ISO/IEC JTC 1/SC 2/WG 2/**IRG N2788**
Date: 2025-02-25

Source:	Judith Chen
Title:	Discrepancies between GB Standards and the UCS
Status:	Individual Contribution on IRG #64
Action:	To be considered by China NB, UK and IRG
Reference:	IRG N2276, IRG N2302, IRG N2542, IRG N2689

In IRG N2302, Mr. Eddie Li discussed several issues with the G3 and G5 values. Based on his contribution, this document further lists more discrepencies between the GB standards and the UCS, and suggests possible changes accordingly. The issues are divided into three categories: source reference value issues, glyph design issues and disunification.

1 Source Reference Value Issues

1.1 U+4FB4

The current G-source reference value of U+4FB4 is G3-327D, as shown by Figure 1.

Figure 1: U+4FB4 in Unicode 16.0

However, in GB 7589—87 (simplified forms of G3), the glyph listed at 18-93 (0x327D) is 剑, while 俞 is separately encoded in another standard GB 7590—87 (simplified forms of G5) at 17-31 (0x313F), as shown by Figure 2 and Figure 3.

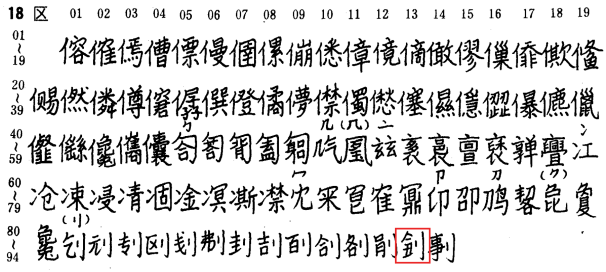


Figure 2: 18-93 (0x327D) in GB 7589—87

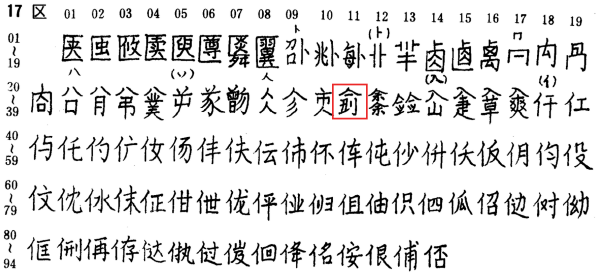


Figure 3: 17-31 (0x313F) in GB 7590—87

Considering that 剑 has been submitted to WS2024, I would like to suggest the following changes to the Unihan database:

- U+4FB4
  - kIRG_GSource: G3-327D → G5-313F
  - kGB3: 1893 → (null)
  - kGB5: (null) → 1731

1.2 U+96DF and U+5DC2

The current kIRG_GSource property of U+96DF is G3-3970, while the current kIRG_GSource property of U+5DC2 is G5-3F37, as shown by Figure 4.

Figure 4: U+96DF and U+5DC2 in Unicode 16.0

In contrast, in GB 7589—87, the radical of 25-80 (0x3970) is 山, whereas in GB 7590—87, the radical of 31-23 (0x3F37) is 屮, as shown by Figure 5 and Figure 6.

Figure 5: 25-80 (0x3970) in GB 7589—87

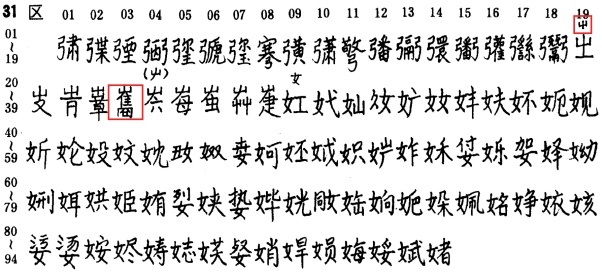


Figure 6: 31-23 (0x3F37) in GB 7590—87

Therefore, I would like to suggest the following changes to the Unihan database:

- U+5DC2
  - kIRG_GSource: G5-3F37 → G3-3970
  - kGB5: 3123 → (null)
  - kGB3: (null) → 2580
- U+96DF
  - kIRG_GSource: G3-3970 → G5-3F37
  - kGB3: 2580 → (null)
  - kGB5: (null) → 3123

1.3 U+58AB and U+58FF

The current G-source reference value of U+58AB is G5-3722, while the current kIRG_GSource property of U+58FF is GE-3541, as shown by Figure 7.

Figure 7: U+58AB and U+58FF in Unicode 16.0

In contrast, in GB 7590—87, the radical of 23-02 (0x3722) is 士, as shown by Figure 8.

