ISO/IEC JTC1/SC2/WG2/**IRGN2430EisoReview**
2020-06-13

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

Doc Type:	**Ideographic Research Group Document**
Title:	**Review on Submission of Macao's Vertical Extension (UNC Characters), Horizontal Extension, and IVSes Registration for MSCS**
Source:	**Eiso Chan (陈永聪, Culture and Art Publishing House)**
Status:	**Individual Contribution**
Action:	**For consideration by IRG and Macao SAR**
Date:	**2020-06-13**

This document includes 7 parts.

1. Feedback on 2.1

In the end of 2.1, the proposal shows "MSARG only maintains the glyphs and provides fonts for characters in MSCS and in principle does not maintain the glyphs of the characters in Big-5 and HKSCS as a whole. For characters in Big-5 and HKSCS, MSARG will use the fonts in current computer systems. For IVSes registration of variants, MSARG maintains and provides the glyphs of both variants and base characters since the corresponding base characters should also be registered. Since some base characters are in Big-5 or HKSCS, MSARG will also supply their glyphs for IVSes registration." This paragraph looks ambiguous for the IRG experts, font designers and end users. It is not easy to know what means "the fonts in current computer systems". There are so many default fonts in Windows and MacOS, but different fonts follow different regional conventions.

For the IRG experts, we need to check the glyph design in the encoding review works. If the Macao SAR conventions are ambiguous, it will make the expert hard to do, especially in the IVD/IVS review works. We will get different results in IVD/IVS when Macao SAR chooses the Hong Kong SAR conventions as their conventions or TCA conventions as their conventions. Maybe some pairs of IVSes should be removed, and others should be added.

For the font designers, I know there have been someone or vendors who are waiting to generate a whole set of font for Macao use when MSCS and Macao SAR conventions became stable. The current statements will make the designers be put in a tight spot.

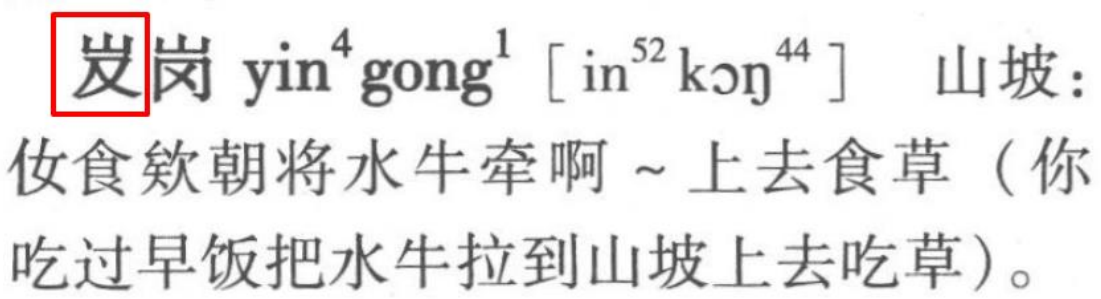
For the end users, something is similar to the font designers. If the supplementary MSCS font follows the TCA conventions, and users use the basic font which follows the Hong Kong SAR conventions, it will make the typography work inconsistent although the baseline would be consistent, and vice versa. As we know, we need to use format 14 (UVS) in the `cmap` table when a font includes IVSes, and under the current policy, the glyph of the basic character must be also registered in IVSes, so the ambiguous conventions will take the fallback work in a

mess.
The Hong Kong SAR conventions have been systematic for the encoding works, and there have been so many fonts to support it, especially after publishing HKSCS-2016. Macao SAR is near Hong Kong SAR, and the people in these two SARs both speak Cantonese as their native dialect, and there are many, even more, exchanges for economics, culture, entertainment, athletics and so on between them from the past to the future, and the Macao SAR has decided to inherit all the characters from HKSCS-2008 and add almost all the new characters from HKSCS-2016 into MSCS as MDH-Source. It looks it is very convenient to inherit the Hong Kong SAR conventions as the Macao SAR conventions.
I will show the details based on the above analysis as below.

