Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC1/SC2/WG2/IRG N1905

Date: 2012-11-09

Source:	Japan
Title:	Contribution for pending issue of CJK E
Meeting:	IRG #39
Status :	
Actions required	For Discussion
Distribution:	IRG
Medium :	Electronic
Page:	
References:	N1864, N1859

This is a contribution for pending issue of CJK E

(1) scope of item in IRG N954AR and stroke count normalization

The stroke count of components 夂 and 夂 were discussed but no conclusion was drawn. The editorial meeting agreed to remove characters with these components (in CJK_Ev7.0 and v7.1 comments) from CJK_E. The editors were encouraged to prepare written documents for discussion and revision of IRGN954AR.

Information: IRGN954AR

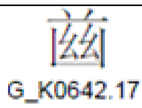
35	夂	丿	4	夂夂

Japan recognizes #35 of IRGN945AR is valid only for "CJK RADICAL RAP"-like shape, on the other hand, at the editorial meeting of IRG#38, HKSAR suggested this to extend to "KANGXI RADICAL GO" and "KANGXI RADICAL GO SLOWLY"-like shape, for example, 00529 ("冬" as a component), 00560("麥" as a component), 01166("峯" as a component), ... so many characters are commented to have wrong stroke count. This is not agreed and all such characters are pended.

00529	冬 00529		冬 V04-4169	SC	SC should be 6 according to item no. 35 in IRGN 954AR and IRGN 1105. KX=0132.201. 0132.071 ㇀ 5| 3		H	postpone
00560	凌 00560		凌 V04-4172	SC	SC should be 12 according to item no. 35 in IRGN 954AR and IRGN 1105. KX=0133.371. 0133.301 ㇇ 11| 1		H	postpone
01166	峰 01166		峰 UTC00562	SC	SC should be 12 according to item no. 35 in IRGN 954AR and IRGN 1105. KX=0209.151.		H	postpone

It is true that no sufficient explanation for #35 of IRG N954AR and ambiguous to interpret but Unihan database and CJKU_SR.txt has a stroke information so this may cause serious inconsistency. In CJKU_SR.txt, Stroke Count (without radical) of "冬" is 3, "凌" is 8 and "棆" is 7.

(2) Unification of「艹」and「䒑」

The character SN00588 (G_K06042.17) 

was removed from CJK_E. China and Japan were invited to prepare written documents for further discussion on unification/dis-unification of SN00588 (G_K06042.17) and U+26C73 

. (See also IRGN1824.)

Since there is no convincing evidences or explanations to unify the components「艹」and「䒑」, the editorial meeting agreed to study the suggested unification cases one by one. We welcome contribution on this issue.

This record is not correct. Japan submitted the explanation to unify these differences *before* IRG#37 (see Japan feedback to CJK E 7.0), however China and TCA repeatedly say to no objection was submitted as a written document during IRG #38. So this issue remains as pending.

J comments on remaining issue of SN00588

Comment on the remaining issue recorded in IRG N1824 (Editorial group report in Meeting #37)

- REMAINING ISSUE:

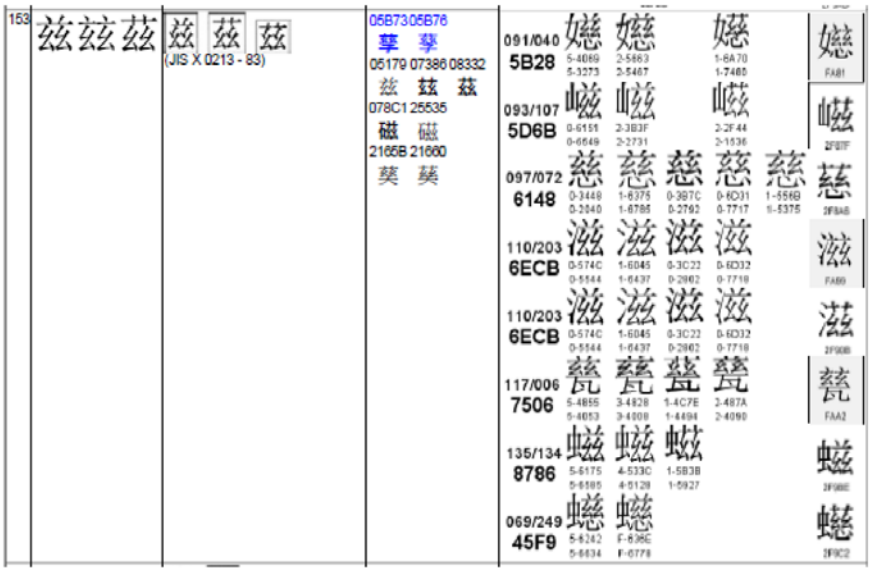
The character SN00588 (G_K06042.17) 茲 G_K0642.17 was agreed (by the editorial meeting) to be coded in CJK_E rather than unified to U+26C73

