IRG N2174 Part 1A

Title: Possible T-source Glyph Mistakes

Author: Henry Chan

Source: Henry Chan, Ken Lunde

Type: Individual Contribution to IRG #47

•••

Issue

The G-source Glyph for U+8669 may be in error:

U+8669's original source GB 7589-87 65-49:

The same left hand component is also found in characters such as 隙 㵾 螉 䃬.

-- End of Document

IRGN2175 Part 1B

Title: Misunification of U+53FD

Source: Henry Chan

Type: Individual Contribution to IRG #47

Date: 2016/10/12

==

Introduction

The Unification of u53FD G-source and T-source is likely in error.

Description

The G-source (G0-5F34) is ▯口几 while the T-source (T4-216A) IDS for 叽 is ▯口儿.

The G-source character is the simplified variant of 嘰 and is pronounced as jī.

The pronunciation for T-source is listed as jiào, which comes from the 教育部罕用字.

It references the Kangxi dictionary as being the variant of 叫:

 (Cropped from http://kangxizidian.com/kangxi/0173.gif)
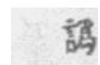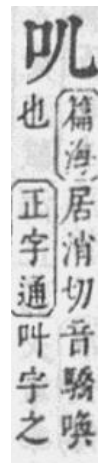
Suggested Action

The two characters (􏰀口儿 vs 􏰀口几) are non-cognate and should be disunified.

As U+53FD is in IICORE with value CG, indicating that the character for G-source is used frequently, it may be better to disunify T4-216A to a new codepoint.

IRG N2174 Part 1C

Title: Radical for U+266B9 should be moon not meat

Source: Henry Chan

Type: Individual Contribution to IRG #47

Date: 2016/10/12

....

Radical for U+266B9 should be moon, instead of meat.

Entry in Hanyu Dazidain:

-- which indicates the meaning is related to "moon"
(http://web.archive.org/save/_embed/http://www.guoxuedashi.com/kangxi/pic.php?f=dzd
&p=2215)

And the entry in Kangxi:

(http://www.kangxizidian.com/kangxi/0506.gif)

Entry in CNS11643:

全字庫字型	發 音		
朚 朚 朚 明體 楷體 宋體	《ㄥˋ 男聲 女聲		
拼音	漢語 gèng		
字義	筆畫	部首	倉頡
相關詞	10	月	BMVM
部件	月,互,部件版本(CNS11643_2_94)		
筆順序	ノ,ㄱ,ㄧ,ㄧ,ㄧ,ノ,ㄱ,丶,丶,ㄧ,筆順序版本 (CNS11643_1_94)		
漢字國屬性	台灣(T5-2E24) 中國(G_HZ)		

Entry in MOE Dictionary:

異體字
朏 胆

&補充資料

字號	C04975	正字	【朏】月-06-10
音讀	ㄍㄥˋ		
釋義	月出。亦作「朏」。見康熙字典‧月部。朏字引篇海。 浟「亙」之異體。		

異體字表

05696
朐亙
月
9
13
朐

http://dict.variants.moe.edu.tw/yitic/frc/frc04975.htm

-- End of Document

IRG N2174 Part 1D

Title: Misunification of U+2F8FA to U+6C4E

Source: Henry Chan

Type: Individual Contribution to IRG #47

Date: 2016/10/15

•••

Issue

The mapping from U+2F8FA to U+6C4E is wrong.

According to CNS11643 website, U+2F8FA is pronounced as wan2 with U+2007D 𠁽 (=== U+4E38 丸 wan2) as the phonetic component:

全字庫字型	發 音
汎 汎 汎 明體 楷體 宋體	ㄨㄢˊ 男聲 女聲
拼音	漢語 wán

Meanwhile, U+6C4E is pronounced as fan4 with U+51E1 凡 (fan2) as the phonetic component:

全字庫字型	發 音
汎 汎 汎 明體 楷體 宋體	ㄈㄢˋ 男聲 女聲
拼音	漢語 fàn

Which is a variant for 氾／泛

By etymology, U+2F8FA could have been unified with U+6C4D 汍 instead.

Action Items

Possible Actions to take:
- Remove the source T6-252E from U+2F8FA
- Add T6-252E to IRG Working Set
- Unify to U+6C4D, or encode new character in Extension G / URO.

