

Date: June. 5, 2007

ISO/IEC JTC 1/SC2/WG2/IRG
Ideographic Rapporteur Group

Source: Taichi KAWABATA, John Knightley, James Kass, Andrew West, Yasuhiro ANAN

Meeting: IRG#28 @ Xi'an

Title: Remaining questionable characters in CJK-C

Keywords: (IDS, Unification)

Status: Expert Contribution

Proposed Conclusion / Requested Action: IRG to review and to make a recommendation to WG2 on CJK C for PDAM5.

Attachments: This document is accompanied by three appendices.

Remaining questionable characters in CJK-C

A-1 : [Unification pattern](#): extra or omission of strokes such as 丿 一 etc;

Component comparison :

夸(𠃉大𠃉一𠃉)---(𠃉大𠃉)

- Unifiable candidate

U+2177B		TF-336E,,,,;		
---------	---	--------------	--	--

U+2A988	媵	TC-553A	kua	Character used in Person's name
---------	---	---------	-----	---------------------------------

N.B.

✓ WG2#50 unified the above pair.

A-2 : [Unification pattern](#): extra or omission of strokes such as 丶 一 etc;

Component comparison : 凡 vs 几

● Unifiable candidate

U+21B36	尋	G_HZ;;;;;KP1-3E BB;		
U+2AA06	尋	TC-556C	xun	Character used in Person's name

A-3 : [Unification pattern](#): 31a, 丶, 乙

Component comparison :

● Unifiable candidate

U+22045	𠄎	G_KX:T6-4836;;;; H-9B4B;;		0327.180 20987.060
U+2AA83	𠄎	TD-2623	xun	Character used in Person's name

A-4 : [Unification pattern](#): 139e

Component comparison :

● Unifiable candidate

U+24814	獼	G_KX;T6-5524;;;; KP1-5846;		0716.200 21362.140 Unified counter part of U+2F927.
U+2F927	獼 2F927	T6-4B42;;;;		Compatibility variant of U+24814.
U+2AEEF	獼	HC100898		同“獼”。A kind of monkey.

N.B.

- ✓ WG2#50 decision: HC100898 comes from non-cognate source thus different from
U+24814 獼 which has a compatibility code in different glyph (U+ 2F927).

A-5 : [Unification pattern](#):

Component comparison :

- Unifiable candidate

U+7A6A	穨	GE-3644;T3-5A67 ;J1-514D;K2-4F2 7;;;KP1-62B4;	CHENG1	kDef: to praise
U+2B079	穨	TE-3C4F	ni; er	Character used in Person's name

A-6 : [Unification pattern](#):

Component comparison :

- Unifiable candidate

U+25BDC	箕	G_KX;T5-4F41;;;;; ;		0891.100 52991.100
U+2B0B6	箕	TD-5274	wei	Character used in Person's name

A-7 : [Unification pattern](#):

Component comparison :

- Unifiable candidate

U+8380	農	G5-682C;T3-3850; J1-5823;K2-5766;; ;KP1-6DED;	CHEN2	to farm; a farmer; agriculture
U+2B25C	農	TD-452A	Chen; nong	Character used in Person's name

A-8 : [Unification pattern](#): 129-4, 62-8, 65a

Component comparison :

- Unifiable candidate

U+8F68	軌	G0-396C;;;;;;	GUI3	track, rut, path Simplified form of 軌 U+8ECC.
U+2B43F	軌	CH401798;UTC0 0058		“軌”的简化字。意思是： 小车车杠前端与车衡相衔接。Simplified form of " 軌".

A-9 : [Unification pattern](#):

Component comparison :

- Unifiable candidate

U+8F68	軌	G0-396C;;;;;;	GUI3	track, rut, path Simplified form of 軌 U+8ECC.
U+2B43F	軌	CH401798;UTC0 0058		“軌”的简化字。意思是： 小车车杠前端与车衡相衔 接。Simplified form of " 軌".

A-10 : [Unification pattern](#): extra or omission of strokes such as 丶 一 etc;

Component comparison :

- Unifiable candidate

U+51E1	凡	G0-3732;T1-443B; J0-4B5E;K0-5B6 D;V1-4C69;;KP0- DDD8;	FAN2	all, any, every; ordinary, common
U+2A702	𠂔	TC-214E	fan	Character used in Person's name

A-11 : [Unification pattern](#): extra or omission of strokes such as 丶 一 etc;

Component comparison :

- Unifiable candidate

U+4EA2	亢	G0-3F3A;T1-446 D;J0-5036;K0-797 1;V1-4A5B;;KP0- F2F1;	KANG4 GANG1	high, proud; violent, excessive; skilled; name; KXI 0088.040 HYD 10280.010
U+2A721	𠂔 [G]	K0422.03; TC-225E	[G] [T]kang	Character used in Person's name KXI: 0422.03

	尢			
	[T]			

N.B.

