Doc Type: **Working Group Document**

Title: **Comments on four error glyphs on CJK Unified Ideographs Ext B & E**

Source: **Eiso Chan (陈永聪), Victor Wong**

Author: **Eiso Chan (陈永聪)**

Status: **Individual Contribution**

Action: **For consideration by JTC1/SC2/WG2/IRG and UTC**

Date: **2017-01-25**

1. Possible error V-source glyph for U+2017B

The IDS of the current UCS2003 glyph for U+2017B is ⿰亡涓, but the IDS of the current V-source glyph for U+2017B is ⿰亡湏. (pls see Fig. 1) 涓 shouldn't be unifiable with 湏.

*Fig. 1 U+2017B on The Unicode Standard, Version 9.0*

We found this character in the offical website of VNPF, Quốc Ngữ is quên. (pls see Fig. 2)

Fig. 2 U+2017B on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
quên	湏	U+2017b	quên lãng	vhn	to fall/sink into oblivion

Fig. 3 U＋2017B on Kangxi-Plus Unicode Han Dictionary

张玉书, 陈廷敬, 王宏源 等 (Zhāng Yùshū, Chén Tíngjìng, Wáng Hóngyuán, et al): 《康熙字典（增订版）》(*Kangxi-Plus Unicode Han Dictionary*), 北京: 社会科学文献出版社 [Běijīng: Social Sciences Academic Press (CHINA)], 2015, ISBN 978-7-5097-7093-1, p. 22
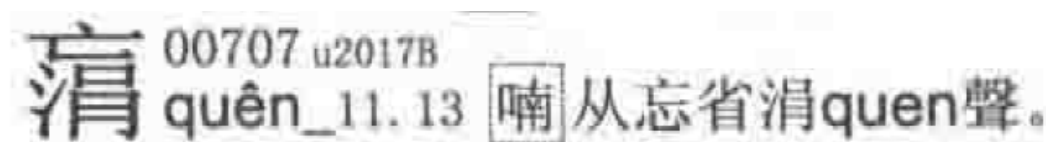
According to Fig. 3, the semantic component (形符) of U＋2017B is 忘 which means to fall or sink into oblivion (pls see Fig. 4), and the other part is the phono component (声符). Quốc Ngữ of 湨 is vờn (pls see Fig. 5), and 涓 is quen, quên, quyên (pls see Fig. 6), so we think V-source glyph for U＋2017B may be wrong.

Fig. 4 U＋5FD8 on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
vong	忘	U+5fd8	vong ân	vhn	to prove ungrateful

Fig. 5 U＋6EB3 on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
vờn	湨	U+6eb3	chờn vờn	vhn	to cruise

Fig. 6 U＋6D93 on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
quen	涓	U+6d93	quen biết	vhn	be acquainted with
quên	涓	U+6d93	quên đi	btcn	to forget it!
quyên	涓	U+6d93	quyên (dòng nước nhỏ): quyên trích (giọt nước nhỏ)	btcn	small stream

We suggest three ways to solve this problem:

A1) Encode V2-6E46 as a new character, and the kRSUnicode value for the new character is 8.14.

A2) Add a new U-source glyph into U+2017B. The evidence of the new U-source ideograph is

Fig. 3.

Table 1 Information for a New Proposed U-source Ideograph

Glyph	IDS	Radical	Radical Form	Strokes	First Stroke	Simplified Form	Variant Form	Variant Code Point	U-source Number
凊	⿱亡涓 U+2FF1, U+4EA1, U+6D93	8	亠	11	5	N/A	N/A	N/A	UTC-xxxxx OR UCI-xxxxx

B1) Keep the present situation of U+2017B, and add a second kRSUnicode value of 8.14 in the

future.

B2) Add two glyphs into SVS like this:

Table 2 Proposed SVSes for U+2017B

Base Character	Variation Selector	Glyph	Description
U+2017B	FE00 (VS1)	凊	UCS2003
U+2017B	FE01 (VS2)	凊	V2-6E46

C1) Keep the present situation of U+2017B, and add a second kRSUnicode value of 8.14 in the

future.