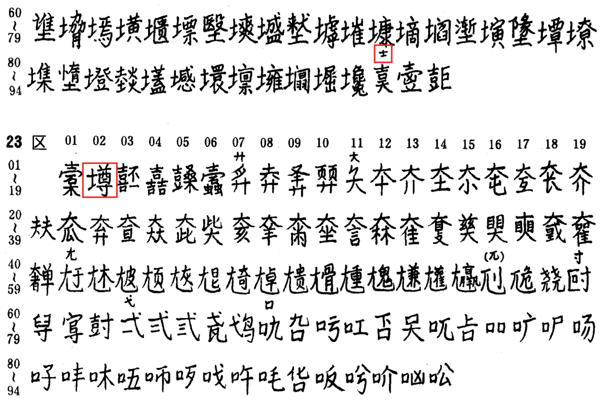


Figure 8: 23-02 (0x3722) in GB 7590—87

Therefore, I would like to suggest the following changes to the Unihan database:

- U+58AB
  - kIRG_GSource: G5-3722 → GKX-0239.03
  - kGB5: 2302 → (null)
- U+58FF
  - kIRG_GSource: GE-3541 → G5-3722
  - kGB5: (null) → 2302

1.4 U+4A9E

The current G-source reference value of U+4A9E is G5-7768, as shown by Figure 9.

Figure 9: U+4A9E in Unicode 16.0

However, in GB 7590—87, the glyph listed at 87-72 (0x7768) is ⿱夭韭, while 䪞 is separately encoded at 87-71 (0x7767), as shown by Figure 10.

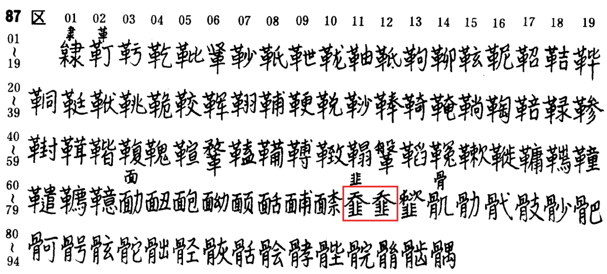


Figure 10: 87-71 (0x7767) and 87-72 (0x7768) in GB 7590—87

Therefore, I suggest that the G-source reference value of U+4A9E be revised to G5-7767.

1.5 U+4E85, U+5570 and U+7CA6

In IRG N2689, China proposed updating the G-source reference values for 1,685 characters. Specifically, most of the changes to the GU-source reference values are based on IRG N2542, a proposal submitted in 2022.

However, in PRI #497 feedback [ID20240218072351], I pointed out that 丿, 啰, 龏 and 覀 do exist in GB 8565.2—88 (as shown by Figure 11), so U+4E85, U+5570, U+7CA6 and U+8980 should be revised back to the G8-source. According to L2-24/067, China agreed with these changes, and the four kIRG_GSource properties were updated in Unicode 16.0.

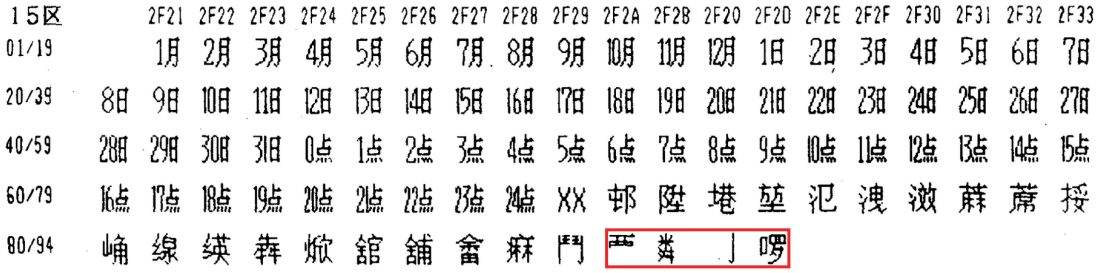


Figure 11: 0x2F7A..0x2F7D in GB 8565.2—88

Since U+4E85, U+5570 and U+7CA6 currently do not have GU-source reference values, they are not supposed to be revised to other sources. Therefore, I suggest that the kIRG_GSource properties of U+4E85, U+5570 and U+7CA6 remain as G8-2F7C, G8-2F7D and G8-2F7B, respectively.

1.6 G7-source Characters

According to UAX #38, the G7-source characters are sourced from 'General Purpose Hanzi List for Modern Chinese Language, and General List of Simplified Hanzi'.

G7-21xx characters were introduced in the very first edition of the UCS. Most of them can be found in the General Purpose Hanzi List for Modern Chinese Language (现代汉语通用字表) published in 1988. However, G7-214B 镕 was not reinstated for use until 1993, so it can not be included in the General Purpose Hanzi List for Modern Chinese Language. Therefore, I suggest that the G-source reference value of U+9555 be revised to GHZR-84572.01.