2. Feedback on 3.2

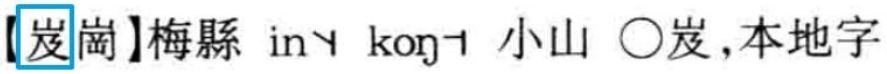
The proposal shows a note on U+5C83 (岁). This is a common Hakka character which means "hillside". I once submitted a document to CLIAC to request adding this character into HKSCS. The most important reason is that it is a place name character used in New Territories.
In the paper *The Origin and Ethnic Identity of Hakka People in Macao* (《澳门客家源流及其族群认同》) written by Yuan Li (袁理) shows that there are 100 thousand Hakka people in Macao, and they are originally from some Hakka habitations in Guangdong Province, such as Meizhou, Xingning, Heyuan, and so on. In the dialect materials in my hand, 岁崗/岁岗 is a common word, which is read as [jin^{52} kɔŋ^{44}] in Meizhou Hakka dialect, but in Meizhou, 岁岗 is always written as 岌岗. 岌 has not been encoded yet, but I trust it is suitable to unify with U+5C83 (岁).

Fig. 2.1 客方言标准音词典, *P. 298*


岌岗 yin^{4} gong^{1} [in^{52} kɔŋ^{44}] 山坡：
仪食欸朝将水牛牵啊～上去食草（你
吃过早饭把水牛拉到山坡上去吃草）。

岌岗 was misread as 岌岗 before Unicode or ISO/IEC 10646 and GBK were used commonly in Meizhou, even in the famous dialect dictionary as below, because so many natives, dialect survey cooperators included, didn't know the regular form of 岌 is U+5C83 (岁). Therefore, the name of the famous bus stop in Meijiang District was used as 赤岌岗, which the real name should be 赤岌岗/赤岁岗. I heard some place names have been changed back to U+5C83 (岁), please see here.

Fig. 2.2 现代汉语方言大词典, *P. 1394*


【岌崗】梅縣 inㄚ kɔŋㄐ 小山 ○岌，本地字

I once tried to search the real uses of U+5C83 (岁) or the variant U+5C7B (岖) in Macao, but failed.
The current character set looks only for the governmental requirements at the current stage, so U+5C83 (岁) is useless for Macao. If Macao experts consider supporting the dialect or other researching use, this character will be needed. On the other hand, MSCS has included some place name characters or person name characters out of Macao for the travelling affairs, for

example, U+2C494 (硝) M(A)C-00111 is a Yue-dialect character used in Foshan and 石硝 is also a famous bus stop in Foshan.

3. Feedback on AppB

In this proposal, Macao SAR submits six unencoded characters which there are two of them have been included in the latest version of IRG WS2017.

3.1. MC-00134

The submitted evidence is unclear, but the same character has been also included in UAX #45 as UTC-00441. Maybe UTC can provide clearer evidence, that will be useful for the IRG encoding works.

I provid one use for MC-00134 in Hong Kong SAR as below. The word “渣喱” or “喳喱” is read as [ʦa^{55} ʦa^{33}] in Cantonese, which means one kind of Nyonya-Malaysia style desserts with five colored beans, tapioca and others, and this Cantonese word is derived from the Malaysian word “bubur cha-cha”. Some studies show the Malaysian word “bubur cha-cha” is derived from a Pali word, but I cannot find out the etymology now.

Fig. 3.1.1 Bubur cha-cha

https://www.malaysianchinesekitchen.com/wp-content/uploads/2016/10/BuburChaCha-1.jpg

Fig. 3.1.2 One use for MC-00134 in Yau Ma Tei, Hong Kong SAR

http://gattin.world.coocan.jp/kanji/025455v.jpg

3.2. MC-00135

According to UCV #337, MC-00135 should be unified with U+8B67 (譧).

Fig. 3.2.1 UCV #337

337	廉 廉	**Disunified Ideographs** 4EB7 廉 廉 廉 5EC9 廉 廉 廉 廉 廉 廉 6FC2 濂 濂 濂 濂 濂 6FD3 濂 濂 濂 濂 71EB 爕 爕 爕 24484 爕 爕 880A 蠊 蠊 蠊 蠊 蠊 274B2 蠊 蠊 942E 鐮 鐮 鐮 鐮 鐮 鐮 4965 鐮 鐮 2214F 幭 幭 幭 22156 幭 幭 24143 灨 灨 24144 灨 灨 27E16 黶 黶 黶 27E1C 黶 黶 29F14 鱸 鱸 29F16 鱸 鱸 **Compatibility Ideographs** 2300A 厳 厳 2F8CA 厳

In Kangxi Dictionary, there are two fanqies under the entry of U+8B67 (譧), one is 離鹽切, the other is 直陷切.