✓ This pair is not unified because the G-source refers to the different KXI.

A-12 : [Unification pattern](#):

Component comparison :

- Unifiable candidate

U+90A6	邦	G0-306E;T1-4B5 A;J0-4B2E;K0-5B 40;V1-696E;;KP0- DDC8;	BANG1	nation, country, state; KXI 1269.050 HYD 63756.140
U+2B497	邦 [G] 邦 [T]	XC201924; TC-2D59	[G] [T]bang	[G] “邦”的简化字形式, 意思是①国。②姓。 Simplified form of 邦. [T] Character used in Person's name

--- end of document ---

Appendix A

In Annex S there are many examples of a difference of one small stroke between two characters with the same horizontal shape. A method to find these is:-

- (1) Look at the reading or meaning of a proposed character.
- (2) Look for already encoded characters with the same reading or meaning but only one small stroke different.

Type One -- A difference of a dot or small stroke

Annex S examples:-

1.a inside 国 U+56EF vs 国 U+56FD

1.b on edge 戾 U+623B vs 戾 U+623E

1.c touching 鄉 U+90F7 vs 鄉 U+9109

Example candidates

1. U+2A8C3 (TC-2B79) sheng vs 圣

72	05496	圣	TC-297B	sheng	Character used in Person's name
----	-------	---	---------	-------	---------------------------------

2. U+2A8BA (TC-297B) sheng, G_ZJW00193 ?? vs 𡗗

99	05552	𡗗	TC-2B79	sheng	Character used in Person's name
----	-------	---	---------	-------	---------------------------------

3. U+(TD-3F3B) luo vs 𡗗

305	13076	𡗗	TD-3F3B	luo	Character used in Person's name
-----	-------	---	---------	-----	---------------------------------

Type Two - A difference horizontal line inside a character

Annex S example:-

皞 U+76A1 vs U+76A5 皞

Example candiates

1. U+2A7F5 (TC-2D27) zhi vs U+76F4 𡗗

116	03211	𡗗	TC-2D27	zhi	Character used in Person's name
-----	-------	---	---------	-----	---------------------------------

IRGN1232_C1_evidences_TCA-Part1.pdf also shows the traditional version of the above TC-3834.

232	03243	𡗗	TC-3834	zhi	Character used in Person's name
-----	-------	---	---------	-----	---------------------------------

Appendix B

Several G_HC sources need to be rechecked. In particular U+2B29E does not seem to be

in Hanyu Dacidian, only the very similar (and probably unifiable) U+452D 膏 .

U+2A81B (G_HC300170) @ HYDCD vol.1 p.946

definition: same as U+5EE8 懈 jiè xiè

Codepoint	Character	Source Ref.	Hanyu Dacidian 漢語大詞典	Reading & Definition	Notes
U+2A81B	懈	G_HC300170	vol.1 p.946	= U+5EE8 懈 jiè xiè	

U+2ABD8 (G_HC100598 vol.6 p.962

reading: jià

definition: “駕駛”

U+2ABD8	駕	G_HC100598	vol.6 p.962	jià -- "駕駛"	
---------	---	------------	-------------	-------------	--

U+2ADA5 (G_HC200793) @ HYDCD vol.5 p.1213

reading: wāng

definition: name of a Korean prince

U+2ADA5	汪	G_HC200793	vol.5 p.1213	wāng -- name of a Korean prince	
---------	---	------------	--------------	---------------------------------	--

U+2AEDF (G_HC100308) @ HYDCD vol.5 p.3

reading: bá

Already unified with U+72AE 𠂔 bá

U+2AEDF	𠂔	G_HC100308	vol.5 p.3	bá	= U+72AE 𠂔 bá
---------	---	------------	-----------	----	---------------

U+2AEEF (G_HC100898) @ HYDCD vol.5 p.96

definition: same as U+7328 猿 yuán

U+24814 獼 = U+733F 猿 yuán, so not cognate.

U+2AEEF	獼	G_HC100898	vol.5 p.96	= U+7328 猿 yuán	N.B. U+24814 獼 = U+733F 猿 yuán, so not cognate.
---------	---	------------	------------	-----------------	---

U+2B29E (G_HC101428) -- not in HYDCD ?!