Other G7-source characters were introduced with the release of CJK Unified Ideographs Extension A. All G7-22xx characters can be found in the General List of Simplified Hanzi (简化字总表), while G7-23xx characters do not appear in either of the two lists. If these characters are actually sourced from a different list, the description of the G7-source should be updated accordingly, otherwise their G-source reference values should be revised to other sources. Listed below are possible changes to the kIRG_GSource values of G7-23xx characters (also available as a PDF attachment):

UCS	Glyph	Current Source	Recommened Source
U+3437	伲	G7-2326	GHZR-10148.01
U+3438	伙	G7-2321	GKX-0094.10
U+343B	纳	G7-2327	GKX-0096.09
U+343D	伬	G7-2325	GIDC-E006
U+343E	伝	G7-2328	GZH-0066.28
U+3448	佟	G7-2323	GZH-0069.16
U+344E	侷	G7-232A	GKX-0102.19
U+3454	侦	G7-2322	GZH-0072.01

UCS	Glyph	Current Source	Recommened Source
U+3455	㑥	G7-2324	GU-03455
U+3460	㑬	G7-232C	GU-03460
U+346B	㑫	G7-232B	GZH-0079.43
U+3482	㑽	G7-232D	GKX-0118.23
U+362E	㘮	G7-232E	GIDC-E01D
U+364B	㙋	G7-232F	GIDC-E022
U+364C	㙌	G7-2331	GZH-0232.01
U+36AF	㚯	G7-2335	GU-036AF
U+36BF	㚿	G7-2332	GHZ-21040.06
U+36C0	㛀	G7-2334	GHZR-21114.04
U+36E3	㛣	G7-2336	GZH-0689.13
U+36E4	㛤	G7-2337	GHZR-21124.11
U+371F	㜟	G7-2338	GU-0371F
U+3877	㡷	G7-2339	GKX-0344.12
U+38AB	㢫	G7-233D	GU-038AB
U+38B3	㢳	G7-233B	GKX-0358.15
U+38B7	㢷	G7-233A	GIDC-E059
U+38BA	㢺	G7-233C	GKX-0359.08
U+38DD	㣝	G7-233E	GKX-0366.08
U+392D	㤭	G7-233F	GHZR-42458.10
U+393D	㤽	G7-2343	GHZR-42464.14
U+3957	㥗	G7-2344	GU-03957
U+396A	㥪	G7-2340	GHZR-42495.19
U+3988	㦈	G7-2341	GZH-0605.45
U+39D1	㧑	G7-2345	GHZR-41950.02
U+39F0	㧰	G7-2346	GHZR-41962.04
U+3A2C	㨬	G7-2347	GZH-0353.31
U+3A4F	㩏	G7-2348	GZH-0362.41
U+3B4F	㭏	G7-234A	GZH-0740.14
U+3B64	㭤	G7-2349	GZH-0742.19
U+3B65	㭥	G7-2352	GU-03B65
U+3B6A	㭪	G7-234F	GKX-0527.11
U+3B75	㭵	G7-234C	GDM-?????[1]

[1]㭵 is included in SJ/T 11239—2001. According to IRG N2689, GDM-source is used for SJ/T 11239—2001 characters. However, the specific decimal value still requires China's confirmation.

UCS	Glyph	Current Source	Recommened Source
U+3B76	楞	G7-2351	GZH-0752.40
U+3B77	楛	G7-2353	GU-03B77
U+3BA0	楠	G7-2350	GIDC-E06F
U+3BC1	楹	G7-234E	GKX-0548.07
U+3BE8	槔	G7-234D	GU-03BE8
U+3CD4	浒	G7-235D	GHZR-31713.05
U+3CD5	浐	G7-235A	GZH-0536.23
U+3CE1	浞	G7-2354	GHZR-31717.03
U+3CE2	浟	G7-2355	GZH-0539.22
U+3CF0	浡	G7-2357	GHZ-31631.11
U+3CF1	浥	G7-235C	GIDC-E080
U+3CFE	浼	G7-2358	GKX-0631.07
U+3D0B	涑	G7-2356	GHZR-31783.05
U+3D28	涨	G7-2359	GKX-0640.13
U+3D47	涓	G7-235B	GU-03D47
U+3EC5	珍	G7-235E	GHZR-21189.10
U+3EFC	璋	G7-235F	GU-03EFC
U+4099	邙	G7-2361	GHZ-63762.03
U+40B1	硪	G7-2362	GKX-0830.22
U+40B5	硕	G7-2363	GZH-1024.46
U+40C4	碌	G7-2364	GU-040C4
U+40C5	碑	G7-2365	GHZR-52612.01
U+410A	祗	G7-2367	GU-0410A
U+4114	禡	G7-2366	GKX-0844.29
U+41E5	笠	G7-2368	GKX-0881.21
U+41F4	筢	G7-236A	GZ-4442202
U+4244	篾	G7-236D	GKX-0897.35
U+4338	纬	G7-236E	GHZR-63591.02
U+4339	纠	G7-2370	GHZR-63605.08
U+433A	纰	G7-236F	GHZR-63614.05
U+433B	纲	G7-2371	GZH-1319.24
U+4880	乾	G7-2372	GIDC-E0E1
U+4881	钥	G7-2373	GIDC-E0E2
U+48B6	邦	G7-234B	GHZ-63757.06