26C73 艹 140.8 茲 UCS2003 茲 GKX-1042.15 茲 T6-4E38 茲 J4-7644

, and the J glyph 茲 J4-7644 of U+26C73 will be kept in CJK_B.

However, China and TCA proposed to move J4-7644 from U+26C73 to CJK_E. This proposal was postponed for more research.

At IRG #37, Japan compromised to keep G_K0642.17 in CJK E candidates for the moment because China showed a dictionary as a non-cognate character. However, Japan didn't agree to change of the rule. In IWDS we already have #153 in UCV (Unifiable Component Variations).

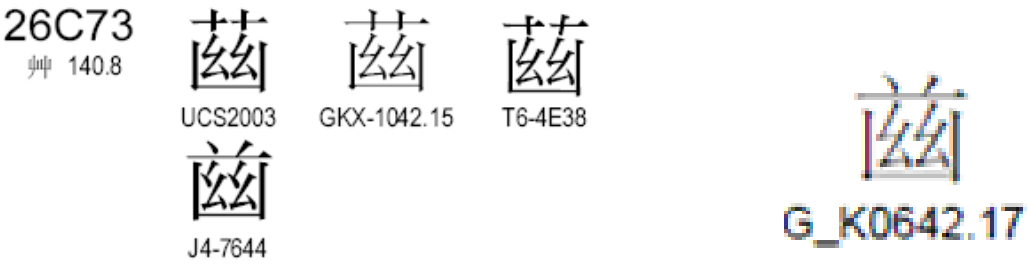


In addition, Japan would like to request China to show the evidence as an IRG document showing G_K0642.17 to separate from u+26c73. Otherwise Japan again insists to unify them.

IRG editors should understand this unification has been made for more than 10 years and applied to several groups of characters which have this difference. Dis-unifying this difference at this moment will cause confusion for CJK Unified Ideographs.

Below is another comment introduced by further study of dictionaries.

In ISO/IEC 10646:2003, there was an unification of the glyphic difference between the component looking like "艹" and one looking like "⺍".



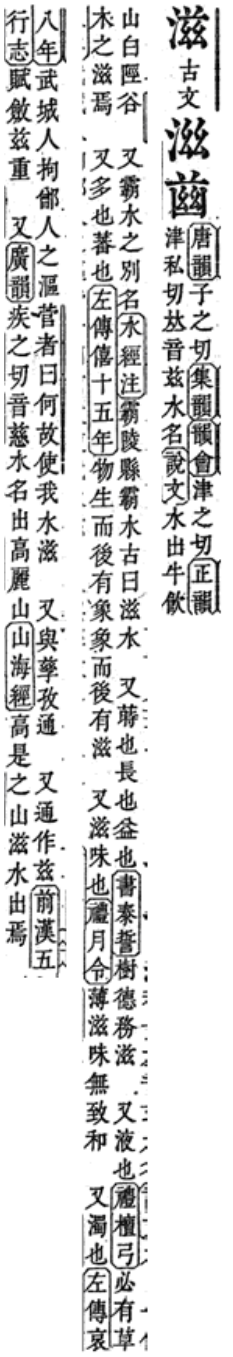
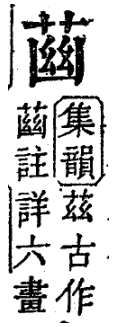
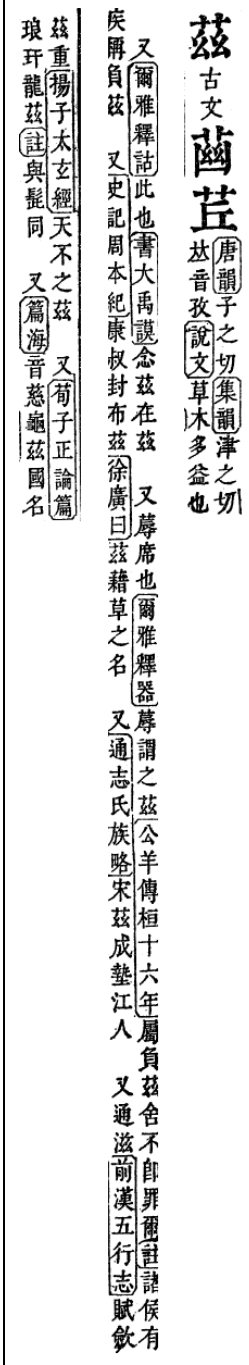
In IRG#37 and #38, there was a discussion that this unification was mistakenly unified, and another character proposed to CJK Unified Ideograph Extension E, G_K0642.17 should be coded at different code position. One reason is that "艹" and "⺍" are semantically different components, and the unification of them may be a barrier against the distinction between semantically different characters. Another reason is that U+26C73 (GKX-1042.15) and G_K0642.17 are shown in different positions in Kangxi Zidian.
In this document, the objections to these 2 points are explained.