Presumably a variant form of U+452D 甯 nìng = U+85B4 寧 nìng. Phonetic

component may be written as any of 甯 寧 甯 甯 甯 .

甯 may be unifiable with 甯 .

U+2B29E	甯	G_HC101428	NOT FOUND	Presumably a variant form of U+452D 甯 nìng = U+85B4 寧 nìng. Phonetic component may be written as any of 甯 寧 甯 甯 甯 . 甯 may be unifiable with 甯 .
---------	---	------------	-----------	---

U+2B36B (G_HC101560) @ HYDCD vol.10 p.1383

definition: same as U+9B3A 觴 shāng

U+2B36B	觴	G_HC101560	vol.10 p.1383	= U+9B3A 觴 shāng	
---------	---	------------	---------------	------------------	--

U+2B4A9 (G_HC101951) @ HYDCD vol.10 p.694

reading: zī

definition: ancient family name

U+2B4A9	鄮	G_HC101951	vol.10 p.694	zī -- ancient family name	
---------	---	------------	--------------	---------------------------	--

U+2B4AC (G_HC201953) @ HYDCD vol.10 p.696

definition: same as U+913D 鄮 chán

U+2B4AC	鄮	G_HC201953	vol.10 p.696	= U+913D 鄮 chán	
---------	---	------------	--------------	-----------------	--

U+2B556 (G_HC202193) @ HYDCD vol.12 p.20

reading: chèn

definition: “撑”

U+2B556	𠵹	G_HC202193	vol.12 p.20	chèn -- “撑”	
---------	---	------------	-------------	-------------	--

U+2B566 (G_HC102194) @ HYDCD vol.12 p.125

definition: same as U+95A6 𠵹 chù

Dunhuang variant

U+2B566	𠵹	G_HC102194	vol.12 p.125	= U+95A6 𠵹 chù	Dunhuang variant
---------	---	------------	--------------	----------------	------------------

U+2B5CB (G_HC102277) @ HYDCD vol.12 p.213

definition: same as U+97B4 𨮑 bài bèi bù fú

cf. U+29341 𨮑 and U+29358 𨮑 which are also variants of U+97B4.

U+2B5CB is probably a confusion of 𨮑 and 𨮑

U+2B5CB	𨮑	G_HC102277	vol.12 p.213	= U+97B4 𨮑 bài bèi bù fú	cf. U+29341 𨮑 and U+29358 𨮑 which are also variants of U+97B4. U+2B5CB is probably a confusion of 𨮑 and 𨮑
---------	---	------------	--------------	--------------------------	---

U+2B67F (G_HC102550) @ HYDCD vol.12 p.733

definition: same as U+9AE7 髧 dàn

May be unifiable.

U+2B67F	髧	G_HC102550	vol.12 p.733	= U+9AE7 髧 dàn	May be unifiable.
---------	---	------------	--------------	----------------	-------------------

U+2B684 (G_HC102559) @ HYDCD vol.12 p.744

definition: same as U+9B10 髧 qí

cf. U+4C07 髧 which is also a variant of U+9B10. Not unifiable because of different layout of components.

U+2B684	髧	G_HC102559	vol.12 p.744	= U+9B10 髧 qí	cf. U+4C07 髧 which is also a variant of U+9B10. Not unifiable because of different layout of components.
---------	---	------------	--------------	---------------	--

U+2B687 (G_HC102568) @ HYDCD vol.12 p.752

definition: same as U+9B18 髧 mán

cf. U+29BBA 髧 which is another variant of U+9B18. U+2B687 and U+29BBA

may be unifiable.

U+2B687	髧	G_HC102568	vol.12 p.752	= U+9B18 髧 mán	cf. U+29BBA 髧 which is another variant of U+9B18. U+2B687 and U+29BBA may be unifiable.
---------	---	------------	--------------	----------------	---

U+2B6C7 (G_HC102588) @ HYDCD vol.12 p.1268

definition: same as U+9BD4 𩺰 zī

U+2B6C7	𩺰	G_HC102588	vol.12 p.1268	= U+9BD4 𩺰 zī	
---------	---	------------	---------------	---------------	--

--- end of document ---

Appendix C : CJK C hard to resolve IDS

Type-I : Characters which IDS representations contain “?” .

Type-II : Characters which IDS representations contain one or more overlapping IDC “□” .

Type-III : Characters which IDS representations contain more than three IDC characters.

This does not necessarily mean the character shape in question is weird, but we could make use of them as part of a systematic review process, which would complement IDS review and increase our confidence.