2 Glyph Design Issues

2.1 U+6F78

The current G-source representative glyph for U+6F78 is 潸, while the current G-source representative glyph for U+23F7D is 潸, as shown by Figure 12.

Figure 12: U+6F78 and U+23F7D in Unicode 16.0

However, in GB 2312—80, the glyph listed at 68-90 (0x647A) is 潸 (as shown by Figure 13)

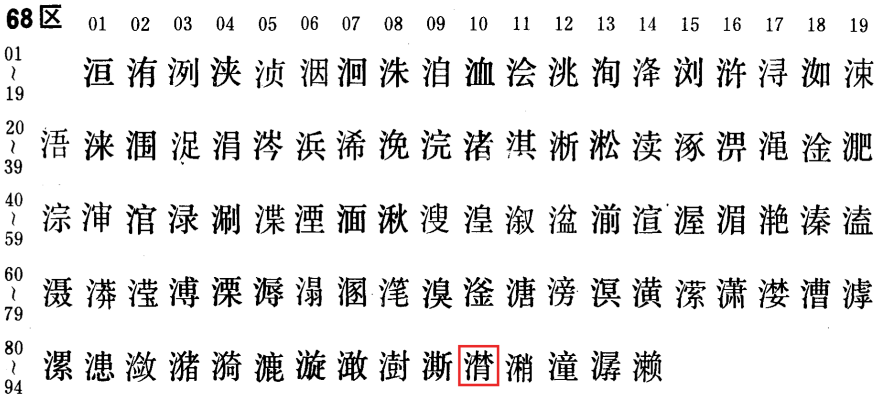


Figure 13: 68-90 (0x647A) in GB 2312—80

China's GB 18030—2022 standard also adopts the shape 潸 for U+6F78, as shown by Figure 14.

Figure 14: U+6F78 in GB 18030—2022

Since 潸 is listed in China's 通用规范汉字表 and its mapping to U+6F78 is specified in Appendix E of GB 18030—2022 (as shown by Figure 15), I suggest that the G-source representative glyphs for U+6F78 and U+23F7D be swapped[2], and the G-source reference value for U+23F7D be revised to GKX-0651.04.

GB 18030—2022

表 E.1 《通用规范汉字表》汉字的字形和代码位置（续）

序号	本文件的码位	本文件对应的 GB/T 13000 码位	字形
5998	E4F9	6F8C	澌
5999	E4FA	6F78	潸
6000	C1CA	6F66	潦

Figure 15: TGH 5999 潸 is mapped to U+6F78 in Appendix E of GB 18030—2022

[2] L2/23-106 has pointed out that such situation is not unique as demostrated by the T-source representative glyphs of U+81F4 and U+26936.

2.2 U+3594

The current G-source representative glyph for U+3594 is 唔, as shown by Figure 16.

Figure 16: U+3594 in Unicode 16.0

However, in GB 7589—87, the glyph listed at 22-16 (0x3630) is 唔, as shown by Figure 17.

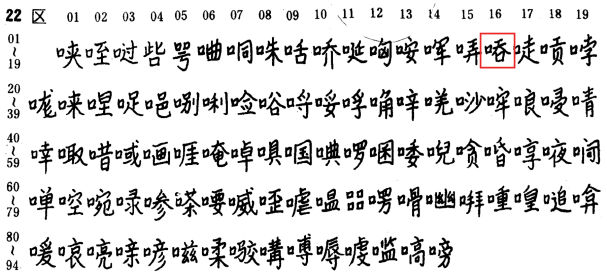


Figure 17: 22-16 (0x3630) in GB 7589—87

China's GB 18030—2022 standard also adopts the shape 唔 for U+3594, as shown by Figure 18.

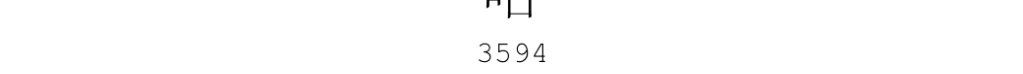


Figure 18: U+3594 in GB 18030—2022

Therefore, I suggest that the G-source representative glyph for U+3594 be normalized to 唔.

