ISO/IEC JTC1/SC2/WG2 IRG N2840

DATE: 2025-4-5

Title: Request to revise the glyphs for 4 G-source ideographs

Source: Ma Shijie

Status: Individual Contribution

Action: For China NB, IRG and WG2

More than 50 glyphs have been duplicate-encoded like 炯 (U+3DB7) and 炯 (U+2420E) (though this may not be exhaustive—see the appendix).

Although the duplicate encoding of CJK ideographs in Unicode is intended for compatibility and historical reasons, it introduces several drawbacks, primarily affecting text processing, consistency, and user experience.

1.Increased Complexity in Text Processing

Different encodings of the same ideograph are treated as distinct, causing problems in search, sorting, and comparison operations. For example:

2.Data Redundancy and Storage Waste

Duplicate encodings occupy extra code points, increasing storage and transmission overhead (especially in large texts).

Database indexes may bloat due to redundant ideograph representations.

3.Challenges in Natural Language Processing (NLP)

Tokenization and Encoding: Duplicate-encoded ideographs may be misinterpreted by tokenizers, affecting machine learning models.

Text Analysis: For example, frequency counts may treat 炯 (U+3DB7) and 𠀀炯(U+2420E) as separate, skewing results.

4.User Experience Problems

Copy-Paste Anomalies: Text copied from different sources may contain duplicate-encoded ideographs, causing formatting issues.

Collaboration Conflicts: In collaborative editing, different encodings of the same ideograph may be flagged as changes (e.g., Git diff detection).

Here, I have compiled data on four of these ideographs and

suggest that China revise the glyphs in the right column of the table based on new reference sources.

20457 人 9.19 僊 GKX-0121.39 僊 T4-662F 僊 僊 說文·人部 同“仙”。《廣韻·仙韻》:“僊”,“仙”的古文。 GHZR-10280.06	34A8 人 9.19 僊 G5-3329 僊 T7-574C 僊同“仙”。《集韻·僊韻》:“僊,隸作僊,通作仙。”《漢書·郊祀志》:“(求)僊人羨門之屬。”顏師古注:“古亦以僊為仙字。” GHZR-10281.02
821D 舛 136.7 牽 牽 牽 牽 牽 牽 G5-3039 HB2-DFA3 T2-4648 J1-5674 K2-564E KP1-6C7E 牽 [xiá《广韵》胡瞎切, 入鎋, 匣。] ❶车辖。古代为固定车轮而插在车轴两端的键。《诗·邶风·泉水》:“載脂載牽,還車言邁。”孔颖达疏:“古者車不駕則脱其牽……今將行,既脂其車,又設其牽。”又《小雅·车牽序》:“車牽,大夫刺幽王也。”陆德明释文:“車軸頭鐵也。”❷星名。《史记·天官书》:“房爲府,曰天駟。其陰,右驂。旁有兩星曰衿;北一星曰牽。” GHC-010797.02	246C9 牛 93.10 牽 GHC 牽 T5-4730 246C9 牛 93.10 牽 GHC 牽 T5-4730 Unicode 5.2 牽 xiá 音狹] 同“辖”,穿在车轴两端孔内使车轮不脱落的键。字见《篇海》。 GHZ-0854.55
23515 木 75.8 桀 GHZ-21235.06 桀 T6-4A39 桀 說文·桀部 同“乘”。《龍龕手鑑·入部》:“桀,古文。車桀也。”	204F2 入 11.10 桀 GHZ-10199.02 桀同“乘”。《篇海類編·人事類·人部》:“桀,通作乘。” GHZR-10237.05

GHZR-10237.06	
249BC 玉 96.9 堅 T5-473E	249E9 玉 96.9 堅 GHZ-21130.04 堅 説文·玉部 《説文》:“堅,石之似玉者。从玉,取(旧字形作‘取’)聲。” wàn 《集韻》烏貫切,去換影。元部。 似玉的石。《説文·玉部》:“堅,石之似玉者。” GHZR-21200.09

END.

