

Integration Panel Considerations on the 20151125 Chinese Root LGR

Integration Panel

December 17, 2015

1 Summary

This document analyses the content of the current Chinese LGR (CLGR5) as provided by the Chinese Generation Panel, and specified in the following two files dated November 25th 2015:

- CGP LGR 20151125.xlsx,
- CGP20151018-mod.xml

The Integration Panel analysis can be summarized as follows. The CGP LGR repertoire:

- is very close in code point repertoire to the full set of CJK Unified Ideographs in MSR-2 (CLGR5: 19 731 versus MSR2: 19 850).
- includes all the Hong Kong core characters that are part of the HK IICORE subset included in the Basic Multilingual Plane (BMP) (as do MSR-2 and the .asia ZH set)
- does not include the same category of HK characters, if they are encoded on the Supplementary Ideographic Plane (SIP),
- does not include some Singapore characters and other miscellaneous characters that are part of the .asia ZH set (7 characters),
- adds 94 Kanji core characters (J0 subset), leaving only 50 excluded from the set (from a total of 6 356 characters) without providing a rationale for the split decision.

This leads to two observations:

1. In light of the rather small difference between the latest CGP LGR draft and the MSR-2 repertoire of Unified CJK Ideographs, the integration panel is interested in obtaining a more detailed rationale for the selection of code points to be excluded, as none was found in the files provided.
2. Alternatively, the Chinese GP might consider adopting a much smaller and more restricted repertoire for the root; for example, using the Chinese subset of IICORE, which could be simpler to specify, in particular with regards to variants.

For a more complete set of questions by the Integration Panel, please end of Section 4.2 (p. 8).

It should be noted that the variant set has not been analyzed by the Integration Panel. This will be done when the repertoire is agreed and confirmed.

2 Definitions

2.1 IICORE collection

The International Ideographs Core (IICORE) is a fixed collection of CJK Ideographic code points deemed essential to all IRG Asian constituencies except Vietnam (a total of 7 sources). It contains 9 810 code points and is part of both ISO/IEC 10646 and Unicode. It was created by IRG based on priority (A to C, A being the highest) among its 7 sources.

2.2 MSR-2 CJK repertoire

The CJK repertoire in MSR-2 consist of 19 850 CJK Unified Ideographs, corresponding to the union of the following sub-repertoires:

- 1) .asia Japanese https://www.iana.org/domains/idn-tables/tables/asia_ja_1.1.txt
- 2) .asia Chinese https://www.iana.org/domains/idn-tables/tables/asia_zh_1.1.txt
- 3) IICORE as defined in Unicode 6.3
- 4) Code point U+9DC0.

The .asia Chinese repertoire is itself the result of a merge of the repertoires corresponding to various Chinese sources such as China PRC, Hong Kong SAR, and Taiwan.

Note that MSR-2 contains a few additional code points tagged with “sc:Hani” but that are not considered CJK Ideographs (for example U+3005 IDEOGRAPHIC ITERATION MARK and U+3006 IDEOGRAPHIC CLOSING MARK).

3 Existing IDNA practice for Chinese repertoires

3.1 IDNA .cn and .tw CJK Unified Ideograph repertoire

There was no single, well-established standard on which to establish a script-based repertoire for Chinese. Chinese repertoires have been independently created for Traditional and Simplified Chinese. However, the need to harmonize the process of simplified and traditional variants has lead to common repertoires in the context of IDNA.

The Chinese repertoires defined by the .cn registry [CNNIC] and the .tw registry [TWNIC] are identical and contain 19 520 Hanzi ideographs from two Unicode blocks: CJK UNIFIED IDEOGRAPHS and CJK UNIFIED IDEOGRAPHS EXTENSION A. They include the core repertoires for both simplified and traditional Chinese, expressed as follows:

IRG Source name	size	Included in .cn and .tw	Excluded
G0	6 763	6 763	0
G1	2 202	2 201	1
HB0	10	10	0
HB1	5 401	5 401	0
HB2	7 650	7 650	0

T1	5 414	5 413	0
T2	7 650	7 650	0

Other secondary Hanzi sources, such as G3, G5, GE, GH, GK, and T3 are also widely represented in that repertoire. Sources listed are not mutually exclusive, except for levels inside the same group G, T, H.