2.3 U+3ADA and U+66F6

The current G-source representative glyph for U+3ADA is 舀, while the current G-source representative glyph for U+66F6 is 舀, as shown by Figure 19.

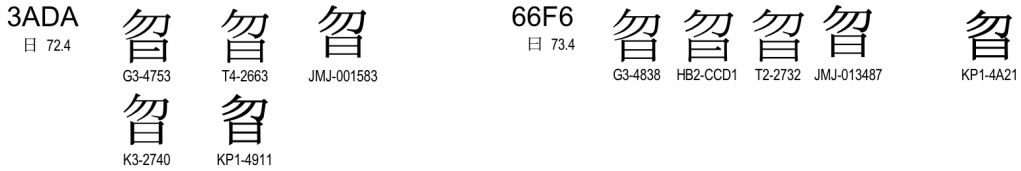


Figure 19: U+3ADA and U+66F6 in Unicode 16.0

However, in GB 7589—87, the glyph listed at 39-51 (0x4753) is ⿰勿日, while the glyph listed at 40-24 (0x4838) is 舀 (actually ⿰勿曰), as shown by Figure 20.

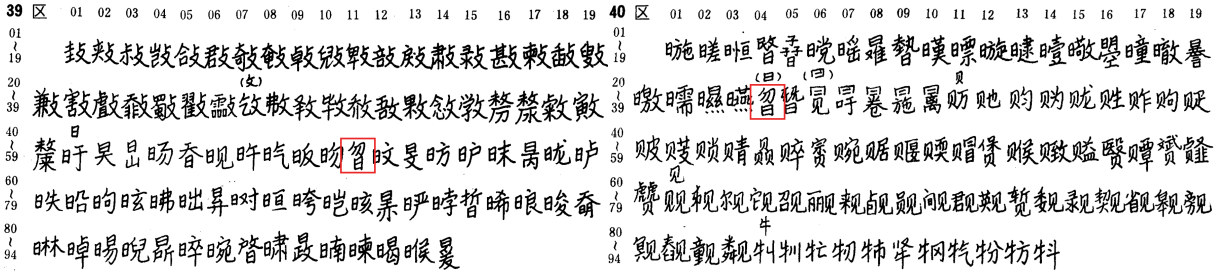


Figure 20: 39-51 (0x4753) and 40-24 (0x4838) in GB 7589—87

Just as Ms. CheonHyeong Sim pointed out on Zhihu, U+3ADA is highly confusable with U+66F6. Based on their rationale, rather than revising the G-source representative glyph for U+3ADA to 𫶇, it might be better to swap the G-source glyphs for U+3ADA and U+66F6.

2.4 U+9D56 (and U+3125D)

The current G-source representative glyph for U+9D56 is 鵖, as shown by Figure 21.

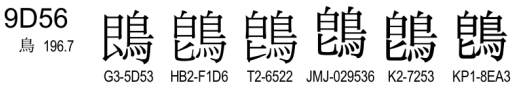


Figure 21: U+9D56 in Unicode 16.0

However, in GB 7589—87, the glyph listed at 61-51 (0x5D53) is 𫶇, as shown by Figure 22.

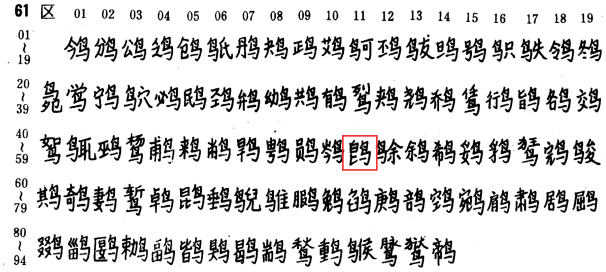


Figure 22: 61-51 (0x5D53) in GB 7589—87

In addition, the G-source glyph for U+9D56 is 鵖 in ISO/IEC 10646-1:1993, as shown by Figure 23.

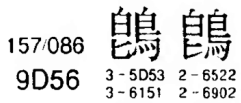


Figure 23: U+9D56 in ISO/IEC 10646-1:1993

Since the simplified character in the orginal evidence has not undergone normalization, there is no need to normalize its traditional form. Therefore, I suggest that the G-source representative glyph for U+9D56 be revised back to 鵖 (or normalized to 鵖, which aligns more closely with the original evidence).

Besides, the current UK-source representative glyph for U+3125D is 𫶇, as shown by Figure 24.

Figure 24: U+3125D in Unicode 16.0

On the contrary, the shape in its original evidence is 𫶇, as shown by Figure 25.