Appendix

序号	A	B
1	34A8 人 9.19 僊 G5-3329 僊 T7-574C	20457 人 9.19 僊 GKX-0121.39 僊 T4-662F
2	3588 口 30.6 耆 TF-295C 耆 K3-224C	439B 老 125.3 耆 K3-2F51
3	363D 土 32.6 㦸 GHZ-10437.05 㦸 K3-2321 㦸 KP1-3B9C	39B3 戈 62.5 㦸 K3-2623
4	3DB7 火 86.6 焗 K3-2A24	2420E 火 86.6 焗 TF-3070
5	5341 十 24.0 十 G0-4A2E 十 HB1-A451 十 T1-4432 十 J0-3D3D 十 K0-6428 十 KP0-E5EC 十 V1-4D63	2C09B 木 75.-2 十 JK-58711
6	5E50 肉 130.9 腃 G3-4D56 腃 T4-3D2A 腃 J1-3C33 腃 KP1-40A7	3B3A 月 74.9 腃 K3-275E
7	69E9 木 75.10 槩 GE-3D65 槩 H-9B4C 槩 T3-4C64 槩 J3-7623 槩 K1-576B 槩 KP1-4CFB 槩 V2-8D3A	3BA3 木 75.9 槩 GHZ-21257.01 槩 T3-4167
8	7468 玉 96.10 瑨 G3-3F6B 瑨 H-947A 瑨 T3-4822 瑨 J1-4C45 瑨 K0-7253 瑨 KP0-EBA5	24A01 玉 96.10 瑨 TF-4978

9	43DE 肉 130.5 肺 GKX-0977.09 肺 T5-2B33 肺 JA-257C	2669C 肉 130.5 肺 TF-2D2D
10	7B93 竹 118.8 篆 G8-7E28 篆 T3-485E 篆 K2-5036 篆 KP1-6457 篆 VN-07B93	25B7C 竹 118.7 篆 TF-4448 篆 VN-25B7C
11	7BF9 竹 118.11 簋 GE-4637 簋 HB2-EABF 簋 T2-592B 簋 J1-5258 簋 K2-505B 簋 KP1-64F0	25CBB 竹 118.11 簋 GHZ-53008.05 簋 T7-3A2A
12	809E 肉 130.3 肟 GLK-4212074 肟 T4-2470 肟 J1-556E	2E30B 肉 130.4 肟 JMJ-058379 肟 SAT-60289
13	821D 舛 136.7 犟 G5-3039 犟 HB2-DFA3 犟 T2-4648 犟 J1-5674 犟 K2-564E 犟 KP1-6C7E	246C9 牛 93.10 犟 GHC 犟 T5-4730
14	8654 虍 141.4 虔 G0-722F 虔 HB1-B040 虔 T1-5825 虔 J0-694A 虔 K0-4B73 虔 KP0-CEDA 虔 V1-656D	2AA90 广 53.6 虔 TC-3B23
15	902F 辵 162.8 逯 G0-6556 逯 HB2-DC4F 逯 T2-4074 逯 J1-6153 逯 K2-6572 逯 KP1-7DA0	284FB 辵 162.8 逯 GKX-1260.15 逯 T4-3B5B
16	9094 邑 163.3 邔 GE-4F53 邔 HB2-CA55 邔 T2-2275 邔 KP1-7E2C	2866C 邑 163.3 邔 GKX-1268.05

17	9646 阜 170.5 陆 G0-423D	28E38 阜 170.5 陆 TF-287E
18	2001F 一 1.5 束 GHZ-80007.04	233BA 木 75.2 束 T5-2274
19	200A6 丿 4.5 用 GHZ-10038.02	24C03 用 101.1 用 T6-253A
20	200C1 丿 4.10 衷 GHZ-10041.08	24C6E 田 102.6 衷 T6-422E
21	2028D 人 9.8 金 GHZ-10169.14	28940 金 167.2 金 T6-3C62
22	20306 入 11.10 㚇 GHZ-10169.02	204F3 入 11.10 㚇 T6-465A
23	204F2 入 11.10 桀 GHZ-10199.02	23515 木 75.8 桀 桀 GHZ-21235.06 T6-4A39
24	20500 八 12.1 廾 八 GKX-0126.27 T6-2137	2232C 廾 55.0 廾 T6-213F

25	20553 冂 13.5 肯 T5-233A	21D0A 屮 45.4 肯 GHZ-21022.03
26	20727 刀 18.7 劒 GHZ-10338.11	20730 刀 18.7 劒 T6-3549
27	20947 匚 22.7 匿 GHZ-10086.08	20975 匚 23.7 匿 T5-2834
28	21018 口 30.13 嚶 GHZ-10694.07	2103C 口 30.13 嚶 T5-5376
29	21D0C 屮 45.5 命 命 GKX-0305.12 T6-2A71	24C12 田 102.0 命 GKX-0758.03
30	21F37 山 46.12 嵒 GHZ-80016.20	2439A 火 86.11 嵒 T6-5E4E
31	2211A 巾 50.10 冪 T5-3E7B	26971 臼 134.6 冪 GHZ-53039.17
32	22472 彐 58.14 㚇 T7-3775	26133 糸 120.11 㚇 GHZ-53449.05