Note: The one exception to the inclusion of the core sets is the code point U+9DC0 (鸚) part of the G1 set which should have been the traditional variant of U+9E5A (鸚), but has been commonly replaced by U+9DBF (鸚) in that role. See a specific consideration below.

3.2 DotAsia .asia ZH repertoire

Because the repertoire defined by .cn and .tw does not fully cover the need of some Chinese constituencies such as Hong Kong SAR and Singapore, a new IDNA repertoire was created by DotAsia (.asia registry) for Chinese use (.asia ZH). It is a superset of these sets. It adds 163 Hanzi ideographs, of which 156 are part of HKSCS (Hong Kong Supplementary Character Set) included in the IICORE collection, 4 are GS (Singapore Characters), and the remaining 3 are part of various other Chinese sources but necessary to insure full transitivity in variant processing.

This makes a total of 19 683 Hanzi ideographs.

Note: This LGR contains 62 code points from the block: CJK UNIFIED IDEOGRAPHS EXTENDED B, added through the inclusion of the 156 characters from HKSCS.

Note: By being such a large union of various Hanzi ideographs, these repertoires include almost all the core Japanese Kanji set (6 212 Kanji ideographs out of a total of 6 356 for the J0: JIS X 0208-1990 set).

4 Current Chinese Root LGR 2015-10-18 (CLGR5)

The latest repertoire proposed by the GP has a simple relationship to the .asia ZH set.

4.1 69 code points removed compared to other tables

The current LGR draft does not include the 62 code points from the IICORE HKSCS subset which are located in CJK UNIFIED IDEOGRAPHS EXTENSION B, and the 7 non-HKSCS Hanzi that were added by DotAsia to the union of .cn and .tw sets. These 7 code points are all part of CJK Extension A and are shown here:

39DB
手 64.5

拏
GS-2240

43D3
肉 130.4

肱 肱 肱
GKX-0974.19 T3-2A50 K3-2F69

3BA3
木 75.9

槩 槩
GHZ-21257.01 T3-4167

4443
肉 130.14

朦 朦 朦
GKX-0997.08 T3-5866 K3-3039

4882
车 159'.5
𨋖
GS-235A

4C9E
鱼 195'.5
𩺰
GS-224E

4C9D
鱼 195'.4
𩺱
GS-224C

Four of them (U+39DB, U+4882, U+4C9D, and U+4C9E) are part of the GS set (Singapore Characters), and were likely added to satisfy requirements from the Singaporean community.

The last three (U+3BA3, U+43D3, and U+4443) have variants with other .cn code points, although they may have been simply added on their own for basic needs. They all have T3 source which is a well-represented source in MSR-2 and .cn IDNA table (over 50% coverage).

4.2 117 code points added compared to other tables

CLGR5 adds 117 characters for a total of 19 731 code points. (Note that this is only 119 less than the full set of CJK Unified Ideographs in MSR-2).

These 117 code points are classified as follows:

In CJK UNIFIED IDEOGRAPHS EXTENSION A (10 code points)

3960
心 61.9
惓 惓 惓
G5-5435 T3-3B5F JA-2329

惓
K3-2554

4137
禾 115.5
秝 秝 秝
G5-5B24 T3-3275 K3-2D4F

48B5
邑 163.4
邳 邳 邳
G5-6F4F T3-2736 J4-7A27

邳
K3-3368

48BC
邑 163.6
邳 邳
G3-706A T4-2B56

48C5
邑 163.7
郟 郟
G3-6F29 T4-3034

48D3
邑 163.10
郟 郟
G3-7B67 T4-4134

49D1
阜 170.8
阽 阽
GKX-1352.16 T3-395D

4A12
雨 173.6
霪 霪 霪
G5-734C T3-4A4A K3-3455

4CB3
鳥 196.4
鴝 鴝 鴝
GKX-1484.04 T3-5028 JA-282A

4D08
鳥 196.13
鷓 鷓 鷓
G3-5E36 T4-6C52 V2-8A27

In CJK UNIFIED IDEOGRAPHS (107 code points)