For first, the relationship between "艹" and "⺍" is not constant. "艹" is basically the radical meaning "grass" and sometimes confused with eyebrow "⺿". But "⺍" has no basic meaning; it may mean the horn (of sheep), like, "羊", "⺶", etc, or, it may mean the foot (or stop), like, "前", or, sometimes a transformation for easier calligraphy, like "益", "並", "従", etc. Thus, some characters should be distinguished semantically by this component difference, but it is not that the component difference makes the difference always. It should be considered in case-by-case manner, based on the evidence of the submission.

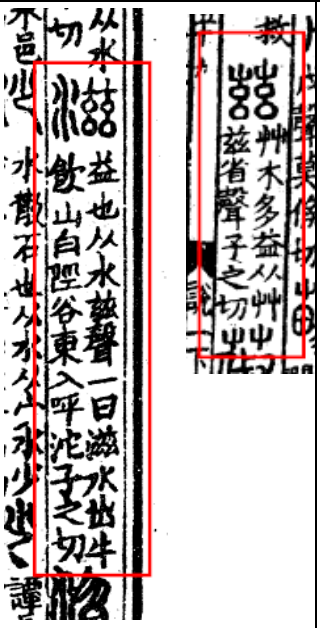
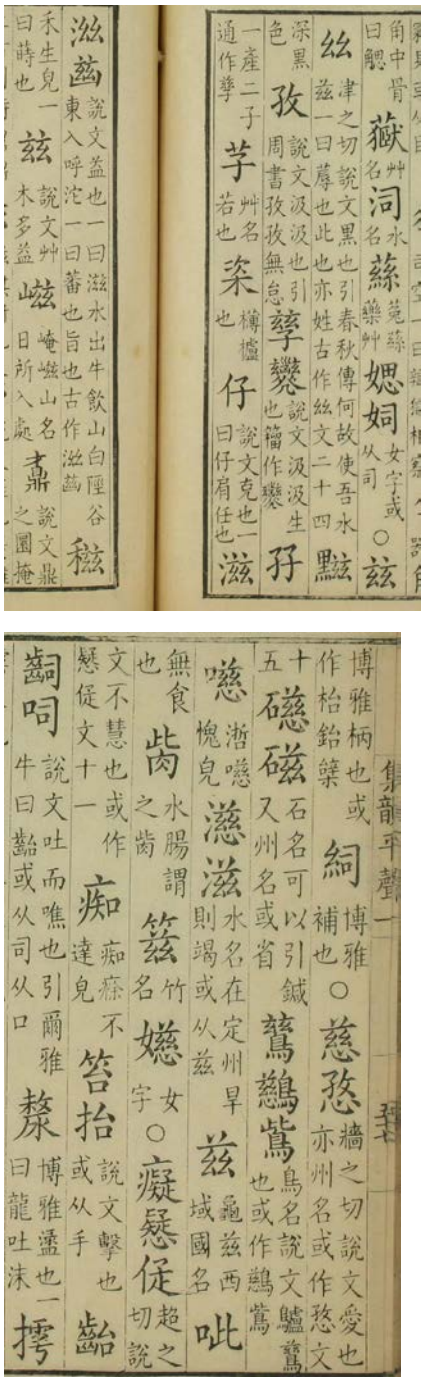
For second, "different position in Kangxi" is insufficient evidence of the semantic difference. G_K0642.17 is shown as an ancient variant form of "滋". The description of "滋" refers various materials and does not clarify which material uses G_K0642.17, but the candidates would be some dictionaries available at that time; 広韻, 集韻, 説文. Both of 説文 and 広韻 do not show an ancient shape for "滋" nor "**玆**", so 集韻 is the only candidate of the source of G_K0642.17.

