

Universal Multiple-Octet Coded Character Set  
International Organization for Standardization  
Organisation Internationale de Normalisation  
Международная организация по стандартизации

**Doc Type:** ISO/IEC JTC1/SC2/WG2/IRG

**Title:** UK Response to Comments on UK Submission to IRG Working Set 2015 (IRGN2133)

**Source:** Andrew West

**Status:** Member Body Contribution

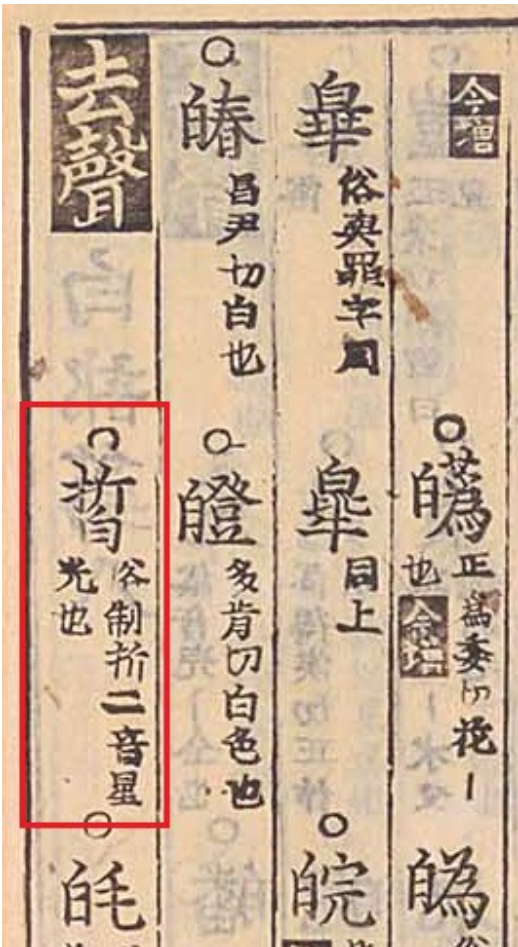
**Action:** For consideration by IRG

**Date:** 2016-05-17

This document provides the UK responses to comments relating to the UK submission to IRG Working Set 2015 (IRGN2107), as well as responses to comments relating to those characters in the UTC submission for IRG Working Set 2015 (IRGN2091) that were proposed to the UTC by Andrew West. See IRGN2133 Consolidated Comments On CJK2015 v1.1 Parts 1 and 2.

## 1. Evidence Issues

### 1.1 Buggy Evidence


S/N	By	Comment	Response
02658	J	UTC-02660 is written incorrectly in the evidence.	<p>Not accepted. The early edition of 龍龕手鑑 <i>Lóngkān Shǒujiàn</i> held at Kyoto University shows this character correctly under the 白 'white' radical.</p> 

S/N	By	Comment	Response
02671	J	UTC-02813 is written incorrectly in the evidence.	We agree that UTC-02813 is written as U+22F1B 𠂇 in 廣韻 <i>Guǎngyùn</i> , but this is only the secondary reference for this character in 汉语方言大词典 <i>Hànyǔ Fāngyán Dà Cídiǎn</i> . The primary meaning of this character is ‘to hit’ in modern Shanghai dialect, and there is no evidence to suggest that this character is not written as shown in the dictionary when used in modern Shanghai dialect. It is quite likely that the linking of this modern word to a character attested a thousand years earlier in 廣韻 <i>Guǎngyùn</i> is a mistake. However, on the basis that the character attested in 汉语方言大词典 <i>Hànyǔ Fāngyán Dà Cídiǎn</i> is used in modern Shanghai dialect, we suggest that this character should not be unified with U+22F1B.

### 1.2 Multiple Characters

No UTC characters mentioned.

### 1.3 Unclear Evidence

S/N	By	Comment	Response
01832	J	Is UTC-01315 written with a single stroke or two separate strokes at the top?	The index entry for this character on p. 580 of the source clearly shows that the character consists of two separated 正 characters:  Therefore the glyph for UTC-01315 is correct.
02857	J	Evidence for UTC-02956 shows a completely different glyph.	The evidence for UTC-02956 is Fig. 1696; and the evidence for UTC-02957 is Fig. 1697 (both on p. 342 of IRGN2107_evidence.pdf). The reviewer seems to have confused the evidence for UTC-02956 with that of UTC-02957.

#### 1.4 Derived Simplified Characters

The Japanese comments on 133 UTC characters (not individually listed here) and one G character (GHZ42306.03) are unreasonable, and should not be accepted.