C2) Add two glyphs into IVD like this:

Table 3 Proposed IVSes for U＋2017B

Base Character	Variation Selector	Glyph	Description
U＋2017B	E0100 (VS17)	�	UCS2003
U＋2017B	E0101 (VS18)	瀆	V2-6E46

2. Possible error V-source glyph for U+204AF

The IDS of the current UCS2003 glyph for U+204AF is ⿰肉兒, but the IDS of the current V-source glyph for U+204AF is ⿰肉充. (pls see Fig. 7) 兒 shouldn't be unifiable with 充.

Fig. 7 U+204AF on The Unicode Standard, Version 9.0

We found this character in the offical website of VNPF, Quốc Ngữ is nhóc. (pls see Fig. 8) The glyph on VNPF is as the same as UCS2003, not V0-4422. The supportive font of VNPF is Nom Na Tong Regular V4.2 produced by VNPF themselves now. (pls see Fig. 9, and the software used in the figure is BabelMap produced by Andrew West)

Fig. 8 U+204AF on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
nhóc	𠒯	U+204af	nhóc con, thằng nhóc	vhn	kid, brat

Fig. 9 The Information about Nom Na Tong Regular V4.2
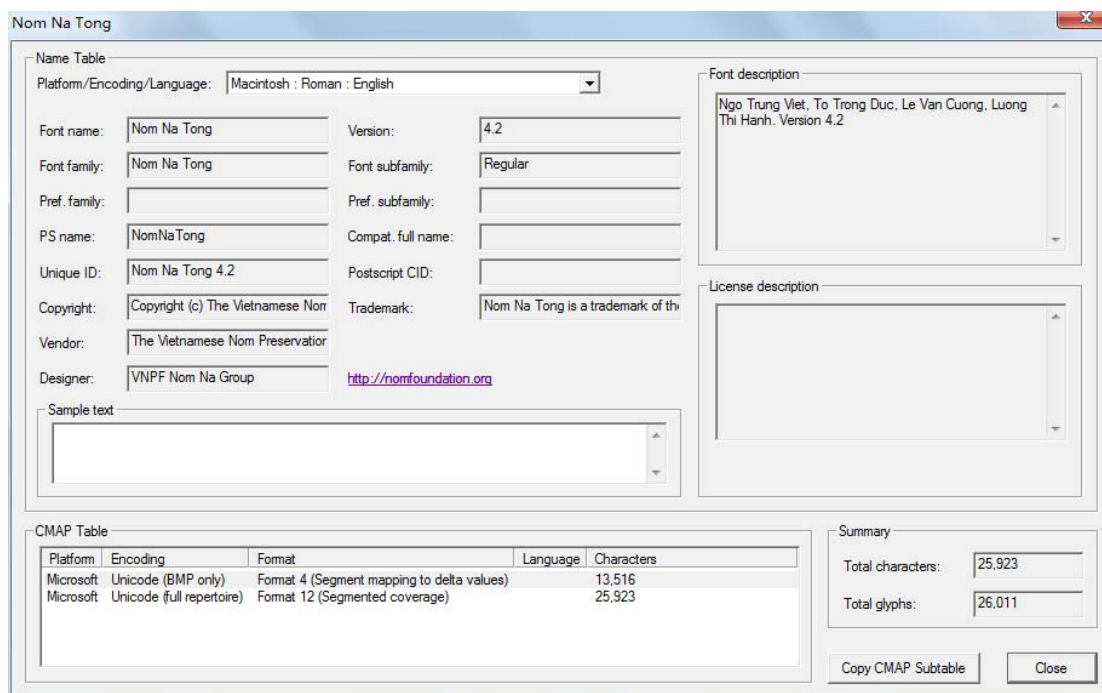
We think the phono component of U+204AF is 肉, because Quốc Ngữ of 肉 is nhục which is similar with nhóc (pls see Figs. 8 & 10), and the other part is the semantic component. According to VNPF, U+204AF means kid and brat, and it's similar with 兒 (pls see Fig. 11), not 充 (pls see Fig. 12). So we think V-source glyph for U+204AF may be wrong.