Fig. 3.2.2 Kangxi Dictionary, P. 1183

The submitted evidence shows the Cantonese reading is LIM, which should be [lim^{21}] and related to 離鹽切.
Based on the current framework, MC-00135 should be changed to ME_8B67_001, and added into IVD.

3.3. MC-00137
3.3.1. Glyph
There are two different glyphs under WS2017-02004. Macao suggests unifying MC-00137 to WS2017-02004, which the submitted glyph is as the same as GDM-00085, but the glyph in the evidence is as the same as UTC-02993.
It is better to modify the MC-00137 to match the evidence like UTC-02993.
3.3.2. Unification
Ken suggested disunifying GDM-00085 and UTC-02993 based on the different glyph in V5 review cycle, but Henry and I disagreed.
Two pieces of evidence in IRG WS2017 are related to Foshan City, and the evidence for MC-00137 is also related to the same city, so the unification proposed by Macao should be

accepted.

3.4. MC-00138

The evidence shows it is a character used for a company name, which the address of the company is 澳門青州大馬路美居廣場. It looks easy to get much information of 美居廣場 by Internet. The unencoded character used in the company name looks uncommon, but there is no more other information. It will be not easy to confirm how to solve this character.

The proposal shows the similar character is U+2512E (旺), which is a Nôm character. VNPF shows three readings: nhòm, nhắm, nhằm; TĐCNDG shows five readings: nhăm, nhằm, nhắm, nhầm, nhòm; Kho chữ Hán-Nôm Mã hoá shows four readings: dòm, nhằm, nhắm, nhòm. All in all, the phonetic element for U+2512E (旺) is U+58EC (壬) not others.

Based on the context in the evidence, it looks this character is an error form of U+65FA (旺).

I suggest Macao provide more information or explanation for MC-00138 later, otherwise this character should be postponed.

BTW, MC-00138 is also included in CNS 11643 as 12-402C.

3.5. Code points

Based on the above analysis, one character should be unified, one character should be postponed, and two characters have been included in IRG WS2017, so there are two new characters which are needed to find the encoding slots.

Could we add MC-00134 and MC-00136 into IRG WS2017 or CJK Ext. H like what we did for Macao in CJK Ext. E?

4. Feedback on AppA and AppC_1

AppA includes all the characters in MSCS, and the glyphs for the MC-Source characters which the previous source is MAC-Source have been re-designed.

AppC_1 is related to the M-Source horizontal extensions. I review this part to follow the Hong Kong conventions.

4.1. MD-5C2D glyph and MD-6681 glyph

MD-5C2D is the component of MD-6681, but they have minor differences.

Fig. 4.1.1 MD-5C2D

54	MD-5C2D		U+5C2D

Fig. 4.1.2 MD-6681

69	MD-6681		U+6681

It is better to make the glyph of same component consistent.

4.2. MD-6245 glyph

The outside component of U+6245 (彥) is U+6236 (戶) or U+6237 (户), and Hong Kong SAR chooses U+6237 (户) as their standard glyph.

I suggest modifying the MD-6245 glyph to follow Hong Kong SAR conventions and change the current MD-6245 glyph to ME_6245_001 and add it to IVD.

4.3. MD-67BC glyph

In Hong Kong SAR conventions, the glyph of the bottom component U+6728 (木) is the real U+6728 (木), the last two strokes should be close to the wood body and the last stoke should be right-falling (捺) not dot (點). Please see U+505E (傞), U+558B (喋), U+8776 (蝶) and so on.

