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

**Title:** ISO/IEC 10646 6th edition CJK matters

**Source:** Michel Suignard (project editor)
**Project:** JTC1.02.10646.00.00.00.06
**Status:** For review by WG2 and IRG
**Date:** 2019-05-11
**Distribution:** WG2
**Reference:** SC2 N4635 N4646 N4654 N4660 WG2 N5021R IRG N2291
**Medium:** Paper, PDF file

This document is an extract of the CJK related ballot comments made for ISO/IEC 10646 6th edition CD and CD.2. The dispositions of the comments for the CD are also included along with proposed dispositions for CD.2.

IRG is encouraged to consider these comments and their dispositions (final or proposed) to help the project editor prepare the next phase of the 6th edition. This is especially important for CJK Extension G. Many of the comments were not accepted into CD.2 to allow IRG to review them before incorporating any related changes into a new ISO/IEC 10646 ballot.

However, note that some changes were already made between the CD and CD.2 as follows:

**Repertoire reordering (minor) through this ballot**
30000..31349    CJK UNIFIED IDEOGRAPHS EXTENSION G (see details in comment GE2 from Japan)

**CJK Repertoire with K source updates (see details in ROK ballot disposition):**
3EAC (delete K source), 8C6C (modified K source, same glyph), 248F2 (new K source, new glyph),
27CEF (new K source, new glyph), 2EB7E (modified K source, same glyph), 2EB89 (modified K source, same glyph)

Details concerning CJK Extension G, GonChe characters, and other CJK related matters are presented in the following pages.

# Extension G


## Ireland:


### T7. Page 2671, Row 3050: CJK Unified Ideographs Extension G

Ireland requests the removal of U+3050C (UK-02790) from CJK Unified Ideographs Extension G as it is unifiable with U+22C3A. See attached image. (The T glyph form is anomalous, but as it has the reading *fù* it should be the same character as UK-02790.).



**Not accepted**

*It is true that U+3050C 損 looks a lot like V2-7729. But the alternative solution may be to move V2-7779 to Ext G at the position U+3050C. This should be discussed by IRG experts at IRG #52 before deciding.*


## Japan:


### GE2. Page 2655, 33.5 Code charts and lists of character names – CJK Extension G

CJK Unified Ideograph Extension-G ("CJK-G" in short) code chart has been updated before SC 2/WG 2 meeting #67 in London following the discussion by IRG meeting #50, then it is adopted in this ballot text.
However, Japan has the concerns with the following.

- Japanese experts reviewed updated CJK-G code chart and found that so many things were not correctly updated following the discussion record from IRG meeting #50. Note Japan expert sent this feedback to IRG convener and technical editor with copying WG2 project editor.
- IRG, then, reviewed above feedback at IRG meeting #51 and issued new report with discussion records. Japanese experts carefully checked this report and confused very much because it concluded that there were so many incomplete or wrong records in the report from IRG meeting #50. On the other hand, many changes has been made on the rules IRG decided before, such as the unification rules, the way of stroke counting and the choice of indexed radicals.

Japan is seriously concerned with such ad hoc operation by IRG. Therefore, Japan still thinks Ext G are not matured yet. Japan would like to suggest IRG re-considering its procedure before proposing new CJK Unified Ideographs Extensions.
By the way, during this CD ballot period, CJK-G was discussed in IRG meeting #51 held in October 22nd thru 26th. See the recommendation IRG M51.2 in SC2 N4641. This obviously violates the following in ISO/IEC Directives, Part 1.

> JA.1.1 Discussion during ballot period
> ... Documents out for ballot at Committee Stage or any later stage shall not be subject to formal discussion at any working level of JTC 1 during the balloting period. ….