Fig. 10 U+8089 on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
nhục	肉	U+8089	cốt nhục	vhn	bone and flesh

Fig. 11 U+5152 on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
nhi	兒	U+5152	thiếu nhi, bệnh nhi	vhn	child
nghê	兒	U+5152	ngô nghê	gdhn	foolish
nhẻ	兒	U+5152	nhỏ nhẻ	gdhn	son, child, oneself; final part

6

Fig. 12 U+5145 on VNPF

Quốc Ngữ	Hán-Nôm	Codepoint	Context	Ref.	English
sung	充	U+5145	sung sức	vhn	in good form, fully fit

We suggest four ways to solve this problem:

A) Change the V-source glyph to ⿱肉兒 like the glyph of Nom Na Tong Regular V4.2. (pls see

[IRGN1784 VietnamResponse](#) or Figs.13 & 14)

Fig. 13 IRGN1784 VietnamResponse

| U+204AF | V | 204AF 儿 10.12 | 㝮 UCS2003 | 㝮 V0-4422 | J: glyph not revised G: and not unifiable, need V glyph to change back. Already mentioned in IRGN1775_A4 (J comments). | V-font is not correct correct font as UCS2003 | 㝮 |

Fig. 14 The Right V-source Glyph for U+204AF

⿱肉兒

B1) Encode V0-4422 as a new character, and the kRSUnicode value for the new character is

10.10.

B2) Add a new U-source glyph into U+204AF. We feel sorry not to find out a suitable evidence

at this time.

Table 4 Information for a New Proposed U-source Ideograph

Glyph	IDS	Radical	Radical Form	Strokes	First Stroke	Simplified Form	Variant Form	Variant Code Point	U-source Number
⿱肉兒	⿱肉兒 U+2FF1, U+8089, U+5152	10	儿	12	2	N/A	N/A	N/A	UCI-xxxxx

C1) Keep the present situation of U+204AF, and add a second kRSUnicode value of 10.10 in the

future.

C2) Add two glyphs into SVS like this:

Table 5 Proposed SVSes for U+204AF

Base Character	Variation Selector	Glyph	Description
U+204AF	FE00 (VS1)	奐	UCS2003
U+204AF	FE01 (VS2)	奐	V0-4422

D1) Keep the present situation of U+204AF, and add a second kRSUnicode value of 10.10 in the future.

D2) Add two glyphs into IVD like this:

Table 6 Proposed IVSes for U+204AF

| Base Character	Variation Selector	Glyph	Description
U+204AF	E0100 (VS17)	奐	UCS2003
U+204AF	E0101 (VS18)	奐	V0-4422

3. Error UCS2003 glyph for U+24CB2

The current UCS2003 glyph for U+24CB2 is error, it's different from G-source & T-source glyph for U+24CB2. (pls see Fig. 15)

Fig. 15 U+24CB2 on The Unicode Standard, Version 9.0

We show the evidence in one of the original sources: GHZ-42549.08 which is in the 《汉语大字典》. (pls see Fig. 16)

Fig. 16 GHZ-42549.08 on Hànyǔ Dàzìdiǎn V1
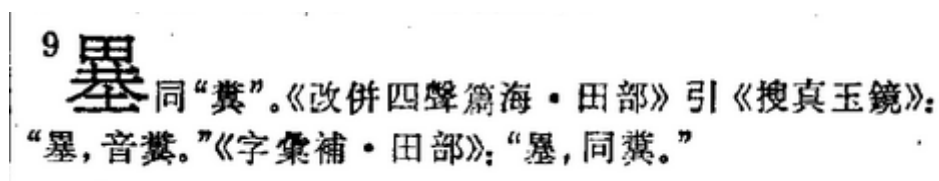
The glyph in 《汉语大字典》V2 is the same. (pls see Fig. 17)