The IRG Proposal Summary Form that accompanies the submission states at B. 4 “Evidence”:

*Do all the proposed ideographs have a separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)?*

The submissions for these characters satisfies this requirement, and nothing on the form suggests that additional evidence is required for simplified Chinese characters. Indeed, there is ample precedent in Exts. C through F for encoding simplified Chinese characters on the basis of dictionary evidence alone. The proposal summary form even states “**preferably dictionaries**”, and the sources for these characters are important and reputable PRC dictionaries such as 漢語大字典 *Hànyǔ Dàzìdiǎn*, 中华字海 *Zhōnghuá Zìhǎi* and 现代汉语词典 *Xiàndài Hànyǔ Cídiǎn*. Usage in these dictionaries is evidence of use, and no further evidence should be required.

The statement (e.g. for S/N 01490 UTC-02752) that simplified forms of pre-modern characters should not be encoded shows a complete misunderstanding of the use of simplified characters in PRC, where simplified characters are regularly used to represent pre-modern or ancient texts. If a simplified character is used in the modern orthography of PRC then it needs to be encoded, regardless of when the character in question was first used.

Although it should have no impact on the question of whether these characters should be encoded or not, we note that we were asked to help encode the simplified Chinese characters in 漢語大字典 *Hànyǔ Dàzìdiǎn* by several people in China working on an electronic database of characters in this dictionary, who are currently unable to represent these characters in Unicode. There is therefore a real user need to encode these characters, and their encoding should not be blocked for spurious reasons.

#### 1.5 Scanned Dictionary without Further Evidence

As discussed above, dictionary evidence is not only sufficient, but is preferred according to the IRG Proposal Summary Form. Where there are no glyph issues, a single dictionary entry is sufficient evidence for encoding. Therefore the Japanese objections to encoding these 24 G characters and 2 UTC characters should be rejected.

S/N	By	Comment	Response
02648	J	UTC-02648 no reference is given?	Dictionary evidence is provided that states that this character is used as surname. The evidence is sufficient.
02261	J	UTC-01083 no reference is given?	Dictionary evidence is provided that states that this character is equivalent to U+7130 焰. The evidence is sufficient.

### *1.6 Possibly Proper Name Characters*

The 52 UTC characters (not individually listed here) objected to by Japan are characters used as names by members of the imperial family during the Ming dynasty. These were mostly specially invented characters to fit the naming convention for the generation (each generation shares a character with the same radical), and they may have no semantics outside the use as a name character for a particular individual. The individuals with names using these unencoded characters are historically significant, and are mentioned in the historical sources cited as evidence. In order to represent these historical texts in electronic form it is necessary to encode these characters.

In summary, the evidence of use of these characters is sufficient, and there is no reason not to encode them.

### *1.7 Possibly Old Hanzi Transliteration Characters*

There are 84 UTC characters that Japan requests more information for to evaluate the stability of these characters. We consider that it is unnecessary to provide additional evidence as the evidence of use in printed materials is sufficient. We were requested to encode these characters by scholars in the UK and US who are working on oracle bone inscriptions and on early Chinese bamboo and silk texts.

Please see IRGN2103 “UK Activity Report” for an explanation as to why the encoding of these characters is an important activity that we are involved in.

## 2. Unifications

In addition to the unifications and duplicates discussed in the comments below, we have discovered one more duplicate character in the UK submission (IRGN2017):

S/N 05493 (UTC-02592 𠂔齐贝) is a duplicate of U+2CE73 𠂔 (𠂔文 𠂔 𠂔 𠂔 𠂔 𠂔) in Ext. E.

S/N	By	Comment	Response
00002	H	Unify UTC-01671 with U+20006 𠂔	Not cognate and different shape, therefore should not be unified. UTC-01671 is a variant of U+4E32 串, but U+20006 is a variant of U+5343 千.
00003	H	Unify UTC-00969 with USAT05803	Not cognate and different shape, therefore should not be unified. UTC-00969 is the second stage simplified form of U+8650 虐, but U+20006 is a variant of U+4EA1 亡 / U+4EBE 𠂔.
00384	UK	Unify UTC-01600 with U+2A7CA 寡	Withdraw UTC-01600.
00453	H	Unify UTC-01333 with U+209AA 𠂔	Agree.
00553	H	Unify UTC-02957 with U+20C3C 𠂔	Not cognate and different shape, therefore should not be unified. The right side and phonetic of UTC-02957 is U+5E02 𠂔 shì; the right side an phonetic of U+20C3C is U+5DFF 𠂔 fú.
00807	Taichi	Unify UTC-01344 with U+21249 𠂔	Agree.
00986	UK	UTC-01422 is duplicate of Ext. F JMJ-058783	Withdraw UTC-01422.
01182	Taichi	Unify UTC-01005 with U+5F50 𠂔	Not cognate and different shape, therefore should not be unified. UTC-01005 is the second stage simplified form of U+96EA 雪, but U+5F50 is the 'pig snout' radical.
01458	Taichi	Unify UTC-02851 with U+3A30 𠂔	Agree.
01605	H	Unify UTC-01586 with U+2327B 𠂔	Agree.