Proposed change by Japan.
Remove CJK-G from this CD
**Not accepted**
*The opinion of Japanese IRG experts was apparently not shared by other IRG experts. It was felt that the result out of IRG meeting #50 was good enough to progress for ballot in the CD as reflected in recommendation M67.18 (unanimous) and endorsed in SC2 resolution M23.04 (although the adoption of M23.04 was not unanimous, nobody objected or abstained on the CJK Extension G part of the resolution).*

*The review at meeting IRG#51 was extremely limited, resulting only of code point re-ordering due to radical/stroke update. The bulk of CJK Extension G feedback is still due. While it is not clear whether the level of Ext G discussion during IRG #51 rose to the level of 'formal discussion' the editor will make sure the repertoire is not under ballot during the next IRG meeting (#52 in May 2019 Hong Kong, China).*

*Note also the text in the IRG#51 editorial report (http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg51/IRGN2327WS2015EditorialReport.pdf )as:*

> Most of the inconsistencies among IRG#50 discussion record (IRGN2291) and IRGN2308RWS2015v6 raised by Japan were confirmed. Their further process were also decided.
>
> …
>
> The following records are all about discussions on Japan's comments (inconsistencies among IRG#50 discussion record IRGN2291 and IRGN2308RWS2015v6). There are 6 stroke counts and 1 glyph to be changed, the corresponding conclusions are marked in red.

*Based on this (extracted from the document linked above), the following changes will be made on the CD.2:*

3010B ⺡ 15.12 達 GZ-3542301 *SC goes to 13, code point moved to U+3010D*

3041C 弓 57.9 嶤 GZ-1912202 *SC goes to 8, code point moved to U+3041A*

309CE 石 112.9 磽 GZ-1861301 *SC goes to 8, code point moved to U+309C9*

30A35 禾 115.10 穋 GZ-3271501 *SC goes to 11, code point moved to U+30A37*

30DCA 貝 154.10 賵 GZ-4412301 *SC goes to 12, code point moved to U+30DD1*

30E21 足 157.9 踩 GZ-1231301 *glyph change to* 踩*, no R or SC change*

30E37 足 157.17 蹄 GZ-2171101 *SC goes to 18, code point unchanged*

*Resulting code chart is in http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg51/IRGN2327Codechart.pdf*
*IRG members have accepted to conduct further review of CJK Extension G and corresponding updates can be agreed upon at the next IRG meeting (#52) and incorporated into the 6$^{th}$ edition.*


# U.K.

### TE.38. 33.5 Code Charts and lists of character names – CJK Ext G – 30E21

UK submitted UK-01969 to IRG Working Set 2015, but as it was misprinted in the source (*Hànyǔ Dà Zìdiǎn*, 2nd ed., 2010) it was excluded from the CJK Ext. G submission. We have recently provided additional evidence for this character which verifies its glyph form (see http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg51/IRGN2308_UK-01969.pdf ).

As this character completes the set of derived simplified characters in *Hànyǔ Dà Zìdiǎn* (2nd ed., 2010), it would be best to include it in Ext. G with the rest of the characters in this set
<u>Proposed change by U.K.</u>
Add UK-01969 (RS 140.9) between 30C5C (T13-3068) and 30C5D (KC-03602).
**<u>Not accepted</u>**

*(IRG#51 editorial report,
it was decided that:*

> A former WS2015 character (SN03617 UK-01969 〼艹项) is accepted for
>
> inclusion in WS2017 as an exceptional case. (Also recorded in IRGN2328 Editorial
> Report on WS2017.)

*Typically, IRG is not amenable at reinserting a character that was removed from an extension in an update of such
extension. The UK experts are invited to submit this feedback to IRG#52.*

### TE.39. 33.5 Code Charts and lists of character names – CJK Ext G – 30E21

There is some confusion over the glyph form of 30E21 (GZ-1231301). The evidence shows {𧾷泰}, and this was the
glyph form shown in the previous version of the Ext. G code chart. This was realised to be a mistake in the evidence,
and so the character has been corrected to {𧾷忝} in the current code chart. However, some experts have
questioned this correction.

Additional evidence from *Nánníng Pínghuà Cídiǎn* 南寧平話詞典 [Dictionary of Nanning pinghua dialect] (Jiangsu
Jiaoyu Chubanshe, 1997) page 175 confirms that {𧾷忝} is the correct glyph form:

【□】timˇ 突然將腳或手抬起：你踩中我

　腳喇，快啲～腳起來 ‖ 壯語有 diem[tiː

　mˡ]這個詞，古壯字作'踗、躡、點"是抬

　（腳）的意思

The RS for 30E21 is given as 157.9, but it should be 157.8 to match the correct glyph form.
<u>Proposed change by U.K.</u>
Do not change the glyph for 30E21.