4E55
丩 4.6
𠄎 𠄎
GE-2130 J0-6949

4FAD
人 9.6
𠄎 𠄎
GE-2168 J0-4B79

51E6
几 16.3
𠄎 𠄎 𠄎
GE-227E J0-3D68 K2-244F

5227
刀 18.5
刼 刼 刼 刼 刼
GE-2330 H-89A6 T3-245D J0-5167 K1-582E

524F 刀 18.7	𠄎	𠄎	𠄎	𠄎	𠄎	63BB 手 64.8	搔	搔	搔
	GE-2339	H-87BC	T3-2B43	J0-516C	K2-246B		GE-2B54	J0-415F	K2-3569
5271 刀 18.9	劒			劒		6442 手 64.10	掇	掇	
	GE-2340			J0-517B			GE-2B6A	J0-405D	
5368 卜 25.9	高		高		高	663B 日 72.5	昂	昂	昂
	G3-3055		T4-3064		K0-6059		GE-2C64	T3-2D22	K0-6469
56A2 口 30.15	囊			囊		663F 日 72.5	昞	昞	
	GE-2529			J0-4739			GE-2C65	J0-5B26	
576E 土 32.5	坩		坩	坩	坩	66FD 日 72.7	曾	曾	
	G5-3575		T3-2828	J1-3747	K0-535D		GE-2D33	J0-413E	
5817 土 32.9	堧		堧		堧	685C 木 75.6	桜	桜	桜
	G5-364F		T4-3669		K1-5D50		GE-2D78	J0-3A79	K2-3A6C
5841 土 32.9	罊			罊		685F 木 75.6	棧	棧	
	GE-2572			J0-4E5D			GE-2D7B	J0-3B37	
58CC 土 32.13	壤			壤		6AAA 木 75.13	櫟	櫟	櫟
	GE-2633			J0-3E6D			GE-2F2A	T3-5529	J0-5D2C
5BFE 寸 41.4	对		对	对	对	6C17 气 84.2	气	气	气
	GE-2774		T3-255C	J0-4250	K2-2E26		GE-2F75	J0-3524	K2-3F43
5C02 寸 41.6	専			専		6D9C 水 85.7	澆	澆	澆
	GE-2775			J0-406C			GE-303A	J0-4642	K2-4066
5CC0 山 46.5	岯		岯		岯	6E13 水 85.8	溪	溪	
	GE-2835		T4-2634		K0-617C		GE-304F	J0-374C	
5D5C 山 46.9	寄			寄		6E7A 水 85.9	淝	淝	淝
	GE-2851			J0-563F			GE-305B	TF-3C5D	K0-5C48
5E3F 巾 50.9	幬		幬		幬	6E8C 水 85.9	澆	澆	澆
	GE-287A		T5-377C		K0-7D2C		GE-3065	J0-482E	K2-4158
5F10 弋 56.3	弋		弋	弋	弋	7114 火 86.8	焰	焰	焰
	GE-2944		TF-224E	J0-4675	K2-3122		GE-316E	J0-316B	K2-444A
6060 心 61.6	恠		恠	恠	恠	713C 火 86.8	烧	烧	烧
	GE-2A21		T3-2C5F	J0-575F	K1-5A26		GE-3177	J0-3E46	K2-4459
60A9 心 61.7	恼		恼	恼	恼	7155 火 86.9	熙	熙	熙
	GE-2A30		T3-3121	J0-473A	K2-324D		GE-317E	J0-5F66	K2-4466
61F4 心 61.15	懺			懺	懺	72A0 牛 93.13	犧	犧	犧
	GE-2A6C			J0-5873	K2-337A		GE-3271	J0-353E	K2-4637
633F 手 64.7	插		插	插	插	731F 犬 94.8	狺	狺	狺
	GE-2B40		T3-313C	J0-415E	K2-3538		GE-3323	J0-4E44	K2-4676