The rest of characters are supposedly regular in structure and they would have fewer problems assuming that submitters have thoroughly reviewed them.

Type I : Characters which IDS representations contain “?” or non-Ideographic components.

I-1 : U+2A700

Source : V04-4021

IDS : ?

𠄎			
---	--	--	--

I-2 : U+2A9B2

Source : zjw00872

IDS : □女?

媼		同“祢”。	
---	--	-------	--

I-3 : U+2AA39

Source : TC-2A68

IDS : 𠂔 ? 山

𠂔	shan	Character used in Person's name	
---	------	---------------------------------	--

U+2AA18 (TC-2A67) 𠂔 : 𠂔尸 𠂔一山

shan; Character used in Person's name

I-4 : U+2AC0B

Source : JMK65202

IDS : 𠂔 を 方

𠂔	HAKATA	Characters appear in other documents	
---	--------	--------------------------------------	--

I-5 : U+2AC49

Source : TD-286B

IDS : ?

鼎	ding	Character used in Person's name	鼎 DING3 ZHEN1; large, three-legged bronze caldron
---	------	---------------------------------	--

I-6 : U+2B1B9

Source : TD-4374

IDS : 𠂔 耳 𠂔 由 𠂔

U+31C9

CJK STROKE SZWG

聘	pin ping	Character used in Person's name	聘 PIN4 PING4 engage, employ; betroth
---	-------------	------------------------------------	---

I-7 : U+2B513

Source : zjw01754

IDS : □金?

鎏		同“敦”	
---	--	------	--

Type II : Characters which IDS representations contain one or more overlapping IDC “□”.

II-1 : U+ 2A704

Source : zfy00582

IDS : □ㄥ ㄣ 一 口 ㄩ

𣎵	the phonetic 'gā';	the def 'Throw';	(軋 GA2; 釵 GA2)
---	--------------------	------------------	----------------

II-2 : U+ 2A706

Source : zjw00022

IDS : □木□、、井

𠂇		同“深”	
---	--	------	--

II-3 : U+ 2A7CC

Source : TC-5261

IDS :   百一分; Alt : (寡 - 宀)

寡	gua	Character used in Person's name	寡 GUA3
---	-----	------------------------------------	--------

II-4 : U+ 2A7EF

Source : TC-2F3F

IDS :  匚  | 巨

匚	chen	Character used in Person's name	臣
---	------	------------------------------------	---

II-5 : U+2A848

Source : JMK65063

IDS :  王 𠂇

𠂇	wani	Character used in Person's name	
---	------	------------------------------------	--

II-6 : U+2A8DB

Source : TC-4574

IDS : 日廿口土; Alt : (堇- 一);

堇	jin	Character used in Person's name	堇 JIN3; yellow loam; clay; season; few
---	-----	---------------------------------	--

II-7 : U+2A93D (and this is also Type III)

Source : BK100304

IDS : 日夕口水口天口

畚			
---	--	--	--

、十五祀小臣餘犀尊、小子畚、邲其诸

II-8 : U+2AA14

Source : TC-2137

IDS : 冫 丩 丩

尢	wang	Character used in Person's name	尢 WANG1 YOU2; weak;
---	------	---------------------------------	------------------------

092/034 尢 尢 尢 尢
5C22 0-5E4C 1-4452 0-5577 2-2E2F
0-6244 1-3650 0-5387 2-1415

尢

2F875 (T4-2134;,,,,)

II-9 : U+2AA4E

Source : TC-563B

IDS : 𠄎山 壘; alt : 𠄎山 壘;

壘	wu	Character used in Person's name	
---	----	------------------------------------	--

II-10 : U+2AACF

Source : TC-223F

IDS : 𠄎 戍; alt : (戍- 丩);

戍	wu	Character used in Person's name	戍 WU4; 5th heavenly stem
---	----	------------------------------------	-----------------------------

II-11 : U+2AAD0

Source : TC-2378

IDS : 𠄎 夷人; alt : (夷- 一);

夷	Yi	Character used in Person's name	夷 YI2; ancient barbarian tribes
---	----	------------------------------------	---------------------------------------

II-12 : U+2AADF

Source : TC-3B33

IDS : 𠄎 𠄎 十口 參

參	zhen	Character used in Person's name	
---	------	------------------------------------	--

II-13 : U+2ACCF

Source : TD-296A

IDS : 𣎵木𣎵𠄎月

𣎵	yong	Character used in Person's name	𣎵 YONG2
---	------	------------------------------------	------------