Correct 30E21 (GZ-1231301) RS to 157.8.

Reorder 30E21 (GZ-1231301) between 30E12 (GHR-73955.06) and 30E13 (UK-02121).
**Accepted in principle**
*However, to avoid further disruption in the Ext G code chart, this change will need to be validated by IRG#52 and
if accepted by IRG experts it will be incorporated.*

### TE.40. 33.5 Code Charts and lists of character names – CJK Ext G – 300C6, 3029A, 30773, and 323C

The following characters have incorrect Radical/Stroke values (see IRGN2269_KR_Resp2R.pdf):
300C6 (KC-00229): RS 9.17 should be 9.16
3029A (KC-00729): RS 32.11 should be 32.12
30773 (KC-02200): RS 85.18 should be 85.19
3123C (KC-04718): RS 196.12 should be 196.13.
<u>Proposed change by U.K.</u>
Correct 300C6 (KC-00229) RS to 9.16. No reordering required for 300C6 (KC-00229).

Correct 3029A (KC-00729) RS to 32.12. Reorder 3029A (KC-00729) between 302A6 (GZ-4921103) and 302A7
(UTC-01219).

Correct 30773 (KC-02200) RS to 85.19. Reorder 30773 (KC-02200) between 30774 (KC-05297) and 30775
(GHR-31928.04).

Correct 3123C (KC-04718) RS to 196.13. Reorder 3123C (KC-04718) between 31240 (GHR-84968.22) and 31241
(GHR-84976.08)..
**Not accepted**
*IRG#51 editorial report at <u>http://appsrv.cse.cuhk.edu.hk/~irg/irg/irg51/IRGN2327WS2015EditorialReport.pdf</u>
does not address these concerns. The editor has no generic issue with the stroke count and/or the re-ordering.
However, all these CJK Ext G matters should be addressed as a whole by IRG.*

# GongChe characters

## Japan (comments on CD)

**TE.3. Page 1076, 33.5 Code charts and lists of character names - Gongche" characters**
These "Gongche" characters used for musical notation as the symbol have different characteristic and different usage from CJK unified ideographs.  Further, the differences of shape from base characters cannot be considered in the framework of existing unification process. The rule that the character having tiny slash should not be unified with the base character is not consistent with the unification rule adopted so far. Therefore, these characters should not be encoded as CJK Unified Ideograph.
<u>Proposed change by Japan.</u>
Assign these characters into the code point on the block for script and symbols, not CJK ideographs
**<u>Not accepted</u>**
*There have been several opinions on how to encode these characters, either as symbols, CJK characters on their own block, or integrated into an existing CJK block. The 2<sup>nd</sup> solution was proposed in WG2 recommendation M67.10. This proved unpractical for code chart production reason. After all, these characters have radical and stroke count like any other CJK ideograph, information which cannot be easily conveyed in a symbol block. After further feedback, it was considered easier to add them to the more convenient CJK addition area (i.e. end of the URO block starting at U+9FF0).*

*A further consideration is found in quoting WG2 N4967:*

> *Many characters for Gongche Notation have the same appearances with Chinese ideographs exactly, so the vast majority of them could be used isomorphic ideographs to indicate. Hong Kong SARG once submitted two ideographs which are just used in the lyrics of Yueju Opera as UNCs in IRGN1405R. .... And UTC has submitted other two ideographs which are used in the lyrics of Kunqu Opera and traditional Gongche Notation to WS2017.*

*This means that there no intrinsic differences between ideographs used in Gonche Notation that are isomorphic or not. After all, many CJK Unified ideographs are also used in symbolic context (such as digits).*

*Based on this, it seems preferable to keep the proposed encoding as it is.*

## Japan (comments on CD.2)

**TE.2. Page 1079, 33.5 Code charts and lists of character names - Gongche" characters**
These Gongche characters used for musical notation as the symbol have different characteristic and different usage from CJK unified ideographs. Further, the differences of shape from base characters cannot be considered in the framework of existing unification process. The rule that the character having tiny slash is not be unified with the base character is not consistent with the unification rule adopted so far. Therefore, these characters should not be encoded as CJK Unified Ideograph.
For the comments Japan raised at CD ballot, the disposition of comments says as follows.
> "*This proved unpractical for code chart production reason. After all, these characters have radical and stroke count like any other CJK ideograph, information which cannot be easily conveyed in a symbol block. After further feedback, it was considered easier to add them to the more convenient CJK addition area (i.e. end of the URO block starting at U+9FF0).*"
The block where the character is encoded should not be decided from the code chart production reason. The criteria should be if it could be categorized as CJK unified ideograph. And, if so, it should be disunified with other variant characters based on the unification procedure in Annex S.
<u>Proposed change by Japan.</u>
Assign these characters into the code point on the block for script and symbols, not CJK ideographs

