which are sometimes quite different from the official sources. Now it is possible to use the IRG source glyphs as shown in the chart pages.*

Based on these considerations, these two sub-clauses will be redone comparing the 2003 version and the IRG source glyphs.

For example, looking at what are the exact two T source glyphs for 5B0E and 5B14 as published in the charts:

T- source 5B0E	T-source 5B14
娩	娩
T3-4B5F	T2-565F

These represents even a better case of source separation rule than the original shown above. Clearly these two characters would have been unified if the source separation rule did not exist.

E71. Many – Use of this clause or this annex

6.6.7.3.1 of ISO/IEC Directives, part 2, 2011 specifies as follows:

Imprecise references such as "this Clause" and "This Annex" shall not be used.

The wording is ambiguous, but the general understanding of this sentence is

References of the form "this Clause" and "This Annex" are considered imprecise and shall not be used.

and is not

References of the form "this Clause" and "This Annex" shall not be used when imprecise (and they may be used when precise).

Proposed change by Japan
Avoid use of the wording "this Clause", "this subclause" and "this Annex" entirely, and replace them with references using explicit numbers, e.g., "Clause X", "X.X", "Annex X".

Accepted in principle

There are many cases where using references using explicit numbers within their own clause looks very awkward. Common usage dictates that 'this' is a good pronoun to point to what is about to be stated when there is no confusion. The text of the standard will be proofed to make sure that the use of 'this' is not imprecise.

E72. Many – Prohibition of hanging paragraphs

5.2.4 of ISO/IEC Directives, part 2, 2011 prohibits hanging paragraphs, although 10646 contains many. (See EXAMPLE in 5.2.4 of the Directives for hanging paragraph.)

Proposed change by Japan
Avoid hanging paragraphs.

Accepted in principle

To see example of hanging paragraphs, please see Clause 8 and 9 which both contains introductory text before the first embedded sub-clauses. Typically, the method to avoid hanging paragraphs is to create a 'General' sub-clause to encapsulate these paragraphs. The issue is that very often the term 'general' is meaningless in context. Each occurrence of 'hanging paragraphs' will be evaluated and possibly modified. For example, it makes sense to suppress hanging paragraphs in clause 8 and 10 and to keep it in clause 9. Basically, hanging paragraphs should not exist when there may be a need to reference its specific content.

E73. Many – Enclose NOTEs in table

ISO/IEC Directives, Part 2, 2011 does not explicitly specify it by words, but the general understanding of the clause 6.6.6.6 (Notes to tables) is that any NOTEs to a table shall be enclosed in the table border, as in the EXAMPLE to 6.6.6.6.

Proposed change by Japan
Enclose any NOTEs to a table in the table border.

Partially accepted

See also comment E53.
There are 5 tables in the main body of the standard and one table in Annex I. The change requested here is already addressed in comment E53.
That leaves a single table (Table 5 page 39: Elements of Hangul syllable and additional information) with associated notes. This is a table where the row/column information is terse and the note content is verbose. It makes no sense to try to embed the note content in the table.

South Korea: Positive with comments

(The following comments were made through document WG2 N4421, not through ballot comments)

Editorial comments:

E1. On page viii,
"JIEx.txt" is a typo of "JIExt.txt"
Accepted

E2. In NOTE of clause 24.3 on page 37,
"namespace" seems a typo of "name space".
Accepted

E3 - In clauses 24.5, 24.5.1, 24.5.2, and 24.5.3,
There is a space between "name" and "space".
Accepted

USA: Positive with comments

Technical comments:

TE.1. Latin Extended-D

Justification for removing this character is contained in WG2 N3678, with further rationale in WG2 N4340 "Comments to Irish Comments on Middle Dot" and WG2 N4339 "Examples of Collation Tailoring for U+00B7 MIDDLE DOT."

Proposed change by US:

The US again requests the removal of U+A78F LATIN LETTER MIDDLE DOT. We deem the character unnecessary, is a damaging duplication for the standard, and should be removed from the amendment.
If this change is made, along with te.3, the USNB will change its vote to Yes..

Not accepted

See also comment T1 from Ireland.
The name for A78F is changed to LATIN LETTER SINOLOGICAL DOT with a larger and raised dot.

TE.2. Old Italic

The proposal WG2 N4395 has demonstrated that Raetic can amply be covered by the Old Italic script.

Proposed change by US:

The US requests the addition of U+1032F OLD ITALIC LETTER TTE, as proposed in WG2 N4395.

Accepted in principle

The character will be proposed in a future amendment to the 4th edition.