Fig. 4.3.1 MD-67BC

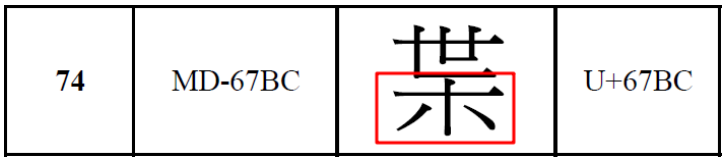


Fig. 4.3.2 U+505E

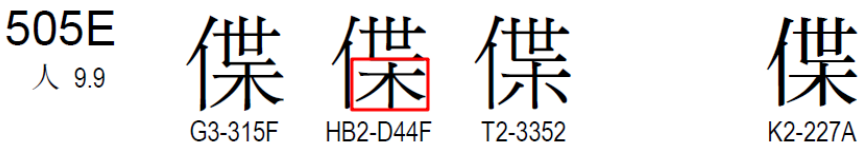


Fig. 4.3.3 U+558B

Fig. 4.3.4 U+8776

4.4. MC-00047 and MD-697D glyph

This is the similar issue to 4.3.

Fig. 4.4.1 MC-00047

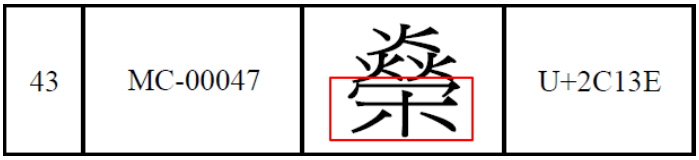
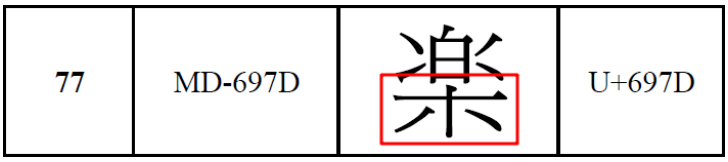


Fig. 4.4.2 MD-697D

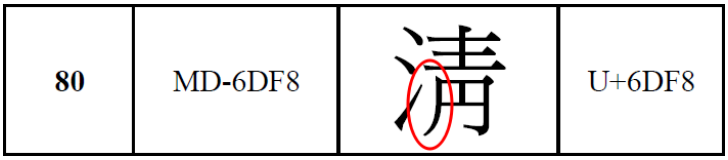


4.5. MD-6DF8 glyph

In Hong Kong SAR conventions, the first stroke of the bottom of U+9751 (靑) is the vertical bar (豎/竪) not the slash (撇).

Fig. 4.5.1 U+9751

Fig. 4.5.2 MD-6DF8



4.6. MD-7399 glyph

In Hong Kong SAR conventions, the first vertical stroke of U+4E0E (与) should be straighter.

Fig. 4.6.1 U+4E0E

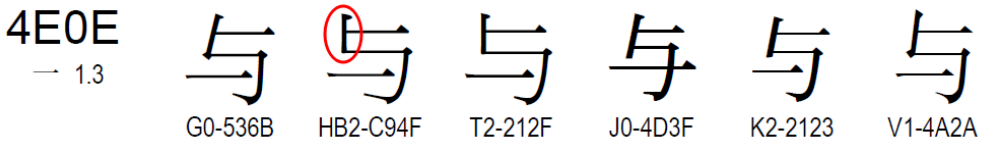
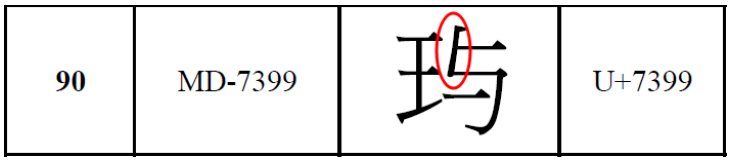


Fig. 4.6.2 MD-7399



4.7. MD-74C7 glyph

In Hong Kong SAR or Macao SAR conventions, the bottom component of U+6182 (憂) should be U+5902 (夂) not U+590A (夊).

Fig. 4.7.1 U+6182

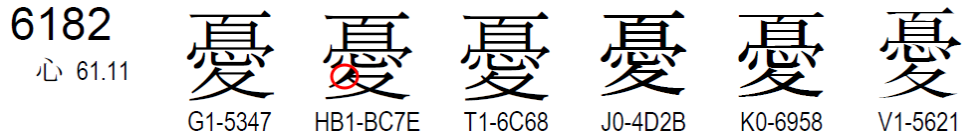


Fig. 4.7.2 U+512A

Fig. 4.7.3 MD-74C7

4.8. MD-784F glyph

In Hong Kong SAR conventions, the second stroke of the middle component should be tí (提/剔).