**IRG discussion**

*There is apparent controversy on whether these characters are CJK ideographs or just 'ideographs. According to recommendation WG2 M67.10, there were accepted as 7 'ideographs', not symbols. The decision to move these characters from a separate block to the main CJK block did not change the essence of that recommendation. Encoding them in a symbol block would be contradictory with the recommendation. At the same time, it is correct that these characters should be formally disunified from existing CJK ideographs based on the unification procedure in Annex S.*

# CJK Ideograph glyph/source reference changes

## <u>Japan</u> (comments on CD)

### TE.4. Page 1849, 33.5 Code charts and lists of character names – U+2278B

The glyph on G column of U+2278B should not be changed, because proposed glyph cannot be considered to be unified with existing one.

<u>Proposed change by Japan.</u>

Do not change the glyph. In case this is error correction, add the information in "Annex P. Additional information on CJK Unified ideographs.".

**<u>Not accepted</u>**

*For reference, code chart extract in ISO/IEC 5th edition:*



*As proposed in CD 6th edition:*



*While an entry to Annex P may be desired, it does not prevent fixing what is clearly a glyph error for the original source. See WG2 N4988 for further details.*

### TE.5. 33.5 Code charts and lists of character names – New G sources

As proposed in ISO/IEC JTC 1/SC 2/WG 2 N4988, the source reference in G column of following CJK unified characters are changed.

- from GKX-0631.02 to GHZ-31665.01 for U+3CFD
- from GE-313D to GZFY-28665 for U+6FF9
- from GE-3952 to GLK-421274 for U+809E
- from GE-3D37 to GHZR-63304.09 for U+891D
- from GHZ-10761.12 to GHF-0229 for U+21D4C

The source reference is normative information and this change may cause the compatibility problem on the data transformation table with existing implementation, therefore this change should be avoided.

<u>Proposed change by Japan.</u>

Do not change the source reference of these 5 glyphs on G column.

**<u>Not accepted</u>**

*There are precedents to this kind of changes (including changes affecting J sources), and the requested was made by China (WG2 N4988) which should be well aware of possible impacts on data processing for its own sources. Source information typically determines the identity of encoded characters and may not necessarily be used as primary source for data transformation.*

## <u>Japan</u> (comments on CD.2)

### TE.3. Page 1858, 33.5 Code charts and lists of character names – U+2278B

When this character was originally assigned on U+2278B with the G source twenty years ago, the glyph of it was the one with radical person as follows.

Page 6

However, it was changed to the one with radical heart one in the second edition of this standard published in 2011 as follows.



Then again, it is proposed to change back to the one with radical person. This is very confusing history to the user, and it affects the reliability of this standard.

Proposed change by Japan.

Do not change the glyph. In case this is error correction, add the information in "Annex P. Additional information on CJK Unified ideographs.".

**Propose acceptance**


# Korea (ROK):

### T1. Page 1997, CJK Unified Ideographs Extension B – U+248F2 and U+3EAC

K6-1002 is currently in U+3EAC. However, it seems proper to move K6-1002 to U+248F2.

If accepted by WG2, KR will send a TTF file containing a modified glyph of K6-1002.

See WG2 N5016 (= IRG N2349) for details.

Proposed change by Korea

Move K6-1002 currently in U+3EAC to U+248F2.