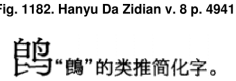


Figure 25: 𫶇 in the original evidence of UK-02441

When reviewing UK-02441, Mr. Henry Chan claimed that the shape 𫶇 'does not match PRC conventions' and 'the left hand component should be left of 即'. However, as mentioned before, the shape can be found in GB 7589—87, a Chinese national standard (Figure 22). Furthermore, it is not hard to find other G-source glyphs that contain 皂, such as U+238BC,

U+23A98, U+2BA43, etc, as shown by Figure 26. Therefore, the UK-source representative glyph for U+3125D 鵾 does not need normalizing and should be revised back to ⿰皂鸟 in order to match its original evidence, and the kkRSUnicode and kTotalStrokes properties should be updated to 196′.7 and 12 accordingly.

Figure 26: U+238BC, U+23A98 and U+2BA43 in Unicode 16.0

2.5 U+4200

The current G-source representative glyph for U+4200 is 箭, as shown by Figure 27.

Figure 27: U+4200 in Unicode 16.0

However, in GB 7589—87, the glyph listed at 68-09 (0x6429) is 箭, as shown by Figure 28.

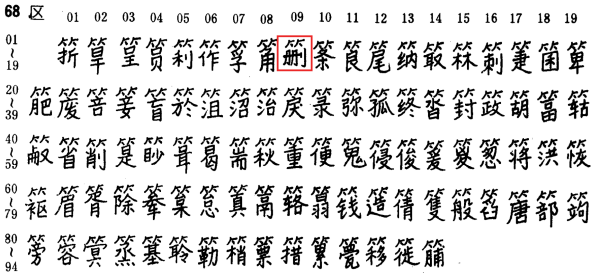


Figure 28: 68-09 (0x6429) in GB 7589—87

The G-source glyph for U+4200 is also 箭 in ISO/IEC 10646-1:2000, as shown by Figure 29.

Figure 29: U+4200 in ISO/IEC 10646-1:2000

Therefore, I suggest that the G-source representative glyph for U+4200 be normalized to 箭. If China wants to keep the current shape, it might be advisable to revise the G-source reference value of U+4200 to GHZ-52973.13 (as shown by Figure 30).

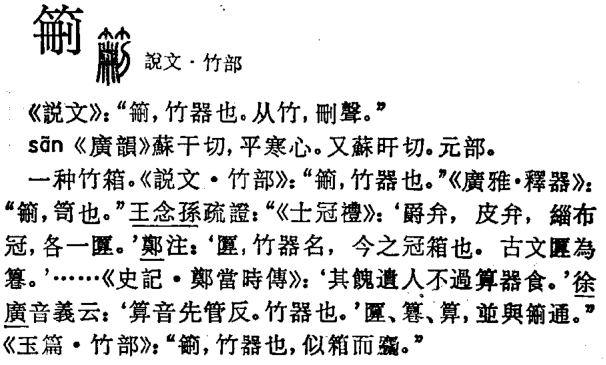


Figure 30: 箭 in 汉语大字典

2.6 U+4BA7, U+9DBE and U+96D7

Mr. Andrew West observed that ‘there is no consistency in the use of ⿰卓𠂉/⿰卓𠂇 in China characters’ when discussing the glyph design of GKJ-00772. He listed a few characters, including U+4BA7, U+9DBE and U+96D7. Their current G-source representative glyph are 鶼, 鶻 and 雔, respectively, as shown by Figure 31.

Figure 31: U+4BA7, U+9DBE and U+96D7 in Unicode 16.0

However, as shown by Figure 32 and Figure 33, the corresponding glyphs listed in GB 7589—87 and GB 7590—87 are 鹌, 鹐 and 雔.

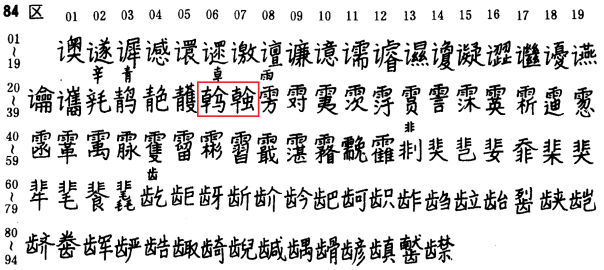


Figure 32: Section 84 in GB 7589—87

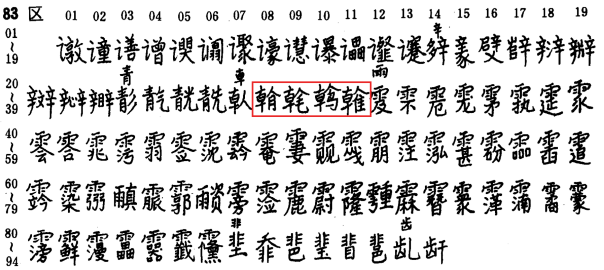


Figure 33: Section 83 in GB 7590—87

It appears that the original Chinese standards favour the shapes containing 臤 over ⿰卓𠂉 (even U+2338D 朝 is normalized to ⿰臤月), but currently only U+8792 and U+3CA6 have G-source representative glyphs that align with the original standards, as shown by Figure 34.