732F 犬 94.9	猫 G5-467E	猫 T3-3D2D	猫 J0-604E	猫 K2-467E	7D9A 糸 120.7	統 GE-383A	統 J0-4233	統 K2-5250
7363 犬 94.12	獸 GE-332F		獸 J0-3D43	獸 K2-473C	7E4A 糸 120.11	織 GE-3858	織 J0-4121	織 K2-5330
74A2 玉 96.13	璫 GE-3353	璫 T3-516B	璫 J0-6065	璫 K2-484E	7E4B 糸 120.13	繫 GE-3859	繫 J0-3752	繫 K2-5331
750E 瓦 98.11	甄 GE-336D	甄 H-FEA9	甄 T3-516E	甄 J0-612F	甄 K1-6C24	8133 肉 130.7	腦 GE-3963	腦 J0-473E
754A 田 102.4	畊 GE-3379	畊 H-FEB2	畊 T3-2E28	畊 J0-6139	畊 K1-5854	81D3 肉 130.14	臟 GE-3974	臟 J0-4221
7573 田 102.7	畷 GE-3427		畷 J0-3E76	畷 K2-493E	8217 舌 135.9	舖 GE-3A24	舖 J0-4A5E	
75E9 疒 104.7	瘦 GE-3435		瘦 J0-4169		8262 舟 137.13	艦 GE-3A2F	艦 H-956B	艦 T3-5A7A
7807 石 112.4	砭 G5-577A	砭 T3-2E3B			839F 艸 140.7	苔 G3-692F	苔 T3-3857	苔 J0-6831
783F 石 112.5	砭 GE-3533		砭 J0-395C	砭 K2-4C56	83B5 艸 140.7	菟 GE-3A5A	菟 J0-6839	菟 K2-5669
784F 石 112.6	研 GE-3534	研 T3-3761	研 J3-7923	研 K0-665A	86CD 虫 142.5	蚩 GE-3C2A	蚩 J0-3756	
7A36 禾 115.10	穢 GE-3634	穢 T4-4B6B			874B 虫 142.8	蠅 GE-3C3F	蠅 J0-4F39	蠅 K2-5C49
7A4F 禾 115.11	穩 GE-363B	穩 T3-522C	穩 J0-323A	穩 K2-4E78	877F 虫 142.9	蠅 GE-3C4B	蠅 J0-4768	蠅 K2-5C65
7A63 禾 115.13	穰 GE-3640		穰 J0-3E77		88B5 衣 145.6	袵 GE-3D29	袵 H-8FC4	袵 T3-3E72
7AD7 立 117.4	竅 GE-365E	竅 TF-2C76			894D 衣 145.12	襪 GE-3D3E	襪 H-95A7	襪 T3-587C
7ADA 立 117.5	竅 GE-365F	竅 H-8E56	竅 T3-3323	竅 J0-636C	8A33 言 149.4	訖 GE-3D63	訖 J0-4C75	訖 K2-6038
7B86 竹 118.8	篔 GE-3724	篔 T3-4860	篔 J0-4A4F	篔 K2-5030	8AAD 言 149.7	誦 GE-3D75	誦 TF-4C41	誦 J0-4649
7C14 竹 118.11	簞 GE-373F	簞 TF-5B5F	簞 J0-6443		8B0C 言 149.10	訶 GE-3E25	訶 H-8FDE	訶 T3-564A
7C4F 竹 118.14	籊 GE-374B	籊 TF-6567	籊 J0-6459	籊 K2-512C	8B72 言 149.13	讓 GE-3E37	讓 J0-6B68	讓 K1-572F
							讓 J0-3E79	