**Accepted**

*The actual reference is K6-1022. The editor needs the font to process this change (another source used in the interim).*

*Note that document WG2 N5015 is asking to swap the K sources between 2EB7E and 2EB89 because it was an incorrect transcription from the original K submission. This will be accommodated in CD.2.*
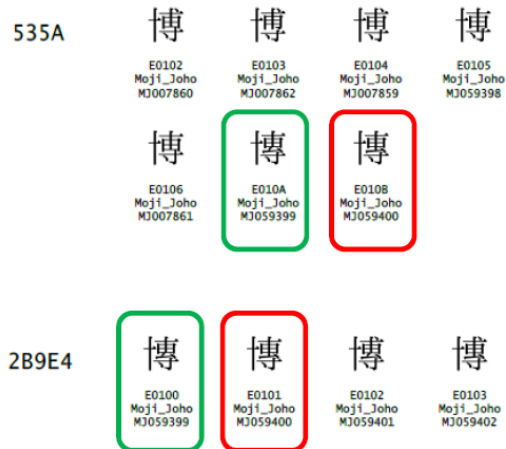


*Finally, the K source K1-6B6B was moved to U+27CEF with the glyph that was formerly at U+8C6C.U+8C6C has now a ghost K source KU-08C6C that will need to be changed or formalized.*

# Miscellaneous comments related to CJK matters

## Japan

### TE.6. Page 2768, A.5.10 390 MOJI-JOHO-KIBAN IDEOGRAPHS-2016 and A.5.11 391 MOJI-JOHO-KIBAN IDEOGRAPHS-2018

In IVD registration of Moji_Joho collection (Registration of additional sequences in the Moji_Joho collection, dated 2017-12-12), it is found that two different IVSs are registered for one glyph redundantly. As shown in the following figure, [535A E010A] and [2B9E4 E0100] are registered for Moji_Joho MJ059399, also [535A E010B] and [2B9E4 E0101] are registered for Moji_Joho MJ059400.



Proposed change by Japan.

- In JMJKI-2018.txt, delete <535A E010A> and <535A E010B> in the line starting from "535A."
- Associating with this change, add the following statement in Page 2768, Annex A.5.10 390 MOJI-JOHO-KIBAN IDEOGRAPHS-2016 as NOTE, such as "<535A,E010A> and <535A,E010B> have been removed in collection 391."

**Accepted**

### ED.7. Page 34, Table 5 – GHR syntax
Proposed change by Japan.
Change "GHR-ddddd.dd" to "GHR-ddddd.dd."
**Accepted in principle**
*The entry was in fact a duplicate of the previous entry (GHZR) and will be removed entirely along with its description in G sources enumeration.*

## United Kingdom:

### TE.4. page 24, 16.6.1 General (Variation selectors and variation sequences)
"A variation sequence whose base character is a CJK Unified ideograph and whose variation selector is from the VARIATION SELECTORS SUPPLEMENT BLOCK is called an ideographic variation sequence."

UTS #37 (since Revision 10) defines an ideographic variation sequence as being "a sequence of two coded characters, the first being a character with the Ideographic property that is not canonically nor compatibly decomposable, the second being a variation selector character in the range U+E0100 to U+E01EF". This does not restrict an IVS base character to a CJK unified ideograph, but may be any character with the ideographic property (as long as it is not canonically or compatibly decomposable), including Tangut ideographs.

Proposed change by U.K.
Change to "A variation sequence whose base character is an ideograph which is not canonically or compatibly decomposable and whose variation selector is from the VARIATION SELECTORS SUPPLEMENT BLOCK is called an ideographic variation sequence."

Add a note that at present the only ideographic variation sequences are for CJK unified ideographs.
**Accepted**
*Though there is an issue with the term 'ideograph' which is not fully specified in ISO/IEC 10646. Doing some research, the editor could find some elements in the original proposal in*
*https://unicode.org/L2/L2016/16291-ivd-proposal.pdf but not in the current UTS #37 at*
*http://www.unicode.org/reports/tr37/ . The original proposal had an enumeration of the proposed ideographs, UTS #37 specifies an 'ideographic property' but without formally establishing the value. Ideally, there should be a reference to an externally defined property (TBD) or a list of those code points.*

*The Note would read (previous note in sub-clause is renumbered NOTE 1):*

> NOTE 2 – At present the only specified ideographic variation sequences are for CJK unified ideographs.