Figure 34: U+8792 and U+3CA6 in Unicode 16.0

In addition, the G-source glyphs for U+9DBE and U+96D7 are 鶼 and 雔 in ISO/IEC 10646-1:1993, as shown by Figure 35.

Figure 35: U+9DBE and U+96D7 in ISO/IEC 10646-1:1993

China’s GB 18030—2022 standard also adopts the shape 鶼 for U+4BA7, as shown by Figure 36.

Figure 36: U+4BA7 in GB 18030—2022

Therefore, I suggest that the G-source representative glyphs for U+4BA7, U+9DBE and U+96D7 be normalized to 鶼, 鶫 and 雔, respectively.

2.7 U+5329

The current G-source representative glyph for U+5329 is 匩, as shown by Figure 37.

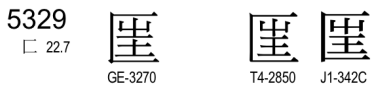


Figure 37: U+5329 in Unicode 16.0

However, in GB/T 16500—1998, the glyph listed at 18-80 (0x3270) is 匩, as shown by Figure 38.

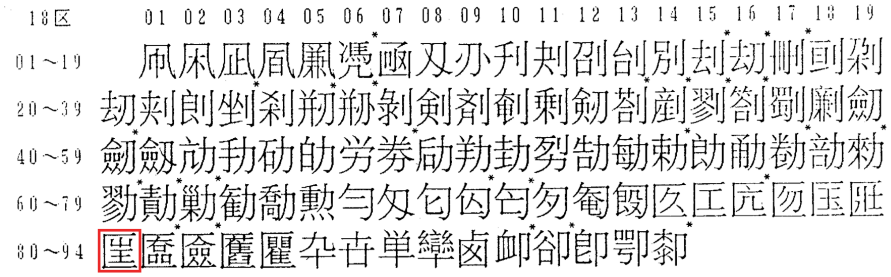


Figure 38: 18-80 (0x3270) in GB/T 16500—1998

The G-source glyph for U+5329 is also 匩 in ISO/IEC 10646-1:2000, as shown by Figure 39.

Figure 39: U+5329 in ISO/IEC 10646-1:2000

Therefore, I suggest that the G-source representative glyph for U+5329 be revised to 匩.

2.8 U+6FF2

The current G-source representative glyph for U+6FF2 is 灂, as shown by Figure 40.

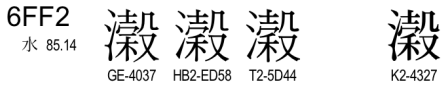


Figure 40: U+6FF2 in Unicode 16.0

However, in GB/T 16500—1998, the glyph listed at 32-23 (0x4037) is 灂, as shown by Figure 41.

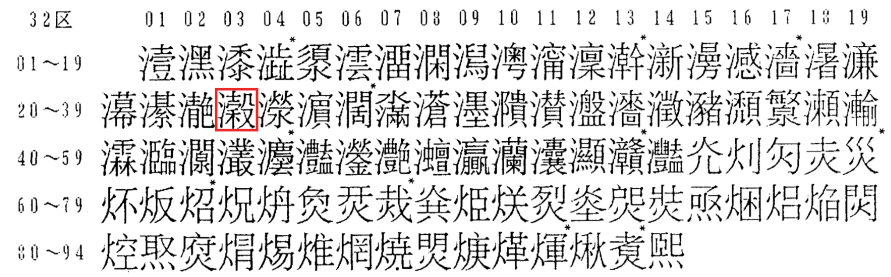


Figure 41: 32-23 (0x4037) in GB/T 16500—1998

China's GB 18030—2022 standard also adopts the shape 瀙 for U+6FF2, as shown by Figure 42.

Figure 42: U+6FF2 in GB 18030—2022

Therefore, I suggest that the G-source representative glyph for U+6FF2 be revised to 瀙.

3 Disunification

Currently, the G-source representative glyph for U+4748 is 豽, while the glyphs of all other sources are 豽, as shown by Figure 43.

Figure 43: U+4748 in Unicode 16.0

Unlike previous cases, the G-source glyph 豽 is consistent with the original standard, as shown by Figure 44.