8E99 足 157.16	躡 GE-3F39	躡 T3-604B	躡 J0-6D38	躡 K1-5F6D	984B 頁 181.9	顛 GE-453C	顛 H-90F8	顛 T3-5959	顛 J0-707B	顛 K1-6569
8F0C 車 159.6	輻 GE-3F6E		輻 J0-6D52		98EE 食 184.4	飲 GE-454E		飲 J0-5D3B	飲 K0-6B66	
8F19 車 159.8	輒 GE-3F73	輒 H-904C	輒 T3-4F36	輒 J0-6D4D	9A12 馬 187.8	騷 GE-462C		騷 J0-417B		
9039 辵 162.8	達 GE-4043		達 J0-6D7D		9A13 馬 187.8	驗 GE-462D		驗 J0-3833	驗 K2-6F2C	
9197 酉 164.9	醜 GE-407E		醜 J0-4830		9A28 馬 187.9	駢 GE-4635		駢 J0-424D		
91A4 酉 164.10	醬 GE-4124		醬 J0-3E5F		9C04 魚 195.9	鰓 GE-4737	鰓 T3-5D68	鰓 J0-725A	鰓 K2-713E	
91B8 酉 164.13	釀 GE-4126		釀 J0-3E7A		9C2E 魚 195.10	鯁 GE-4742	鯁 T3-5F2A	鯁 J0-725B	鯁 K1-6846	
91C8 采 165.4	积 GE-412C		积 J0-3C61	积 K2-6757	9C76 魚 195.15	鱻 GE-4750	鱻 T3-6175	鱻 J0-7268	鱻 K2-717B	
9271 金 167.5	鉉 GE-4167		鉉 J0-395B	鉉 K2-6849	9D0E 鳥 196.4	鷓 GE-475B		鷓 J0-322A		
9421 金 167.12	鐵 GE-4336		鐵 J0-6F45	鐵 K2-6A5A	9D2C 鳥 196.5	鶯 GE-4763		鶯 J0-3229		
945A 金 167.15	鑽 GE-4349	鑽 T3-6055	鑽 J0-6F54	鑽 K2-6A75	9D8F 鳥 196.8	鷄 GE-4825		鷄 J0-375C		
9665 阜 170.7	陷 GE-4429		陷 J0-3459		9DC0 鳥 196.10	鷓 G1-704B	鷓 H-9AA5	鷓 T4-6843	鷓 J3-7E62	鷓 K2-7333
967A 阜 170.8	隄 GE-442D		隄 J0-3831	隄 K2-6B6F	9E78 鹵 197.8	鹵 GE-4857		鹵 J0-3834		
96B2 阜 170.14	隄 GE-443D	隄 T3-5671	隄 J0-702D		9EB9 麦 199.8	麴 GE-486C		麴 J0-396D		
974D 雨 173.19	霽 GE-4464	霽 T3-666F	霽 J3-7D68		9EBA 麦 199.9	麵 GE-486D		麵 J0-4C4D		
982C 頁 181.6	頰 GE-4532		頰 J0-4B4B							

The analysis of these 117 additions is straightforward:

- 22 code points correspond to those IICORE characters included neither in the DotAsia ZH set nor as part of the J0 Japanese core set (see Section 4.2.2 below)
- 94 code points are part of the J0 Japanese core set.

Note: many of these 94 code points have an additional GE source (GB 16500-95), but that simply denotes a horizontal ‘extension’ in the CJK Main block, not a true source. Once this addition is taken into account, only 50 code point of the JO set are excluded from CLGR5 (see Section 4.2.1 below).

- Code point U+9DC0 is included, as required to complete a variant set (see MSR-2).

Questions from the Integration Panel:

- The IP would like to know the rationale for adding the 22 IICORE characters that are not used in existing IDN tables (see Section 4.22). We note that over half of them are included in IICORE for support of Korean.
- The IP would like to learn the reasoning by the CGP with regards to deciding which code points from the JO set to include. We were unable to identify a pattern to distinguish the 94 that were selected from the 50 that were excluded. For example, we can see that a few of these 50 have glyph shapes are distinctively ‘Japanese’ but on the other hand, many are shared with other IRG sources.
- The IP wonders whether it is wise not to include the 7 characters that were deemed useful by DotAsia for Singapore and other Chinese constituencies (listed explicitly in Section 4.1).
- The IP would like to see a more detailed analysis of the 62 characters IICORE encoded in the block: CJK UNIFIED IDEOGRAPHS EXTENSION B and which are not included in the CLGR draft. These are all HKSCS characters and are specific to Cantonese. The main reason that they are encoded in Extension B (instead of the main block or Extension A) is because they were processed later by the IRG. The IP feels that any rationale for excluding them (whether or not this is because of their encoding in a Supplementary Plane), would need to be specifically documented by the Chinese Generation Panel.