S/N	By	Comment	Response
01679	J, SAT	Unify UTC-01746 with U+6803 桡	Not cognate, and not exactly the same shape. Most importantly, UTC-01746 is the simplified form of U+6AD4 橈, and a ToTraditional transformation of UTC-01746 should result in U+6AD4. On the other hand U+6803 should not transform to U+6AD4. As UTC-01746 and U+6803 have different responses to transforming from simplified to traditional form characters they should not be unified.
01684	SAT	Unify UTC-01351 with U+3B5F 棊	Agree.
01752	J, H	Unify UTC-02922 with U+2363B 櫟	We are not convinced that UTC-02922 should be unified with U+2363B as UTC-02922 should have kKangXi value of 0547.100, but U+2363B is not mapped to this Kangxi character.
01796	J	Unify UTC-02923 with U+23839 櫟	Agree.
01807	Taichi	Unify UTC-02830 with U+238A7 𣎵	Agree.
02170	J	Unify UTC-02919 with U+7097 茈	We are not convinced that UTC-02919 should be unified with U+7097 as the two characters have different positions in the Kangxi dictionary.
02204	J, H	Unify UTC-01462 with U+24261 𣎵	Possibly.
02343	J	Unify UTC-02856 with U+2C312 𣎵	UTC-02856 is the simplified form of U+2C312. 将 and 將 are not unifiable components, as evidenced by these pairs: U+5D88 蔣 and U+21E83 蔣 (Ext. B); U+93D8 鏘 and U+9535 鏘 (Ext. B); U+646A 擣 and U+2BF63 擣 (Ext. E); U+2B994 擣 and U+2B98B 擣 (Ext. E); U+8E61 踣 and U+2C9C0 踣 (Ext. E).
02391	J	Unify T13-2D6A (and UTC-01584) with U+80B0 𦞙	We agree that UTC-01584 is a variant of U+80B0 𦞙, although we would prefer not to unify the two characters.
02465	Taichi	Unify UTC-01741 with T13-2E21	Agree.

S/N	By	Comment	Response
02855	H	Unify UTC-01867 with U+254E0 礪	The difference between UTC-01867 and U+254E0 is not an omission of a minor stroke, as the simplified 礪 component cannot be written with an extra horizontal stroke above in Chinese orthography. UTC-01867 is the simplified form of U+78A2 礪, but we have been unable to find any evidence that U+254E0 is cognate with U+78A2; it is more probable that U+254E0 礪 is a variant of U+40D2 礪. Therefore, do not unify.
03085	Taichi	Unify UTC-02793 with U+25AF5 笑 or U+25B07 笑	Not cognate and different shape. The 大 component in UTC-02793 is a phonetic, giving the reading tà, meaning a type of food container; whereas U+25AF5 is read tì and means a type of ancient chariot canopy; and U+25B07 is variant of U+7B11 笑 'to laugh' (xiào). Therefore do not unify.
03555	H, K	Unify UTC-01950 with U+82B2 荅	Not cognate, and not unifiable as U+82B2 is a variant of U+82B1 荅; whereas UTC-01950 is the simplified form of U+83D5 蔀. Applying a ToTraditional operation to U+82B2 should result in no change, but applying the same operation to UTC-01950 should result in U+83D5. Therefore it is inappropriate to unify these two characters.
03601	H	Unify UTC-01176 and UTC-01177 as cognate glyphs with the difference of a minor stroke.	The difference between these two characters is 木 'wood' and 禾 'rice plant'; these are two completely different characters, and not minor stroke differences of the same character. Therefore do not unify.
03769	Taichi	Unify UTC-02896 with Ext. F KC-06499	Possibly.
04151	H	Unify UTC-02669 with U+8ECB 輶	Not cognate and different shape. 乙 and 乚 are not unifiable components. UTC-02669 is a surname read jiá, whereas U+8ECB is read yà and means 'to roll' or is a surname. People with the surname Jiá need to be able to distinguish their surname from the surname Yà. Therefore do not unify.