Fig. 4.8.1 U+5E75

Fig. 4.8.2 U+59F8

Fig. 4.8.3 MD-784F

96	MD-784F	砰	U+784F

4.9. MD-790F glyph

In Hong Kong SAR conventions, the last two strokes of U+696D (業) should be close to the main body and the last stoke should be right-falling (捺) not dot (點).

Fig. 4.9.1 U+696D

Fig. 6.9.2 MD-790F

97	MD-790F	礫	U+790F

4.10. MC-00053 and MD-7A25 glyph

In Hong Kong SAR conventions, the bottom of U+9999 (香) is U+66F0 (曰) not U+65E5 (日), and U+7A25 (香) is the variant of U+9999 (香), so the bottom should be U+66F0 (曰), too.

Fig. 4.10.1 U+9999

Fig. 4.10.2 MC-00053

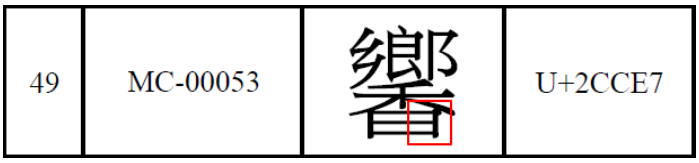
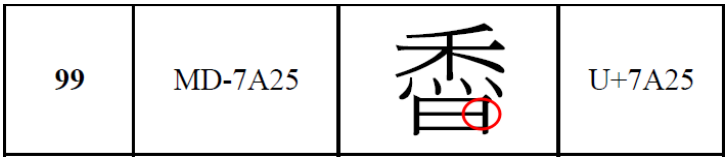


Fig. 4.10.3 MD-7A25



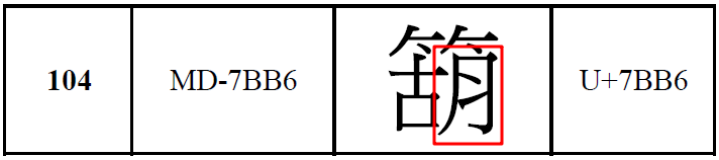
4.11. MD-7BB6 glyph

In Hong Kong SAR conventions, the right part of U+80E1 (胡) should be U+2E9D (月).

Fig. 4.11.1 U+80E1

Fig. 4.11.2 U+846B

Fig. 4.11.3 MD-7BB6



4.12. MD-7D99 glyph

In Hong Kong SAR conventions, the last two strokes of U+7C73 (米) should be close to the main body.

Fig. 4.12.1 U+7C73

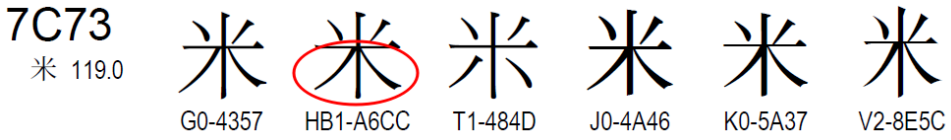
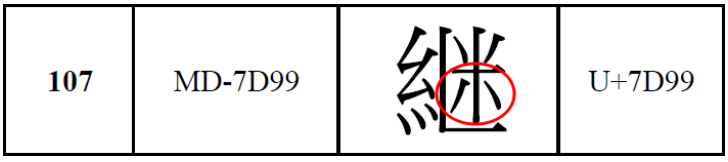


Fig. 4.12.2 MD-7D99



4.13. MD-4058 MD-7E4E glyph

In Hong Kong SAR conventions, the internal two dots of the upper-left component of U+7136 (然) should be parallel.