4.2.1. JO code points not included in CLGR5 (50 code points)

List of JO core Japanese characters that are excluded from CLGR5 (50 code points):

4E8A 亅 6.6	事	事	599B 女 38.4	婁	婁	婁
	GE-213B	J0-502F		GE-265A	T3-2553	J0-562C
4EED 人 9.3	仞	仞	5AD0 女 38.10	媪	媪	媪
	GE-214F	J0-5041 K2-215F		GE-2733	T3-406A	J0-554C K2-2D2D
5050 人 9.9	倭	倭	5BC9 宀 40.8	霍	霍	
	G5-324C	T3-3444 J0-5074 K2-2273		G5-3E2A	J0-5565	
51A9 冫 14.12	冫	冫	5F09 冫 55.7	冫	冫	
	GE-2271	TF-465C J0-5150 K2-2437		GE-2943	J0-5732	
5338 匚 23.0	匚	匚	6121 心 61.9	惛	惛	惛
	G8-287A	H-C6C7 T3-2129 J0-523E K2-2563		GE-2A3E	T3-3B63	J0-583B K2-3279

616F 心 61.11 傷 傷 傷 傷 傷
G5-535D H-A074 T3-466A J0-5853 K2-333C

6282 手 64.4 扌 扌 扌 扌 扌
GE-2B23 H-9FF4 T3-262A J0-5930 K2-3452

65D9 方 70.12 播 播
GE-2C55 TF-5464 J0-5A5A

67A9 木 75.4 忝 忝
GE-2D50 J0-5B40

6859 木 75.6 梓 梓 梓 梓
GE-2D75 T3-3160 J0-5B62 K2-3A6B

68BA 木 75.7 禁 禁
GE-2E28 J0-5B72

69F9 木 75.11 棹 棹 棹 棹 棹
GE-2E67 H-FD57 T3-4C5F J0-5C5F K1-5871

6A0C 木 75.11 檣 檣 檣 檣 檣
G5-444A H-986C T3-4C50 J0-5C6E K2-3C62

6BB1 歹 78.15 殲 殲
GE-2F60 J0-5D54

6E76 水 85.9 淙 淙 淙 淙
G5-5630 H-92B3 T3-3C6C J0-5E59

76B9 皮 107.9 輝 輝 輝
GE-3457 T3-482E J0-6231

7724 目 109.5 昵 昵 昵 昵 昵
GE-3468 H-A0C1 T3-3265 J0-6242 K2-4B2B

799D 示 113.10 褖 褖 褖 褖 褖
G5-5768 H-FEF5 T3-4D50 J0-6335 K1-6E2A

7B7A 竹 118.7 筐 筐
GE-3723 J0-6421

7BCF 竹 118.9 箵 箵 箵
G5-635A H-9F5D J0-643D

7CAB 米 119.6 糲 糲 糲
GE-3769 TF-3E21 J0-6468

7DE4 糸 120.9 縹 縹 縹 縹
G5-6C40 H-8EAE T3-4E24 J0-654A K2-5268

80AC 肉 130.4 肱 肱 肱 肱
G5-4E73 T3-2A59 J0-666F K2-547D

815F 肉 130.9 腔 腔 腔 腔
G5-4F54 T3-437E J0-6735 K2-5560

822E 舟 137.4 舩 舩
GE-3A27 J0-6764

84AD 艸 140.10 葛 葛 葛 葛 葛
G5-6764 H-9AF2 T3-4942 J0-685A K1-704B

8C7C 豸 153.4 豸 豸 豸 豸
GE-3E4B T3-392E J0-6C41 K2-623C

8C8E 豸 153.8 貌 貌 貌 貌
GE-3E4F T3-4E7A J0-6C3F K2-6247

8CCD 貝 154.6 贓 贓 贓 贓 贓
GE-3E5C H-9E48 T3-4467 J0-6C5C K2-6259

8EE3 車 159.4 輿 輿
GE-3F67 J0-6D44

8F5C 車 159.14 輻 輻 輻 輻 輻
G5-477D H-9F78 T3-5E67 J0-6D5F K2-6544

900E 走 162.7 迺 迺 迺 迺
G3-7164 T3-394B J0-6E25 K2-6567

91D6 金 167.2 釳 釳 釳 釳 釳
G5-7537 H-FC5E T3-3430 J0-6E5B K2-675B

91E1 金 167.2 釜 釜
GE-4131 J0-6E5D

91F6 金 167.3 鈹 鈹 鈹 鈹
G5-7539 H-8C50 T3-3959 J0-6E61

91FC 金 167.3 釳 釳
GE-413B J0-6E5F

922C 金 167.4 釳 釳 釳
GE-414C J0-6E65 K2-6828

93E5 金 167.11 鎔 鎔 鎔
GE-427A T3-5B4A J0-6F35

941A 金 167.12 鏹 鏹 鏹 鏹