S/N	By	Comment	Response
04374	TCA	Unify UTC-02093 with U+490D 𪛗	Not unifiable. UTC-02093 is the simplified character corresponding to U+490D. The component 𪛗 in UTC-02093 is a simplified component, and not unifiable with the non-simplified component 𪛗, as evidenced by these pairs: U+748A 𪛗 and U+2B7A9 𪛗 (Ext. D); U+4761 𪛗 and U+2C94A 𪛗 (Ext. E); U+93CB 𪛗 and U+2CB6E 𪛗 (Ext. E); and U+29798 𪛗 and U+2CCCF 𪛗 (Ext. E).
04413	UK, TCA	Unify UTC-02829 with U+289B1 𪛗	Withdraw UTC-02829.
05051	TCA	Unify UTC-02343 with U+9B17 𪛗	Not unifiable. This is the simplified character corresponding to U+9B17. The component 𪛗 in UTC-02343 is a simplified component, and not unifiable with the non-simplified component 𪛗, as evidenced by these pairs: U+748A 𪛗 and U+2B7A9 𪛗 (Ext. D); U+4761 𪛗 and U+2C94A 𪛗 (Ext. E); U+93CB 𪛗 and U+2CB6E 𪛗 (Ext. E); and U+29798 𪛗 and U+2CCCF 𪛗 (Ext. E).
05288	TCA	Unify UTC-02412 and UTC-02413	Not unifiable. UTC-02412 and UTC-02413 are non-cognate, with different shapes. They are the simplified forms of two different characters (U+9D20 𪛗 and U+9D21 𪛗 respectively). The components 旦 dàn and 旦 qiě are non-cognate and not unifiable as they are different phonetics.

### 3. *Glyph Issues*

S/N	By	Comment	Response	Corrected Glyph
00193	UK	Incorrect glyph; should be as shown on the right.	Agree. Corrected glyph will be supplied.	 UTC-02805.bmp
01152	H	The evidence shows that the lower component is 癸. A dot is missing in the glyph.	Agree. Corrected glyph will be supplied.	 UTC-01428.bmp
01205	H	It is suggested that the glyph should be modified to reflect that the encircled part comprises two separate vertical strokes.	Agree. Corrected glyph will be supplied.	 UTC-01482.bmp
01703	H	The evidence presents two different glyphs with [left of 改] and 巳 as their middle components respectively. Both are however different from that of the submitted glyph, which has 己 as the middle component.	Agree.	 UTC-01442.bmp

S/N	By	Comment	Response	Corrected Glyph
02827	UK	Incorrect glyph; should be 𪔐石历, as shown on the right.	Agree. Corrected glyph will be supplied.	 UTC-01860.bmp
04505	H	The evidence shows that the first stroke of the right component is 3 rather than 1.	Agree. Corrected glyph will be supplied.	 UTC-02827.bmp
04543	TCA	Font is not correct. Is“𪔐”not“𪔐”	The font glyph is correct, but the wrong evidence for UTC-02842 was provided. The correct evidence for UTC-02842 is shown below.	
04864	TCA	Font & evidence are different.	There is a printing error in <i>Hanyu Da Zidian</i> v. 8 p. 4777, with the character overprinted. The correct glyph for UTC-02322 is shown in <i>Zhonghua Zihai</i> p. 1626 which is already provided as evidence (see IRGN2107R Evidence Fig. 1436), and also shown below.	
05339	TCA	Font & evidence are different.	The wrong evidence for UTC-02455 was provided. The correct evidence is shown below.	

S/N	By	Comment	Response	Corrected Glyph
05458	TCA	Font & evidence are different.	Agree. Corrected glyph will be supplied.	 UTC-02879.bmp

## Corrected Evidence

S/N 04543 (UTC-02842)

*Hanyu Fangyan Da Cidian* p. 7110

𨾏【錫】dā ㄉㄚˊ; tà ㄊㄚˋ

【錫子】〈名〉门或箱上的搭扣。江淮官话。江苏盐城、阜宁。

S/N 04864 (UTC-02322)

*Zhonghua Zihai* (Beijing, 2000) p. 1626 col. A

𨾏“𨾏”的类推简化字。

S/N 05339 (UTC-02455)

*Hanyu Da Zidian* v. 8 p. 4947

𨾏“𨾏”的类推简化字。

#### **4. Other Issues**

<b>S/N</b>	<b>By</b>	<b>Comment</b>	<b>Response</b>
03489	K	UTC-02968 is part of ideographic	Correct. Many ideographic components like this are already encoded as CJK unified ideographs, so there is no reason not to encode UTC-02968.