G_KX1042.15 is shown as an variant form of "**玆**" (Kangxi p.1028), stated that it was taken from

集韻.

G_K0642.17	G_KX1042.15	Kangxi 1028
		

Checking the appearance of G_K0642.17 or G_KX1042.15 character in 集韻 carefully, it is only shown as a variant of "滋" (corresponding to G_K0642.17), not shown as a variant of "**茲**" (corresponding to G_KX1042.15). However, the shape of the variant is similar to G_KX1042.15. It is suspected that there might be 2 mistakes; unrelated character is shown as a variant, and the shape was better in the position where the mistakenly selected.

說文解字(四部叢刊)	広韻(張氏重刊本)	集韻(棟亭藏本)
		

In summary, the evidence to distinguish G_K0642.17 and G_KX1042.15 is not found in the chains of the dictionaries referred by Kangxi. But, as already described, the unification/dis-unification between "艹" and "⺾" should be considered with the case-by-case manner. If other kind of evidence is provided, it would be possible to distinguish a character looking like G_K0642 as un-encoded character, but Kangxi is really questionable if it is an evidence of the semantic difference.

(3) u+21156 (wrong glyph shape in dictionary influences to other character)

China proposed G_K1612.27 to CJK Unified Ideographs Extension E (CJK E), as un-encoded character. In the finalization of CJK E, Dr. Suzuki posted IRG N1859 as a concern that G_K1612.27 was mistakenly designed by Kangxi, and the correct glyph was already coded in U+21156. As a result, it was postponed and excluded from CJK E.

06601	06601	G_K1612.27			Glyph	request for evidence. suspicious shape by mistakenly hand written of	G character	J

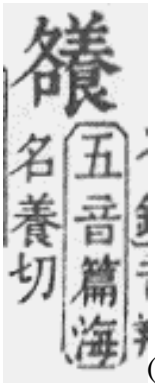
21156 羊 123.15 UCS2003 GHZ-20867.10

During the preparation of CJK F submission, Japan national body found a required character JMJ-032767 similar to G_K16127.27. Japan recognizes these two shapes are not unifiable and u+21156 should keep the current glyph shape to avoid confusion. Japan also expects China to consider if source information of u+21156 need to be modified.

Background

China proposed G_K1612.27 to CJK E. Its shape “各養” looks like as different character from U+21156 shape looking like “名養”.

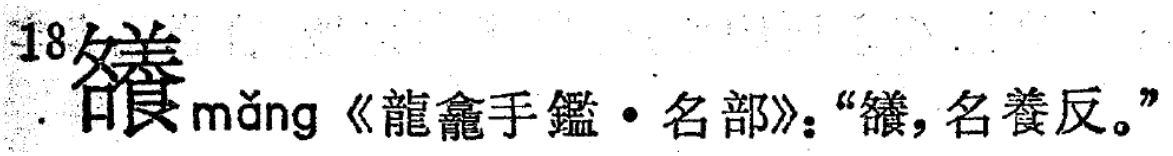
G_K1612.27 (proposed shape)



(Kangxi Zidian p.1612)

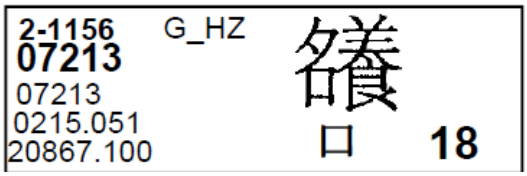
By Backtracking the source of U+21156, it is found that;

- U+21156 is G-single source and taken from Hanyu Da Zidian (漢語大字典).
- The shape in HDZ is different from UCS2003 shape or G-column shape in the CJK B chart in ISO/IEC 10646:2012. The HDZ shape is “各養”.