### TE.7. page 25, 16.6.3 Ideographic variation sequences – ideograph
"Variations sequences composed of a unified ideograph as the base character …"

IVS base character is not restricted to a CJK unified ideograph, but may be any character with the ideographic property (as long as it is not canonically or compatibly decomposable), including Tangut ideographs.
Proposed change by U.K.
Change to "Variations sequences composed of an ideograph as the base character …".
**Accepted**
*See disposition of comment TE.4 about the definition of 'ideograph'.*

### TE.8. page 25, 16.6.3 Ideographic variation sequences – ideographic
"The purpose of IVD registration is to provide a technical solution to represent ideographs variants that are considered unifiable (see Annex S) and should not be encoded in CJK Unified ideographs."

This only refers to CJK unified ideographs, but IVD registration is open to other ideographic scripts such as Tangut.
Proposed change by U.K.
Change to "The purpose of IVD registration is to provide a technical solution to represent ideographic variants that are considered unifiable (see Annex S for the unification rules for CJK Unified ideographs)."

Note change of ungrammatical "ideographs variants" to "ideographic variants".
**Accepted**
*See disposition of comment TE.4 about the definition of 'ideograph'.*

### ED.30. 33.5 Code Charts and lists of character names – 4E00-9FFF
9FD4. Source glyphs are on two rows.
Proposed change by U.K.
Keep two source glyphs on a single row.
**Not accepted**
*This is a limit of the code chart production tool. For the CJK Unified Ideograph main block (4E00-9FFF), the C, J, K, and V are show on the same row, while the U column is on the second row. If the first row is empty that row is collapsed.*

### TE.41. Page 2756, Annex A.1 Collections
Collections 2001 through 2007 are for CJK unified and compatibility ideographs in SIP and TIP (2007 = CJK Unified Ideographs Extension G).

Page 9

Collection 2000 is for SIP, but there is no collection for TIP.

As CJK Unified Ideographs Extension G continues the 2000 series, we suggest redefining 2000 to cover SIP and TIP..

Proposed change by U.K.
Change Collection 2000 to "SIP and TIP" with code points 20000 through 3FFFD.
**Accepted in principle**
*The exact code point ranges would be 20000-2FFFD and 30000-3FFFD.*

### ED.42. page 2785, Annex I.3, Table I.1
The IDS examples are not the same as given in the Unicode Standard, which may be confusing for users of the two standards.
Proposed change by U.K.
Change the IDS examples in Table I.1 to match the examples shown in the Unicode Standard 11.0 Figure 18-8.
**Accepted**
*For reference the figure 18-8 is repeated below:*
**Figure 18-8**. Examples of Ideographic Description Characters

| | | | | |
|---|---|---|---|---|
| U+2FF0 ⿰ | U+4EC1 仁 | → | ⿰亻二 | |
| U+2FF1 ⿱ | U+5409 吉 | → | ⿱士口 | |
| U+2FF2 ⿲ | U+8857 街 | → | ⿲彳圭亍 | |
| U+2FF3 ⿳ | U+58F9 壹 | → | ⿳士冖豆 | |
| U+2FF4 ⿴ | U+56DE 回 | → | ⿴口口 | |
| U+2FF5 ⿵ | U+51F0 凰 | → | ⿵几皇 | |
| U+2FF6 ⿶ | U+51F6 凶 | → | ⿶凵乂 | |
| U+2FF7 ⿷ | U+5321 匡 | → | ⿷匚王 | |
| U+2FF8 ⿸ | U+4EC4 仄 | → | ⿸厂人 | |
| U+2FF9 ⿹ | U+5F0F 式 | → | ⿹弋工 | |
| U+2FFA ⿺ | U+8D85 超 | → | ⿺走召 | |
| U+2FFB ⿻ | U+5DEB 巫 | → | ⿻工从 | |

### ED.44. page 2828, Annex S.1.4.3
Very poor quality font images are used instead of coded characters.
Proposed change by U.K.
Use coded characters instead of images, with different fonts for the variant glyphs, or else use a single font and IVS sequences.
**Accepted in principle**
*It is true that pictures are used for some part of these examples (6 pairs out of the 21 pairs) and could be improved by using fonts (although the current shapes are not 'poor quality'), but the current status already reflects a significant change from earlier versions. At the same time, exact shapes are important to that annex and any changes need to be carefully analyzed by IRG experts. If a font was provided to the editor containing the required exact shapes, the corresponding font images could be incorporated.*