Fig. 4.13.1 U+7136

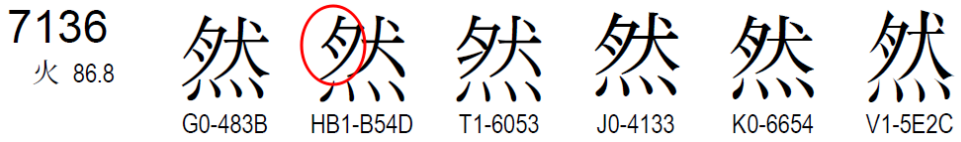


Fig. 4.13.2 U+71C3

Fig. 4.13.3 MD-4058

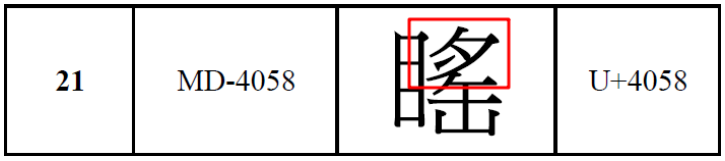
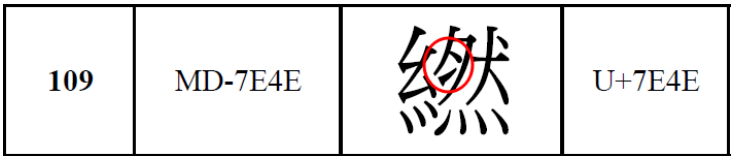


Fig. 4.13.4 MD-7E4E

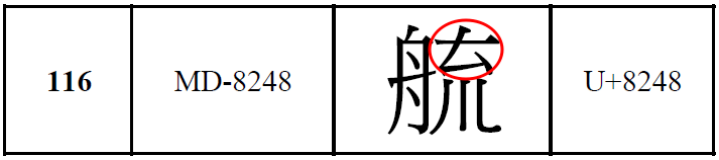


4.14. MD-8248 glyph

In Hong Kong SAR conventions, the top component of U+342C (㐬) is the four-stroke form not the three-stroke form.

Fig. 4.14.1 U+6D41

Fig. 4.14.2 MD-8248



4.15. MD-9759 glyph

In Hong Kong SAR conventions, the first stroke of the bottom of U+9752 (青) is the vertical bar (豎/竪) not the slash (撇).

Fig. 4.15.1 U+9752

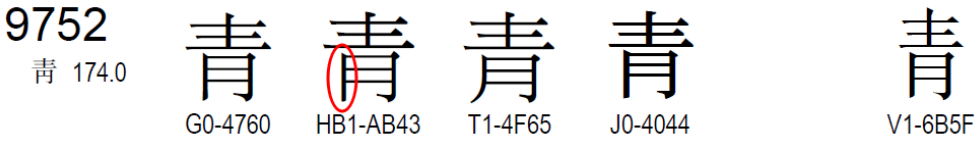
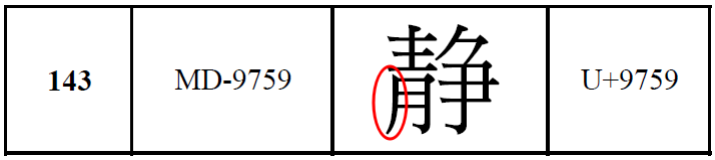


Fig. 4.15.2 MD-9759



4.16. MD-9B30 glyph

In Hong Kong SAR conventions, the bottom left component is different. Notice that U+9B30 (鬱) is the variant of U+9B31 (鬱).

Fig. 4.16.1 U+9B31

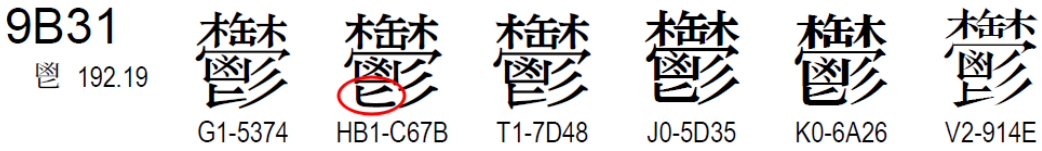
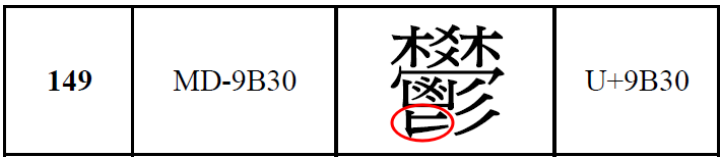


Fig. 4.16.2 MD-9B30

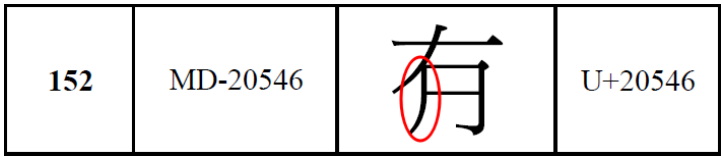


4.17. MD-20546 glyph

U+20546 (𠦂) is the variant of U+20547 (𠦂).