Other Information

There exist two other disunified characters with etymology U+2007D 凡 / U+4E38 丸:

4E38						
丶 3.2	丸	丸	丸	丸	丸	丸
	G0-4D68	HB1-A459	T1-443A	J0-345D	K0-7C2F	V1-4A38

2007D			
丶 3.2	凡	凡	凡
	UCS2003	GKX-0080.14	T6-2132

9AAA				
骨 188.3	骰	骰	骰	骰
	GE-4643	T4-4221	J1-6956	K1-6966

9AAB				
骨 188.3	骰	骰	骰	骰
	G3-796D	HB2-E0E9	T2-486F	K2-6F54

The shape of U+2007D 凡 (wan2) is too easily confused with 凡 (fan2), as evident with U+9AAA. The phonetic of the two characters may be similar in Mandarin but are completely different in Cantonese: jyun2 vs faan4

Possible Treatments:
For future, either (1) add 凡 / 丸 to UCV; or (2) see 凡 as "corrupted" form of 丸 (in reality, the shape of U+2007D is closer to the ancient shape of 丸), and ask for withdrawal if 凡 / 丸 is identified as its phonetic component.

For current glyph, either (1) correct / normalize the G-, T- and K- glyphs for U+9AAA to use 凡 instead of 凡; or (2) leave the forms as is.

-- End of Document

IRG N2174 Part 2 A

Title: Code-point for Full-width Separator Dot (音界號 / 間隔號)

Source: Henry Chan

Type: Individual Contribution to IRG #47

Date: 2016/10/12

==

Introduction

There are currently two code-points used for the separator dot in Chinese punctuation: U+00B7 Middle Dot and U+2027 Hyphenation Point. However, the fullwidth/halfwidth property of these two codepoints are not defined and implemented consistently. These two codepoints are also used in other Latin-based scripts, which is expected to be proportional in those contexts.

Description

For PRC, GB2312 maps the separator dot to U+00B7. For SimSun and Kaiti, U+00B7 is rendered as full-width.

For Taiwan, default operating system fonts do not handle this consistently. U+2027 should be used according to CNS11643, and should be full-width, but 新細明體 displays as proportional, while 標楷體 displays as full-width.

For HK, the big-5 mapping should be to U+00B7, however the neither MSungHK nor LiSung render the code-point as full-width.

In these case of mixed script use (such as Han/Eng), applications must implement complex language detection combined with specific fonts to show the full-width or half-width version in the correct context.

Action Item

In order to ensure that full-width and half-width separator dots are displayed correctly, a different codepoint should be used for the full-width version, as with other CJK Punctuation. It is suggested that the existing "U+30FB Katakana Middle Dot" be unilaterally used for this purpose.

This concerns the mapping of GB2312 0xA1A4, Big5 0xA145, and CNS 1-2126.

Supplementary Information

The request to use U+30FB was also raised at Unicode PRI 316 (http://www.unicode.org/review/pri316/) by Bobby Tung from Taiwan.

(Sometimes the U+FF0E full-width full stop is also used for the separator dot, however, it should be considered as misuse due to its name.)

IRG N2174 Part 2 B

Title: Design for 龜 (qiu1) component

Source: Henry Chan

Type: Individual Contribution to IRG #47

Date: 2016/10/12

●●●

The right hand side for 秋 (龝, 穐, 稝, 龝, 穐, 穐, 龝, 龝, 龝, 龜) and the like have always been thought to be 龜. However, recent oracle bone script studies indicate that the right hand side is not a turtle but likely locusts instead.

The confusion has been evidenced inside shuowen where the ancient form was erroneously drawn with a turtle head instead of a locust head.

Recently some scholars have "modernized" the oracle bone form into 龜 :

(remarks in blue)

At the same time, 䶛 is also encoded at U+2A4D4.
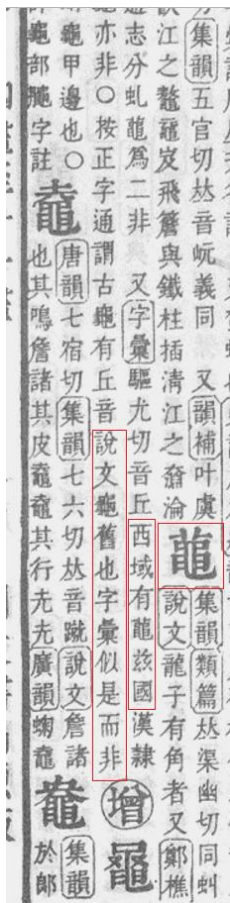
Which also gives a pronunciation similar to 秋 and also quotes <<字彙>> to say that it is the correct form of the 龜 in "龜兹". (though author of Kangxi Dictionary disputes this by referencing shuowen, which has been proved wrong.)

Possible Action items:

A1) Encode 䲰 as a new character; or

A2) Unify it with 𪓔 (U+2A4D4); or

A3) Do not encode

B1) Allow changing of all characters with shape 龜 but with etymology of 秋 to use 䲰 instead (i.e. addiction to UCV)

B2) Do not add to UCV and encode the new variants with 䲰 only when they appear in other (reputable) text.