TE.3. Hatran

In Hatran, RESH and DALETH have fallen together, so there is no need to separately encode the two characters separately. The name change to HATRAN LETTER DALETH-RESH reflects the collapse of the two letters.

Proposed change by US:

The US requests the removal of U+108F3 HATRAN LETTER RESH, moving the following two characters up to fill the hole. We also request U+108E3 HATRAN LETTER DALETH be renamed HATRAN LETTER DALETH-RESH.
If these changes are made, along with te.1, the USNB will change its vote to Yes.

Accepted in principle

The hole for the removed character is maintained.

TE.4. Early Dynastic Cuneiform

A duplicate character had earlier been located at U+124D2, but it has since been removed. Contact with experts has confirmed that there is no reason to leave the hole.

Proposed change by US:

The US requests the hole at U+124D2 be closed up, and all the following characters be moved up by one code point.

Accepted

See also comment T6 from Ireland.

Editorial comments:

E.1. Incorrect characters in annotations

"Hungarian" is the block name of the script under ballot.

Proposed change by US:

The title of the ballot (page 1) currently reads "Old Hungarian". This should be corrected to "Hungarian."

Accepted
See also comment E1 from Japan.

E.2. Sub-clause 16.5

Incorrect characters appear in various annotations for new characters.

Proposed change by US:

The US requests the Editor check all new annotations, because (U+FFFD) now appears in the annotations for various characters, such as at U+111CC (Sharada block) and U+14413 (Anatolian Hieroglyphs block).

Accepted
This was a production issue (UTF-8 versus Latin 1).

.

Based on these dispositions, the US changes its vote to YES.

Project editor comments received through document N4437

Technical comments:

p. 2, Section 3 Unicode Normative References.
These are all accurate currently but should be updated to Unicode 6.3 level which will be stable by the time the DIS for the 4th edition is processed. In addition, the other scattered mentions of Unicode versions in the document should be updated to 6.3.
Accepted

p. 11, Section 6.3.1.
Update 6.1 -> 6.3.
Accepted

p. 21, Section 15.1.
Delete "6.1" in the first paragraph. Let the versioning in such cases all be handled by Clause 3, instead.
Accepted

p. 22, Section 16.3
Delete 17B4 and 17B5 from this list. These are no more format characters.
Accepted

p. 25, Section 20.2.
Remove "6.1".
Accepted

p. 2370.
Add a collection for UNICODE 6.3 (and corresponding A.6.11 clause), since its content is known currently.
Accepted

p. 2375. Section A.3.
"described in A.1" -> "described in A.6"
Accepted

p. 2393.
Add Section A.6.11 313 UNICODE 6.3 here.
Accepted

p. 2400. Section F.1.3.
Name is wrong for "FIRST STRONG ISOLATE" (remove the hyphen).
Accepted

p. 2400. Section F.2.1.
This section on Khmer should be entirely removed. These are no longer "format characters".
Accepted

p. 2402, Section F.4.
The terms "Version 6.0" are out-of-date. The reference can just be generic and Annex M contains the exact references.
Accepted

Editorial comments:

p. viii,
the number of characters is wrong. After the 2nd amendment of 10646:2012 3rd edition the count is already over 113,000 characters. And the 4th edition adds some number beyond that. The new number should be 'over 120,000'.
Accepted

p. 1, Section 1,
drop the note at the end about Unicode 7.1. With Unicode 7.0 not even started, it is premature to announce a synchronization point with a version further out.
Accepted

p. 11, Section 6.2,
small edits: 1st paragraph "as code point" -> "as a code point"; 2nd paragraph "in term of" ->"in terms of"
Accepted

p. 14, Section 7, note:
"Character name alias" -> "Character name aliases"
Accepted

p. 40, Section 25. Note 3
Note 3 is referring to an out-of-date version of UAX #34 (6.0.0). Rather than update to the 6.2 version of UAX #34, since this is a note, just referring to the *data* file directly, which is what is at issue anyway, is preferred. So replace this text with: "…are also listed in NamedSequences.txt in the Unicode character database (http://www.unicode.org/Public/UNIDATA/NamedSequences.txt)."
Accepted

p. 44, Note 2.
The note about Old Italic should be dropped entirely. It is the kind of detailed discussion about a single script which is out-of-place in this kind of overview chart. This predates the new chart format which is a much better location to present this sort of information if needed.
Accepted

p. 2417, Annex M.
Add Unicode 6.2.0. Drop the link for 6.0.0. Add one for 6.2.0 (and 6.3.0 if available).
Accepted

p. 2418, Brahmi.
The Stefan Baums entry has some character hash in it.
Accepted