Fig. 17 GHZ-42549.08 on Hànyǔ Dàzìdiǎn V2

汉语大字典编辑委员会(Hànyǔ Dàzìdiǎn Biānjí Wěiyuánhuì): 《汉语大字典（第二版）》(*Hànyǔ Dàzìdiǎn V2*), 武汉: 湖北长江出版集团崇文书局 (Wǔhàn: Chongwen Publishing House of Hubei Changjiang Publishing Group) & 成都: 四川出版集团四川辞书出版社 (Chéngdū: Sichuan Reference Press of Sichuan Publishing Group), 2010, ISBN 978-7-5403-1744-7, p. 2729
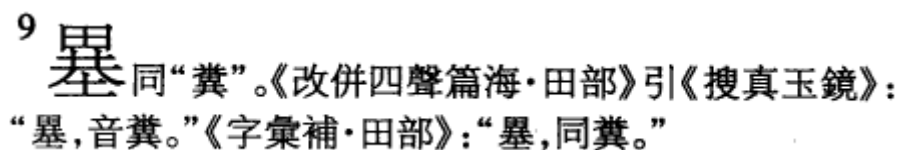
We have found more evidences about this glyph.

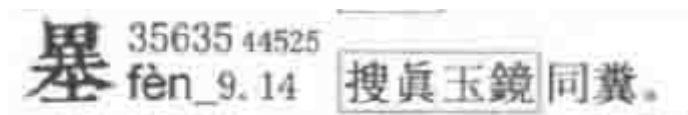*Figs. 18 through 20 GHZ-42549.08 on different versions of KangXi Dictionary*

内府本《康熙字典》（午集備考）, p. 40

同文書局版《康熙字典》（午集備考）, p. 113

张玉书, 陈廷敬, 王宏源 等 (Zhāng Yùshū, Chén Tíngjìng, Wáng Hóngyuán, et al): 《康熙字典（增订版）》(*Kangxi-Plus Unicode Han Dictionary*), 北京: 社会科学文献出版社 [Běijīng: Social Sciences Academic Press (CHINA)], 2015, ISBN 978-7-5097-7093-1, p. 993

On the other hand, we found the evidence for the character 𡕢日坴, too. (pls see Fig. 21)

Fig. 21 𡕢日坴 on Zhōnghuá Zìhǎi

冷玉龙, 韦一心 等 (Lěng Yùlóng, Wěi Yīxīn, et al): 《中华字海》 (*Zhōnghuá Zìhǎi*), 北京: 中华书局 (Běijīng: Zhonghua Book Company) & 北京: 中国友谊出版公司 (Běijīng: Friendship Publishing Corporation of China), 1994, ISBN 7-5057-0630-6/H·1, p. 838

We suggest four ways to solve this problem:

A1) Withdraw UCS2003 glyph for U+24CB2 like U+2384F, and keep the kRSUnicode value of

102.9.

A2) Encode ⿱日㚓 as a new U-source ideograph, and the kRSUnicode value for it is 72.9. The

evidence of the new U-source ideograph is Fig. 21.

Table 7 Information for a New Proposed U-source Ideograph

Glyph	IDS	Radical	Radical Form	Strokes	First Stroke	Simplified Form	Variant Form	Variant Code Point	U-source Number
暮	⿱日㚓 U+2FF1, U+65E5, U+212CD	72	日	9	2	N/A	叁	U+21284	UTC-xxxxx OR UCI-xxxxx

B1) Encode GHZ-42549.08 & T6-5F26 as a new character, and change the kRSUnicode value for

U+24CB2 of 72.9, and the kRSUnicode value for the new character is 102.9.

B2) Add a new U-source glyph into U+24CB2 like Table 7.

C1) Keep the present situation of U+24CB2, and add a second kRSUnicode value of 72.9 in the

future.

C2) Add two glyphs into SVS like this:

Table 8 Proposed SVSes for U+24CB2

Base Character	Variation Selector	Glyph	Description
U+24CB2	FE00 (VS1)	暮	UCS2003
U+24CB2	FE01 (VS2)	暮	GHZ-42549.08
U+24CB2	FE02 (VS3)	暮	T6-5F26

D1) Keep the present situation of U+24CB2, and add a second kRSUnicode value of 72.9 in the future.

