

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation

Международная организация по стандартизации

Doc Type: Ideographic Rapporteur Group Document
Title: Comments on NUCV #401
Source: Eiso Chan(陈永聪)
Status: Individual Contribution to IRG #51, Hà Nội, Việt Nam
Action: For consideration by IRG
Date: 2018-10-13

Ming Fan suggested moving NUCV #401 to UCV. The following is my comments.

Ming Fan mentioned the 《新旧字形对照表》, which isn't an official document but would be included in all the regular dictionaries in P. R. China.

GB 2312-80, the earliest Chinese national standard related to the Han characters encoding, also included the 《新旧字形对照表》.

Fig. 1 List of the "Xīnzìxíng" and "Jiùzìxíng" in GB 2312-80

| | | |
|----|----|----|
| 角⑦ | 角⑦ | 解确 |
| 奂⑨ | 奂⑦ | 换痰 |
| 尒⑧ | 尒⑦ | 敝弊 |

All the pairs of the characters listed there should be the unifiable variants as we know. In the year of 2004, the Commercial Press (商务印书馆) published a book named *Research on the Glyph Form of Hanzi* (《汉字字形研究》), which is an academic collected works on studying the standardization of the modern Hanzi before releasing the TGH-2009 Beta and TGH-2013.

The scholars' comments are following the 《新旧字形对照表》 very clearly. I list 《新旧字形简论》 written by Prof. Lín Zhòngxiāng (林仲湘) and Mr. Lǐ Yìlín (李义琳) and 《新旧字形问题刍议》 written by Chéng Róng (程荣) as below. In fact, when we talk about "Xīnzìxíng" (新字形) and "Jiùzìxíng" (旧字形) in modern Chinese, we mean these pairs of variants are the unifiable variants, and we can also call them as "Yìxiězì" (异写字).

Fig. 2 Research on the Glyph Form of Hanzi, P. 95

2. 从俗、从简

文字本身就是约定俗成的,选择新字形更应如此,要选取群众认可的、使用面广的。如“神—神、直—直”。

从简是汉字规范化的基本原则,新字形也应是取笔画少的,在这点上跟简化字一致的。如“吕—吕 **奂—奂**”。

Fig. 3 Research on the Glyph Form of Hanzi, P. 100

2. 旧字形笔画较多,新字形减少笔画数的,例字也有 9 组:

朗(朗)、郎(郎)、吕(吕)、宫(宫)、印(印) **奂(奂)**、争(争)、盗(盗)、奥(奥)

NUCV shows there are six pairs of disunified ideographs related to 奂 vs 奂 in URO. I list them and their corresponding references as below. Note that * means the reference was used in Unicode, 1.0, ^ means the reference was used in ISO/IEC 10646.1-1993 (aka GB 13000.1-93).

Table 1 References of Several Characters Mentioned by NUCV #401

| UCS | Char. | Ref. | UCS | Char. | Ref. |
|--------|-------|------------------------------------|--------|-------|---|
| U+5524 | 奂 | G0-3B3D G:0-2729*^ A:27366D* | U+559A | 奂 | G1-3B3D G:1-2729*^ HB1-B3EA B:B3EA* J0-342D J:0-2013*^ KP0-F6C6 K0-7C30 K:0-9216*^ T1-5E50 C:1-5E50*^ V1-4F5D A:217152* X:247:125* |

| | | | | | |
|--------|---|------------------------------------|--------|---|---|
| U+5942 | 奂 | G0-5B3C G:0-5928*^ A:4B393B* | U+5950 | 奂 | G1-5B3C G:1-5928*^ HB1-ABB7 B:ABB7* J0-5476 J:0-5286*^ KP0-F6C7 K0-7C31 K:0-9217*^ T1-5059 C:1-5059*^ A:213938* X:265:110* |
| U+6362 | 換 | G0-3B3B G:0-2727*^ A:454146* | U+63DB | 換 | G1-3B3B G:1-2727*^ HB1-B4AB B:B4AB* J0-3439 J:0-2025*^ KP0-F6CB K0-7C35 K:0-9221*^ T1-5F50 C:1-5F50*^ V2-8C51 A:214146* X:245:063* |
| U+6DA3 | 換 | G0-3B41 G:0-2733*^ A:4B4835* | U+6E19 | 換 | G1-3B41 G:1-2733*^ HB1-B541 B:B541* J0-5E52 J:0-6250*^ KP0-F6CF K0-7C39 K:0-9225*^ T1-6047 C:1-6047*^ |

| | | | | | |
|--------|---|------------------------------------|--------|---|--|
| | | | | | A:214835* X:270:301* |
| U+7115 | 煥 | G0-3B40 G:0-2732*^ A:4B4973* | U+7165 | 煥 | G1-3B40 G:1-2732*^ HB1-B7D8 B:B7D8* J0-5F65 J:0-6369*^ KP0-F6D0 K0-7C3A K:0-9226*^ T1-647E C:1-647E*^ A:214973* X:260:131* |
| U+75EA | 瘥 | G0-3B3E G:0-2730*^ A:274C62* | U+7613 | 瘥 | G1-3B3E G:1-2730*^ HB1-BAC8 B:BAC8* J14-715E J:1-4577*^ KP1-5BEB K2-497C T1-696F C:1-696F*^ A:214C62* X:473:346* |

According to the above list, in ISO/IEC 10646.1-1993, these characters were encoded separately because of the G-Source. As we know, G1-Source, GB/T-12345-90, is the traditional version of G0-Source, GB 2312-80. However, 煥, 奂, 換, 渙, 煥 and 瘥 are not the traditional forms of 喚, 奂, 換, 渙, 煥 and 瘥, and 2729, 5928, 2727, 2733, 2732 and 2730 in GB/T 12345-90 are 喚, 奂, 換, 渙, 煥 and 瘥 without any doubt. We can confirm 喚, 奂, 換, 渙, 煥 and 瘥 are not included in the real G1-Source.