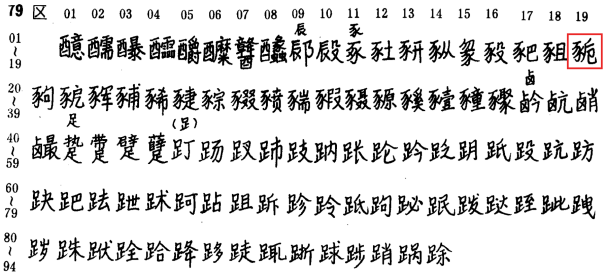
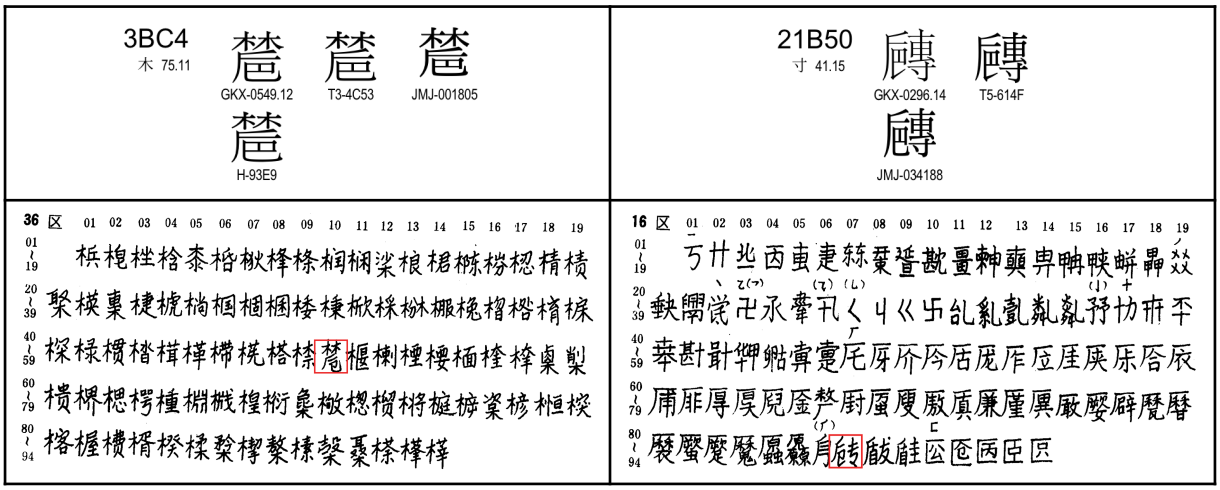


Figure 44: 79-19 (0x6F33) in GB 7589—87

There are several examples where shapes containing 厄 are encoded in UCS, while GB standards favour shapes containing 戹, for China's 第一批异体字整理表 and 印刷通用汉字字形表 prefer 戹 over 厄. For instance, U+3BC4 㦄 corresponds to 36-50 ⿰林戹 in GB 7590—87, and U+21B50 𡜐 corresponds to 16-87 ⿰戹专 in GB 7590—87, as shown below. Similarly, 79-19 豽 in GB 7589—87 should correspond to U+27C52 豽.



According to 康熙字典, ⿰豸巵 is pronounced the same as its phonetic symbol 巵, while 䝈, a variant of ⿰豸⿱宀㔾, is pronounced the same as its phonetic symbol 厄. Clearly, the two shapes, ⿰豸巵 and 䝈, are actually non-cognate.

Therefore, I suggest revising the G-source glyph of U+4748 to 䝈 and changing the kIRG_GSource to GKX-1195.03. ⿰豸巵 could be encoded separately if needed.

Reference

- International Standards
  - ISO/IEC 10646-1:1993 *Information technology — Universal Multiple-Octet Coded Character Set (UCS) Part 1: Architecture and Basic Multilingual Plane*
  - ISO/IEC 10646-1:2000 *Information technology — Universal Multiple-Octet Coded Character Set (UCS) Part 1: Architecture and Basic Multilingual Plane*
- PRC Standards
  - GB 2312—80 信息交换用汉字编码字符集 基本集
  - GB 7589—87 信息交换用汉字编码字符集 第二辅助集
  - GB 7590—87 信息交换用汉字编码字符集 第四辅助集
  - GB 8565.2—88 信息处理 文本通信用编码字符集 第二部分：图形字符集
  - GB/T 16500—1998 信息交换用汉字编码字符集 第七辅助集
  - GB 18030—2022 信息技术 中文编码字符集
  - SJ/T 11239—2001 信息技术 信息交换用汉字编码字符集 第八辅助集
- PRC Specifications
  - 简化字总表
  - 印刷通用汉字字形表
  - 现代汉语通用字表
  - 第一批异体字整理表
  - 通用规范汉字表
- Dictionaries
  - 康熙字典
  - 汉语大字典

(End of Document)