-- End of Document

Title: China Responses to IRGN2170 and IRGN2174
Date: 2016-10-20
Source: China

The G-source Glyph for U+8669 may be in error:

The same left hand component is also found in characters such as 隙 㴩 蟓 隙

<mark>Agree</mark>

 Hanyu Da Zidian(漢語大字典)
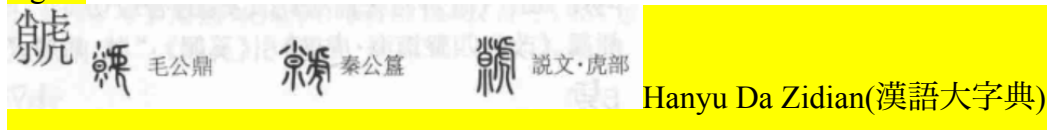
The Unification of u53FD G-source and T-source is likely in error.

Description

<mark>Agree</mark>

Radical for U+266B9 should be moon, instead of meat.

<mark>Agree</mark>

According to the evidence, the two should be separated.

1. Error G-source glyph of U+3B9D

The current G-source glyph of U+3B9D is error:
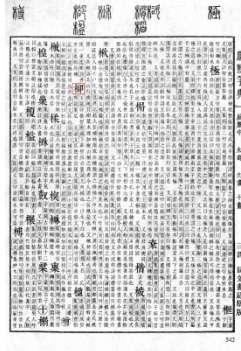
The T source and H-source of U+228C1 have different relative positions of components, but they are encoded at one code point:

According to ISO/IEC 10646,they should be encoded separately, so I suggest to disunify U+228C1 and give a new code point of T-source(TF-5A27).

Agree.

Title: Design for 𪚏 (qiu1) component

The right hand side for 秋 (穐, 龝, 秐, 䁬, 穐, 穐, 龝, 齝, 龝, 蟗) and the like have always been thought to be 龜. However, recent oracle bone script studies indicate that the right hand side is not a turtle but likely locusts instead.

Encode 𪚏 as a new character. And 龜 unifies to 𪚏.

1. 𪚏 is not a normal writing.

The author of this book 《甲骨文字典》頁783：徐中舒主編，四川辭書出版社，1989 also writes 𪚏 （U+9F9C, tortoise） likes 龜.

Comparing with 龜 as the transcription of , should be transcripted to 𪚏 but not

龜 冈 is the normal writing but 目 not.

In fact, 𦳈 has been transcripted to 龜 by Tang Lan（唐蘭） who firstly pointed that

龜 means autumn(秋) in 1934.

Furthermore, 𦳈 is written as 龜 which is the part of 龜 龜 in clerical script(Li style,隸書). 龜 is coded as 𥣨 （U+258E8） which is form Kangxi Zidian (康熙字典)：
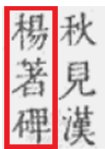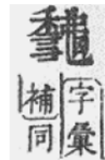
冈 is the normal writing but 目 not.

And 龜 also is written as 龜龜龜 in clerical script(Li style,隸書).

2. 蟉(U+2A4D4) in the ancient documents is not the variant with 秋.

蟉 同"虯"。《廣雅·釋魚》："有角曰蟉龍。"王念孫
疏證："虯與蟉同。"《集韻·幽韻》："虯,《説文》：'龍子
有角者。'或作蟉。"

　qiú 《廣韻》渠幽切,平幽羣。幽部。
❶传说中的有角龙。《説文·虫部》："虯,龍子有角者。"

　qiū 《廣韻》七由切,平尤清。幽部。
❶庄稼成熟。《説文·禾部》："秋,禾穀孰也。"

In the modern pronunciation system, 蟉 pronounces *qiú* and 秋 pronounces *qiū*. Both of them pronounce *qiu*. But their pronunciations are different in the ancient pronunciation

system.

qiú 《廣韻》渠幽切，平幽羣。幽部。 The initial is *g*.

qiū 《廣韻》七由切，平尤清。幽部。 The initial is *ts'*.

Generally speaking, a word pronounced *g* cannot pronounce *ts'*. Although many contrary examples can be found, these contrary examples are very credible. But there is no other evidence to prove that 䜌 can pronounce *ts'* as 秋.

In my opinion, 䜌 is just form which is the variant of 虯, but not form .

龍子有角者

3. But sometimes 䵶(U+2A4D4) can be used as variant with .

Some researchers should use䵶(U+2A4D4) as the transcription of . Sometimes, the distinction between 龜 and 黽 is not very obvious. For example, 鼇=鱉.

(𪓰) also is coded as 𥤠（U+25920） and 𪔁（U+2A501）.

For more example, 囗(U+211A5) is the variant with 囚(U+56DA, prison) and also is the transcription of (disaster or worry).