Fig. 4.17.1 U+20547

Fig. 4.17.2 MD-20546



4.18. MC-00012, MC-00015, MC00052, MC-00094, MC-00123, MD-370F MD-216E9 MD-218DD MD-2BC3E, MDH-5AAA glyph

In Hong Kong SAR conventions, the last stroke of the component U+5973 (女) should be shorter when it is set as the left component.

Fig. 4.18.1 U+5974 and U+5975

Fig. 4.18.2 U+5AAA

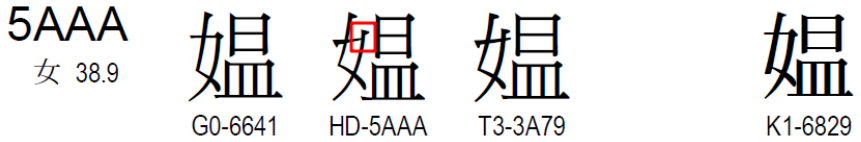


Fig. 4.18.3 MC-00012

11	MC-00012	嬾	U+2A9B4

Fig. 4.18.4 MC-00015

14	MC-00015	孀	U+2BC90

Fig. 4.18.5 MC-00052

48	MC-00052	娕	U+2BC0B

Fig. 4.18.6 MC-00094

73	MC-00094	孀	U+2BC84

Fig. 4.18.7 MC-00123

84	MC-00123	孀	U+2D4A8

Fig. 4.18.8 MD-370F

17	MD-370F	嬪	U+370F

Fig. 4.18.9 MD-216E9 and MD-218DD

159	MD-216E9	𡚩	U+216E9
160	MD-218DD	𡻢	U+218DD

Fig. 4.18.10 MD-2BC3E

183	MD-2BC3E	𢰾	U+2BC3E

Fig. 4.18.11 MDH-5AAA

192	MDH-5AAA	𡚪	U+5AAA

4.19. MD-24327 glyph

This issue is similar to 4.3.

Fig. 4.19.1 U+6851

Fig. 4.19.2 U+55D3

Fig. 4.19.3 MD-24327

167	MD-24327	𤋮	U+24327

4.20. MD-2A9AA glyph

In Hong Kong SAR conventions, the first stroke of the bottom component should be the clear horizontal bar.

And, the left component is related to 4.18.

Fig. 4.20.1 U+7CB5

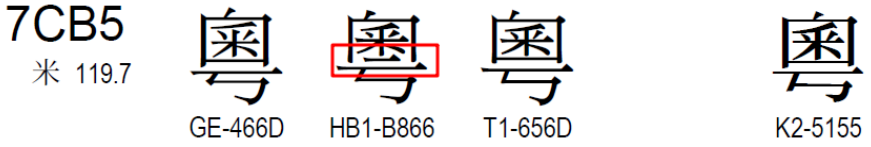


Fig. 4.20.2 MD-2A9AA

180	MD-2A9AA	嫧	U+2A9AA

5. Feedback on AppD

AppD includes the new IVS registered requirements. But now the Macao SAR conventions are ambiguous, it is not suitable to define which one is the "base character". Base character should reflect actual conventions, most unifiable glyphs treated as IVS should be registered as the representative glyph directly.

If Macao SAR conventions follow Hong Kong SAR conventions, the following pairs of IVS could be removed.

U+4058 U+555F U+59F8 U+5C8D U+784F U+7AB0 U+833A U+237C2

And, the glyph for the following pair of IVS should be modified.

U+5029 U+56A4 U+5ACF U+701E U+83C1 U+84A8 U+8534 U+936E U+9759 U+975C

On the other hand, the unifiable glyph for U+8FB6 (辶) used in ME_6A0B_001 and ME_9938_001 is strange.

6. Obsolete MAC-Source reference value

In general, almost all the MAC-Source reference values will be changed to MC-Source reference, but only one is not.

UCS	Char.	Obsolete M ref.	New M ref.
U+21290	坍	MAC-00077	MD-21290

The above information should be written down in the proposal because this is a reference updating issue not a horizontal extension issue.

7. Suggestion

MC-00134 and MC-00136 should be accepted as UNCs, but Macao SAR should provide the explanation or definition of the regional conventions, and then re-check all the details of other parts soon.

(End of Document)