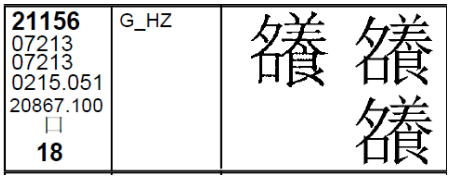


Hanyu Da Zidian (**漢語大字典**) Vol. 2, p. 867.

- The shape in SuperCJKs show “名養” which is different from HDZ shape “各養”.

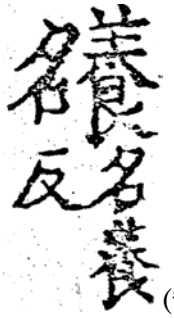
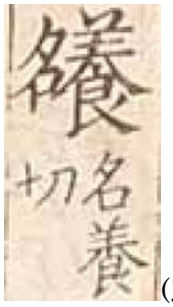


SuperCJK 10.2 (IRG N751)



SuperCJK 14.0 (IRG N802)

- The materials referred by Kangxi or HDZ, Longkan Shoujian (龍龕手鑑) or Wuyin PianHai (五音篇海) show UCS2003 shape "名養".

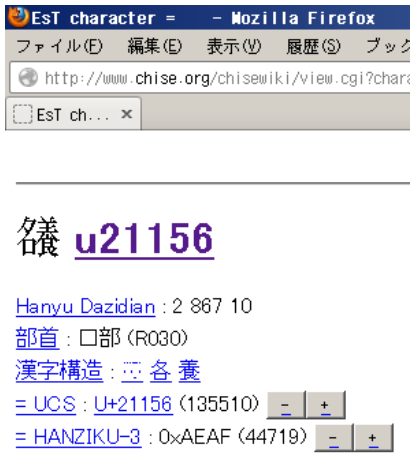

(龍龕手鏡・高麗本)


(五音篇海・萬曆本)

- Also the description of the character "名養切" or "名養反" suggests that the original shape was UCS2003 shape. It is supposed that China expert had submitted "corrected" glyph to CJK B.

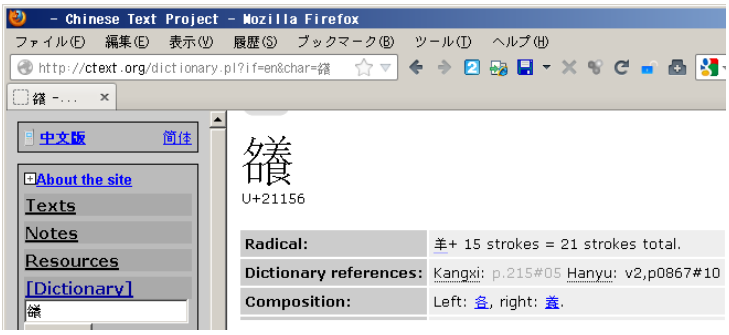
Impact by Encoding G_K1612.27 as New Character

Because the source information of U+21156 is only GHZ-20867.10, there are many documents assuming U+21156 can be used to GHZ-20867.10 or a character at Kangxi Zidian p.1612.27.

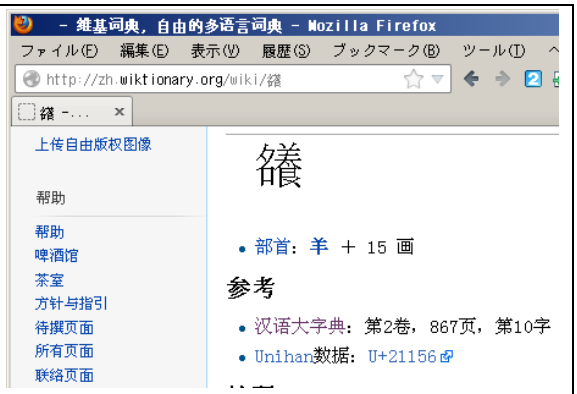


CHISE (Japanese)

