

ISO/IEC JTC1/SC2/WG2/IRG

Ideographic Rapporteur Group

(IRG)

Source/Contribution Identifier : IRG  
 Meeting : 19<sup>th</sup> IRG Meeting in Macau  
 Title : Error Correction  
 Status :

The IRG understands and agrees with WG2's position on the need to maintain stability in the mapping data for CJK Unified Ideographs between ISO/IEC 10646 and other character encoding standards. The IRG recognizes that in principle changes to the mapping data for the Unified CJK Ideographs in ISO/IEC 10646 should be avoided if at all possible, but that there are cases where the mapping data is undeniably incorrect. In such cases, the cost of leaving the data as it is may well be higher than the cost of making a change. Errors reported in these data should be reviewed by the IRG to determine if a change is appropriate. Such a determination can be made on the basis of general principles adopted by the IRG, with an understanding that the facts in individual cases may well override the general principles. The IRG can then make a recommendation to WG2 as to whether or not a change should be made.

The IRG considered errors in mapping and glyph data reported through May 2002 at its nineteenth meeting. The following recommendations are made—

*The G-source for the following characters needs to be changed from G\_KX to G\_HZ:*

U+23B0E, U+23839, U+24BAD, U+26CED, U+26F5F, U+27466, U+23F72, U+27B66

*The G-source for the following characters needs to be dropped:*

U+24B3E, U+24C67, U+29B3B, U+2579B, U+266A9, U+24454, U+239F7

*The glyphs for the following eight characters need to be changed as follows:*

Code point	Correct glyph
U+21E45	𪗗
U+236AD	𪗗

U+2384F	𩇛
U+24FC4	𩇛
U+25D0E	𩇛
U+276EB	𩇛
U+27CF1	𩇛
U+2890F	𩇛

*The following mappings should be changed:*

U+23FF0, the H-source should be changed from 8E6A to FDD8

U+25D20, the H-source should be changed from FDD8 to 8E6A

(That is, the H-sources for these two characters should be exchanged).

The IRG feels that since the H-source and Extension B are both relatively recent, these changes can be made without significant negative impact on existing implementations and data.

*The following two characters have had mapping and glyph errata reported to the IRG. The negative impact of making the change indicated is unclear, and the IRG cannot make a solid recommendation to WG2:*

𩇛

The T-source for U+5B90 should be changed from 4-2625 ( 𩇛 ) to 4-225B

( 𩇛 )

This would fix a clear error in the mappings between CNS 11643-1992 and ISO/IEC 10646-1. Since the mapping is to plane 4 of the standard (which is relatively unused) and the standard is not much used out of Taiwan, the impact of making the change would be minimal. On the other hand, the two characters are variants of one another, so retaining the error would have a minimal impact on the semantic content of texts.

磚

The character with T-source 2-662B should be changed from U+7921 (

磚

) and U+7934 (

). This one is more complicated. The difference between the two characters is subtle, the latter having the three-dot form of the water radical which the former lacks. In point of fact, this error resulted from a typographical error in the original CNS 11643. U+7934 is a not uncommon character in Chinese text, and U+7921 does not actually occur. The clear intent of CNS 11643 is to use its code point 2-662B to represent the same character which is encoded in ISO/IEC 10646-1 as U+7934. However, plane 2 of CNS 11643 is a remapped version of part of the Big 5 character set, which is the predominant encoding for traditional Chinese on the Internet. Big 5 fonts are inconsistent in the glyph they use, some having the three-dot water and some not. Moreover, HC SCS (which is an extension to the Big 5) explicitly includes a character which can map to U+7934.

Making the change therefore would (on the one hand) clarify a confusing situation and correct a clear error in the mapping between what CNS 11643/Big 5 is intended to include and ISO/IEC 10646-1. This will simplify and ease conversion between CNS 11643/Big 5 and other character sets (among them character sets from the PRC, Japan, and Korea, as well as others). On the other hand, it would have an impact on data already encoded using the most common character set for traditional Chinese and create a duplicate character in HK SCS where none existed before.

The recommendation from the IRG to WG2 is that the current glyph for the T-source in

磚

ISO/IEC 10646-1 be changed to the water-free version (

磚

). The delegation from Taiwan to the IRG will request that CNS 11643 be amended to add a new character with the water component (

), and, if that is done, ISO/IEC 10646 can add the mapping to the new character.