GE-4332 T3-5D44 J0-6F3C K2-6A57

95A0 門 169.5 閨 閨 閨 閨
GE-4368 H-90B5 TF-457B J0-6F60

9764 面 176.5 醜 醜 醜 醜
G5-7761 T3-4A4E J0-7052 K2-6C59

9D2A 鳥 196.5	鵠 GE-4761	鵠 鵠 T3-5432 J0-7272	9E95 鹿 198.8	麇 麇 麇 麇 麇 G5-7A22 H-91AB T3-5C29 J0-734B K2-7379
9DC6 鳥 196.10	鷓 GE-4837	鷓 鷓 J0-7336 K2-7336	9F21 鼠 208.0	兕 兕 GE-487D J0-736B

4.2.2. IICORE code points not included in CLGR5 repertoire (22 code points)

By design, MSR-2 set contains the full IICORE set. Of the 9 810 IICORE code points, these 22 are the only ones not part of the union of the two DotAsia repertoires investigated as part of MSR-2. A bit over half of this set is included in IICORE to support Korean.

Extension A (3400-4DBF):

Code point, Radical, Stroke count, Sources	Priority-Source(s)
3960 心 61.9 惓 惓 惓 惓 G5-5435 T3-3B5F JA-2329 K3-2554	C-K
4137 禾 115.5 秝 秝 秝 G5-5B24 T3-3275 K3-2D4F	C-K
48B5 邑 163.4 邳 邳 邳 邳 G5-6F4F T3-2736 J4-7A27 K3-3368	C-G
48BC 邑 163.6 邶 邶 G3-706A T4-2B56	C-G
48C5 邑 163.7 郕 郕 G3-6F29 T4-3034	C-G
48D3 邑 163.10 郟 郟 G3-7B67 T4-4134	C-G
49D1 阜 170.8 隄 隄 GKX-1352.16 T3-395D	C-G
4A12 雨 173.6 霽 霽 霽 G5-734C T3-4A4A K3-3455	C-K
4CB3 鳥 196.4 鷓 鷓 鷓 GKX-1484.04 T3-5028 JA-282A	C-T
4D08 鳥 196.13 鷓 鷓 鷓 G3-5E36 T4-6C52 V2-8A27	C-T

* In this column, '-K' stands for K source

Main block (4E00-9FFF)

Code point, Radical, Stroke count, Sources	Priority-Source(s)*
--	---------------------

Code point, Radical, Stroke count, Sources	Priority-Source(s)*			
5368 卜 25.9 高 G3-3055	高 T4-3064	高 K0-6059	A-KP	
576E 土 32.5 垠 G5-3575	垠 T3-2828	垠 J1-3747	垠 K0-535D	A-KP
5817 土 32.9 垠 G5-364F	垠 T4-3669	垠 K1-5D50	C-P	
5CC0 山 46.5 岙 GE-2835	岙 T4-2634	岙 K0-617C	A-KP	
5E3F 巾 50.9 幙 GE-287A	幙 T5-377C	幙 K0-7D2C	A-KP	
663B 日 72.5 昂 GE-2C64	昂 T3-2D22	昂 K0-6469	A-KP	
6E7A 水 85.9 淝 GE-305B	淝 TF-3C5D	淝 K0-5C48	A-KP	
7807 石 112.4 砭 G5-577A	砭 T3-2E3B		C-K	
784F 石 112.6 研 GE-3534	研 T3-3761	研 J3-7923	研 K0-665A	A-KP
7A36 禾 115.10 穢 GE-3634	穢 T4-4B6B	穢 K0-6973	A-KP	
7AD7 立 117.4 竝 GE-365E	竝 TF-2C76	竝 K0-595F	A-KP	
974D 雨 173.19 雩 GE-4464	雩 T3-666F	雩 J3-7D68	C-J	

* In this column, ‘-K’ stands for K source, and ‘-P’ stands for KP source. Therefore here ‘-KP’ means both North and South Korea sources. In other instances, such as source references, the term ‘KP’ refers to North Korea (DPRK) originated characters.

---end