33	22484 彡 59.3	匚 T6-246F		22485 彡 59.3	匚 GHZ-20852.08	
34	22C93 手 64.10	擧 GHZ-31927.07		26979 臼 134.7	擧 T6-6075	
35	23040 攴 66.25	斷 GHZ-21481.11		27959 角 148.22	斷 T5-7B69	
36	23386 月 74.9	贏 G4K-15394		26760 肉 130.9	贏 GKX-0988.14	贏 T5-4227
37	23761 木 75.14	櫹 G4K-21259		237BC 木 75.15	櫹 TF-624F	櫹 H-93DF
38	23F03 水 85.11	溱 TF-4926		2DC88 水 85.11	溱 SAT-03973	
39	23F41 水 85.11	溧 H-8DEC		23F9E 水 85.12	溧 TF-4F77	
40	24425 火 86.14	燹 GHZ-32243.15		2444B 火 86.14	燹 T7-392A	

41	249BC 玉 96.9 堅 T5-473E	249E9 玉 96.9 堅 GHZ-21130.04
42	24BD2 甘 99.23 鬻 GHZ-42415.04	2A415 黍 202.16 鬻 T7-652A
43	24C0D 用 101.13 蒲 TF-5F61	27061 艸 140.15 蒲 TF-635A
44	251D7 目 109.8 眼 GKX-0810.27	251E4 目 109.8 眼 T6-555D
45	25844 禾 115.10 稄 GHC	2584C 禾 115.10 稄 TF-5072
46	25C94 竹 118.11 篁 GKX-0898.10	25D01 竹 118.12 篁 T7-4132
47	26065 糸 120.8 緬 T6-6038	260B2 糸 120.9 緬 GHZ-53423.10
48	26288 网 122.4 罥 T5-2B29	2668B 肉 130.5 罥 GHZ-32060.06

49	2633D 网 122.18 [illegible] GHZ-42929.12	27B80 言 149.16 [illegible] T7-5C6A
50	26842 肉 130.13 [illegible] GKX-0996.07 [illegible] T4-582A [illegible] VN-26842	26866 肉 130.13 [illegible] V2-8222
51	26913 自 132.6 [illegible] GHZ-53048.11	2EBB4 鼻 209.0 [illegible] JMJ-059276
52	269AD 臼 134.14 [illegible] T7-492F	269A6 臼 134.12 [illegible] GKX-1006.04
53	26B60 艸 140.5 [illegible] GHZ-53194.03	26B6B 艸 140.5 [illegible] T6-4D68
54	26B86 艸 140.6 [illegible] GKX-1031.22 [illegible] T6-4422	26C28 艸 140.7 [illegible] GHZ-53215.12
55	26E41 艸 140.11 [illegible] GHZ-53260.04	26DAB 艸 140.10 [illegible] T6-6149
56	27098 艸 140.16 [illegible] GKX-1068.25 [illegible] T7-4F5B	270F1 艸 140.17 [illegible] GHZ-53331.04

57	285BD 辵 162.12 邉 GKX-1265.16	285ED 辵 162.13 邉 邉 GKX-1266.29 T7-3449
58	285FF 辵 162.13 邈 GHZ-63890.09	28605 辵 162.13 邏 T7-3C5A
59	286FB 邑 163.8 郱 T6-4568	286F2 邑 163.8 郱 GKX-1273.08
60	28EED 阜 170.10 隑 GHZ-64149.08	28EEF 阜 170.11 隑 T6-6364
61	29157 雨 173.11 霧 霧 GHZ-64076.07 VN-29157	2916C 雨 173.12 霧 T7-5132
62	29FCE 鳥 196.4 鵷 J4-7E25	29FD7 鳥 196.4 鵷 鵷 GHZ-74621.13 T7-2B74
63	2B741 一 1.5 丽 JH-JTAD0F	2CEC7 一 1.5 丽 JMJ-059297
64	2C0C4 木 75.8 桌 TD-296B	2DAA2 木 75.8 桌 SAT-02597

65	FA23 走 156.4 赴 赴 赴 GU-0FA23 JA-2728 UTC-00850 • a CJK unified ideograph	27EAF 走 156.4 赴 TF-3862

END.