II-14 : U+2AD2C

Source : TE-2D3A

IDS : 𣎵木𣎵𣎵一曲豆

𣎵	cao	Character used in Person's name	𣎵 CAO2 ZAO1; trough; manger; vat, tank; groove; a distillery
---	-----	------------------------------------	---

II-15 : U+2AED6

Source : TC-4336

IDS : 𣎵𣎵玄𠄎牛

𣎵	qian	Character used in Person's name	𣎵 QIAN1 QIAN4; drag, pull, lead by hand
---	------	------------------------------------	---

II-16 : U+2AEDF

Source : HC100308

IDS : 𣎵犬𠄎

𠂔	bá	2.通“拔”。拔除。 Pulling out. KXI 0706.010	𠂔 ^[G] 𠂔 ^[T] BA2;to pull up KXI 0706.010
---	----	--	---

13547		𠂔	𠂔
0706.010	3		HC100308
𠂔	1		

II-17 : U+2B04D

Source : zjw01368

IDS : 𠂔 日 𠂔 目 | 示

𠂔		8.4130 : “𠂔(敖)叔 微~于西宫”。	
---	--	----------------------------	--

II-18 : U+2B120

Source : TD-5352

IDS : 𠂔 糸 𠂔 𠂔 月

繡	xiu	Character used in Person's name	繡繡 XIU4 embroider; embroidery; ornament
---	-----	------------------------------------	--

II-19 : U+2B129

Source : TD-6239

IDS : 𠂔 糸 𠂔 𠂔 月、

繡	xiu	Character used in Person's name	繡繡 XIU4 embroider; embroidery; ornament
---	-----	------------------------------------	--

U+2B120 (TD-5352)

繡 糸 申月

II-20 : U+2B1CD

Source : V04-5172

IDS : 月 十、

肺			
---	--	--	--

II-21 : U+2B25D

Source : JMK65592

IDS : 廿 一 隶

叢	Suge	Characters appears in the old dictionary (Shinsen-Jikyo)	
---	------	---	--

II-22 : U+2B4ED

Source : JMK65809

IDS : 金 厶 口 儿

鏡	nata	Characters appears in the old dictionary (Shinsen-Jikyo)	
---	------	---	--

II-23 : U+2B5FF

Source : V04-567E

IDS : ☐☐几二末

𠄎			(☐几二) Vietnamese simplification of 風风
---	--	--	---

II-24 : U+2B600

Source : V04-5724

IDS : ☐☐几二彼

𠄎			
---	--	--	--

II-25 : U+2B601

Source : V04-5726

IDS : ☐俞☐几二

𠄎			
---	--	--	--

II-26 : U+2B602

Source : V04-5728

IDS : ☐☐几二美

𨮑			
---	--	--	--

II-27 : U+2B603

Source : V04-572A

IDS :

𠄎

𠄎

 几二海

𨮒			
---	--	--	--

II-28 : U+2B604

Source : V04-572C

IDS :

𠄎

𠄎

 几二退

𨮓			
---	--	--	--

II-29 : U+2B605

Source : V04-572E

IDS :

𠄎

𠄎

 几二退

𨮔			
---	--	--	--

II-30 : U+2B606

Source : V04-5730

IDS :

𠄎

𠄎

 几二愈

characters.

III-1 : U+ 2A778

Source : zjw00715

IDS : 𠂇 今 𠂇 米又 𠂇 又 匕

𠂇		人名用字。People's name.	
---	--	---------------------	--

III-2 : U+2A881

Source : V04-436C

IDS : 𠂇 口 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 一 八 𠂇

𠂇			𠂇 口 舉
---	--	--	-------

III-3 : U+2AEC9

Source : zjw01279

IDS : 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇

𠂇		5.2532: “乃~子乍 (作)厥文考尊彝”。	
---	--	--------------------------	--

III-4 : U+2B2BB

Source : JMK65632

IDS : 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇 𠂇

	ori	Characters appears in the old dictionary (Shinsen-Jikyo)	
---	-----	--	--

III-5 : U+2B41B

Source : V04-5526

IDS :  足  | |  ㄥ、見

			
---	--	--	--

III-6 : U+2B43E

Source : zjw01645

IDS :   車田  𠂔 幺又  車田

		同“申”。	
---	--	-------	--

III-7 : U+2B47C

Source : V04-5557

IDS :  辵  | | 𠂔 一 八 𠂔

			 辵 舉
---	--	--	---

III-8 : U+2B56F

Source : V04-5650

IDS :    | | 𠂔 一 八 𠂔 門

舉門			舉門
----	--	--	----

--- end of document ---