D2) Add two glyphs into IVD like this:

Table 9 Proposed IVSes for U+24CB2

Base Character	Variation Selector	Glyph	Description
U+24CB2	E0100 (VS17)	暨	UCS2003
U+24CB2	E0101 (VS18)	暨	GHZ-42549.08
U+24CB2	E0102 (VS19)	暨	T6-5F26

4. Error M-source glyph for U+2CC43

Fig. 22 U+2CC43 on The Unicode Standard, Version 9.0

IRGN1606 showed MAC-00119 (aka MAC00119) was based on the book named "*Codificado de Romanizacao do Cantonese*" (aka *Silabário Codificado de Romanização do Cantonense* on the official website of Macao SAR). (pls see Fig. 23) Its Macao Cantonese pronunciation is NGAK.

Fig. 23 MAC-00119 on IRGN1606

MAC00119	頟	Based on the book name "Codificado de Romanizacao do Cantonese"	NGAK，Code:9473	名頁		Reserve	Based on the book name "Codificado de Romanizacao do Cantonese"

IRGN1739 Part D showed this character was NO. 9473 in the book named "*Codificado de Romanizacao do Cantonese*" (aka *Silabário Codificado de Romanização do Cantonense* on the official website of Macao SAR). (pls see Fig. 24)

13

Fig. 24 MAC-00119 on IRGN1739 Part D

We found two handwriting glyphs about NO. 9473 on the book named "*Codificado de Romanizacao do Cantonese*" (aka *Silabário Codificado de Romanização do Cantonense* on the official website of Macao SAR) when we review the book, one was on p. 3003 like IRGN1739 Part D showed, the other was on p. 2867. (pls see Figs. 25 & 26)

Fig. 25 MAC-00119 on Silabário Codificado de Romanização do Cantonense (p. 3003)

Fig. 26 MAC-00119 on Silabário Codificado de Romanização do Cantonense (p. 2867)

According to Fig. 26, we could find the left hand side of NO. 9473 was 各 not 名 evidently, and

NO. 9473 should be unifiable with 頷 (U + 981F).

On the other hand, we found the evidence for the character 􀀀名頁. (pls see Fig. 27)

Fig. 27 􀀀名頁 *on Zhōnghuá Zìhǎi*

冷玉龙, 韦一心 等 (Lěng Yùlóng, Wěi Yīxīn, et al): 《中华字海》 (*Zhōnghuá Zìhǎi*), 北京:

中华书局 (Běijīng: Zhonghua Book Company) & 北京: 中国友谊出版公司 (Běijīng:

Friendship Publishing Corporation of China), 1994, ISBN 7-5057-0630-6/H・1, p. 1583

We suggest two ways to solve this problem:

A1) Keep the present situation of U + 2CC43, and add a kSemanticVariant value of U + 981F.

A2) Keep the present situation of U + 981F, and add a kSemanticVariant value of U + 2CC43.

B1) Change MAC-00119 glyph to be 􀀀各頁, and make it unifiable with U + 981F.

B2) Make the source for U + 2CC43 become a new U-source glyph. The evidence of the new

U-source ideograph is Fig. 27.

Table 10 Information for a New Proposed U-source Ideograph

Glyph	IDS	Radical	Radical Form	Strokes	First Stroke	Simplified Form	Variant Form	Variant Code Point	U-source Number
䫴	⿰各頁 U+2FF0, U+5404, U+9801	181	頁	6	3	N/A	頜	U+981F	UTC-xxxxx OR UCI-xxxxx

5. Proposed additional information on CJK Unified Ideographs to ISO/IEC 10646 Annex P

At last, we suggest the four error information could be included to ISO/IEC 10646 Annex P before they will be solved.

Table 11 Proposed Additional Information on CJK Unified Ideographs to ISO/IEC 10646

Annex P

UCS / Glyph	Additional Information
2017B ⼧ 8.11 滴(UCS2003) 瀵(V2-6E46)	V2-6E46 source glyph for this code point was mistakenly designed.
204AF ⼉ 10.12 奐(UCS2003) 奐(V0-4422)	V0-4422 source glyph for this code point was mistakenly designed.
24CB2 田 102.9 壨(UCS2003) 壨(GHZ-42549.08) 壨(T6-5F26)	UCS2003 glyph was mistakenly unified to this code point.
2CC43 頁 181.6 頷(MAC-00119)	MAC-00119 source glyph was mistakenly designed from the evidence submitted by Macao SAR to IRG.

(End of Document)