Han Morphism System (Japanese)



ctext.org (Chinese)

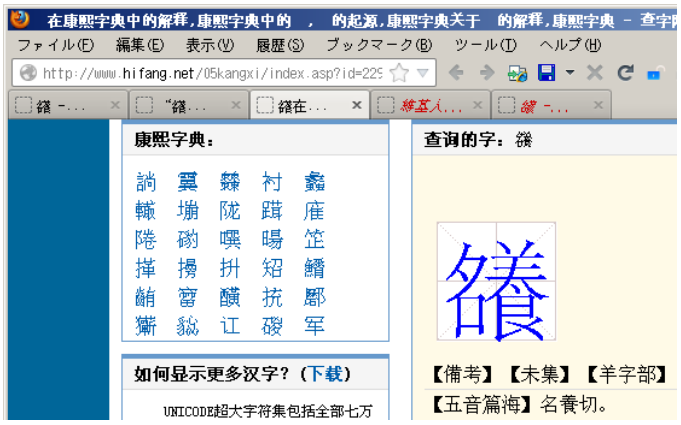


wiktionary.org (Chinese)

Examples of the websites explaining U+21156 as HDZ character

zdic.net (Chinese)

xpcha.com (Chinese)



hifang.net (Chinese)

Wikipedia.org (Chinese)

Examples of the websites explaining U+21156 as Kangxi character

Examples of the dictionary using “名養” shape for G_K1612.27

(康熙字典検索本, 中華書局, 2010, ISBN: 9787101069747)

It seems that the requirement of U+21156 is not so generic and the main usage is the quotation or duplication of HDZ or Kangxi. Considering such situation, it is questionable whether the impact by encoding “各養” as new character is negligible. There is a possibility that the majority of the data using U+21156 are invalidated and requested to change U+21156 to new code point.

Suggestion

Japan understands “名養” and “各養” are not unifiable difference and the glyph shape of u+21156 should not be modified, accordingly Japan keeps JMJ-032767 (“各養” shape) in its CJK F candidates. Japan also requests China to consider if source information of u+21156 needs to be changed to keep inconsistency between glyph shape and source information before finalizing of CJK F work.

(End of document)

Feedback by Jaemin Chung through email dated 14/11/2012 5:03 PM to Rapporteur

Hello, Dr. Lu,
I would like to send my feedback on IRGN 1905.

I do not think it is a good idea to disunify U+26C73. Even if there is a chance of misunification, it is too late to disunify them mainly because of the existing JIS X 0213 implementations.

Moving J4-7644 to another Unicode code point would break existing JIS X 0213 implementations, and this, of course, will cause problems.
JIS X 0213 was first published in 2000, and its JIS X 0213 to Unicode mapping table is completed in 2004, and many vendors support JIS X 0213 using that table since then.

As you may know, there are already several unification and mapping errors on CJK Unified Ideographs Extension B.
http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4173.pdf
But most of those errors are left as they are because changing mappings would break existing implementations.
Instead of moving J4-7644, we can just state that U+26C73 is one of the code points that have unification errors.

Recently, K0-522B and K0-6766 are moved to U+FA2E and U+FA2F from U+F92C and U+F9B8 because U+F92C and U+F9B8 have wrong canonical mappings and canonical mappings cannot be changed. But all KS X 1001 (KS C 5601) implementations have been using U+F92C and U+F9B8 for K0-522B and K0-6766, regardless of their wrong canonical mappings.
Now, since there are U+FA2E and U+FA2F introduced, vendors will have to make changes to their mapping tables. But would this be it? No. All the past data that were using K0-522B and K0-6766 have to be remapped. Also, remember that KS X 1001 is being used since late 1980s; remapping all the data that are piled up for decades cannot be a complete task.
This is why I still think this decision of Korean National Body is not a good idea. We should not repeat this.

Therefore, in my opinion, U+26C73 should be considered as a unification error.

Also, I think changing a mapping is a serious problem, and should not be allowed unless 1) there is no implementation using that mapping, or 2) there is a very urgent problem that must be fixed.

Regards,
Jaemin Chung