Fig. 7 Row 59 in Pseudo-GB/T 12345

| | | | | | | | | | | | | | | | | | | | | |
|-------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 59区 | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | |
| 01/19 | 邸 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | 郅 | |
| | 5B21 | 5B22 | 5B23 | 5B24 | 5B25 | 5B26 | 5B27 | 5B28 | 5B29 | 5B2A | 5B2B | 5B2C | 5B2D | 5B2E | 5B2F | 5B30 | 5B31 | 5B32 | 5B33 | |
| 39 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | 鄞 | |
| | 5B34 | 5B35 | 5B36 | 5B37 | 5B38 | 5B39 | 5B3A | 5B3B | 5B3C | 5B3D | 5B3E | 5B3F | 5B40 | 5B41 | 5B42 | 5B43 | 5B44 | 5B45 | 5B46 | 5B47 |
| 40/59 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | 互 | |
| | 5B48 | 5B49 | 5B4A | 5B4B | 5B4C | 5B4D | 5B4E | 5B4F | 5B50 | 5B51 | 5B52 | 5B53 | 5B54 | 5B55 | 5B56 | 5B57 | 5B58 | 5B59 | 5B5A | 5B5B |
| 50/79 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 | 圪 |
| | 5B5C | 5B5D | 5B5E | 5B5F | 5B60 | 5B61 | 5B62 | 5B63 | 5B64 | 5B65 | 5B66 | 5B67 | 5B68 | 5B69 | 5B6A | 5B6B | 5B6C | 5B6D | 5B6E | 5B6F |
| 80/94 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 | 垠 |
| | 5B70 | 5B71 | 5B72 | 5B73 | 5B74 | 5B75 | 5B76 | 5B77 | 5B78 | 5B79 | 5B7A | 5B7B | 5B7C | 5B7D | 5B7E | | | | | |

RFC 1922 was a document about ISO-2022-CN. It shows the following information. The reason why the escape sequence of GB/T 12345 was incomplete there was that it hadn't become the official part in the International Register, that means it was impossible for UTC to read the ISO-IR version of GB/T 12345 like GB 2312, which was ISO-IR-058.

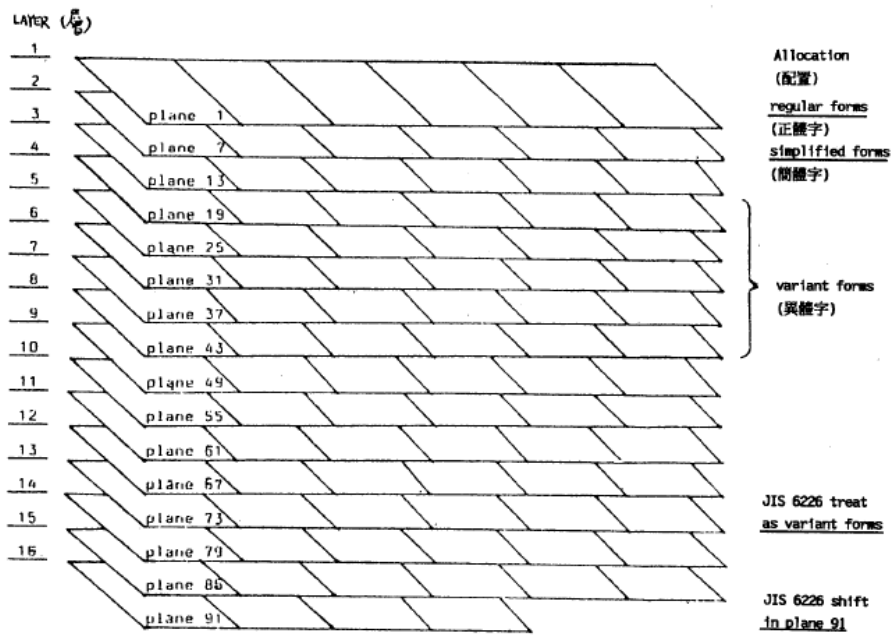
Table 2 Escape Sequences Information in RFC 1922

| | |
|-------------------|--|
| ESC \$) A | Indicates the bytes following SO are Chinese characters as defined in GB 2312-80, until another SO designation appears |
| ESC \$) <X12345> | Indicates the bytes following SO are as defined in GB 12345-90 [GB-12345], until another SO designation appears |

Some issues related to the Pseudo-G1-Source had been solved, but this issue has not been discussed.

On the other hand, the A-Source in Unicode 1.0 means the EACC which is the American version of CCCII. 喚, 奂, 換, 渙, 煥 and 瘼 are all included in Plane 21, Layer 1 of EACC, that was the most important section for the most common traditional forms; 喚 and 瘼 are included in Plane 27, Layer 2, that was prepared for the simplified variants; 換 is included in Plane 45, Layer 7; 奂, 渙, 煥 are included in Plane 4B, Layer 8. According to 《國字整理小組十年》, Layer 3 to Layer 8 of EACC and CCCII are preparing for the variants. It looked the CCCII editors and RLG didn't treat all of 喚, 奂, 換, 渙, 煥 and 瘼 are the simplified form of 喚, 奂, 換, 渙, 煥 and 瘼.

Fig. 8 《國字整理小組十年》, P. 9



圖一B CCCII編碼共分16面(layer)

All in all, I agree with Ming Fan that NUCV #401 should be moved to UCV.