

Title: New Unihan Database property: **kUnihanCore2020**

Author: Ken Lunde

Date: 2019-12-02

Per [L2/18-066R2](#), I previously proposed what I considered to be modest changes to the existing *kIICore* property, mainly to address some shortcomings that were identified in a series of five [CJK Type Blog](#) articles (appended to this document). Given the reluctance on the part of some national bodies to accept such modest changes, I decided to instead propose via [L2/18-279R](#) a completely new Unihan Database property that releases the set from being hampered by memory constraints that may have been applicable 15 years ago, but which arguably no longer apply to modern environments.

The new Unihan Database property name is *kUnihanCore2020*, which includes as part of its name the year in which the first version of Unicode that would include this new property is released, specifically Version 13.0. The attached *unihancore2020-data.txt* data file provides all of the property data, which covers 20,652 CJK Unified Ideographs and 68 CJK Compatibility Ideographs. Compared to the existing *kIICore* property, the proposed *kUnihanCore2020* property includes 10,910 additional ideographs. 22 ideographs that have a *kIICore* property value failed to meet the criteria for the *kUnihanCore2020* property, but have been grandfathered. The following table lists these 22 grandfathered ideographs, their *kIICore* property value, and the source reference that corresponds to their source tag:

Code Point	kIICore Property Value	Corresponding Source Reference
U+3960 憂	CK	K3-2554
U+4137 稹	CK	K3-2D4F
U+48B5 郔	CG	G5-6F4F
U+48C5 郕	CG	G3-6F29
U+48D3 郚	CG	G3-7B67
U+49D1 𩎧	CG	GKX-1352.16
U+4A12 霉	CK	K3-3455
U+4CB3 鳩	CT	T3-5028
U+4D08 鷗	CT	T4-6C52
U+593D 奮	CK	K2-2B54
U+5D44 嵘	CK	K2-2F33
U+5F34 弇	CJ	J13-7436
U+5F45 剻	CJ	J13-743A
U+66A3 瞰	CK	K1-5B6F
U+713F 煢	CT	T3-6552
U+7807 犔	CK	<i>none</i>
U+7A66 積	CK	<i>none</i>
U+974D 鬪	CJ	J3-7D68
U+974F 鬪	CJ	J13-7D6A
U+9964 钝	CG	G8-2D43
U+997E 钽	CG	G8-2D48
U+9AD9 高	CJ	<i>none</i>

Also see the attached *grandfathered-22.txt* data file.

The seven sections that follow describe the scope of each of the seven supported source tags, which are the same as those used by the existing *kIICore* property.

G—PRC

The scope of the “G” source tag is the union of the GB 2312 (6,763), TGH-2013/通用规范汉字表/*Tōngyòng Guīfàn Hànzìbiǎo* (8,105—see the [kTGH](#) property), and 现代汉语通用字表/*Xiàndài Hànyǔ Tōngyòngzìbiǎo* (7,000) standards, which results in 8,241 unique ideographs, all of which are CJK Unified Ideographs. This figure is only 136 ideographs more than TGH-2013 itself. The following six ideographs were grandfathered from the *kIICore* property and use the “G” source tag: U+48B5 邰 (G5), U+48C5 邯 (G3), U+48D3 邼 (G3), U+49D1 院 (GKX), U+9964 钅 (G8) & U+997E 钫 (G8). The total number of ideographs with the “G” source tag is therefore 8,247.

SPECIAL NOTES: 22 existing *kIICore* ideographs with the “G” source tag are excluded, because they are outside the scope of the three specified standards, but are included via other source tags. See the attached *excluded-g-22.txt* data file.

H—Hong Kong SAR

The scope of the “H” source tag is the union of the Big Five (13,060—see the [kBigFive](#) property) and HKSCS (4,603) standards, which results in 17,663 unique ideographs, 11 of which are CJK Compatibility Ideographs. There is no overlap between these two standards.

J—Japan

The scope of the “J” source tag is the union of the JIS X 0208 (6,356), 常用漢字/*Jōyō Kanji* (2,136—see the [kJoyoKanji](#) property), 人名用漢字/*Jinmei-yō Kanji* (863—see the [kJinmeiyoKanji](#) property), and 表外漢字/*Hyōgai Kanji* (1,022) standards, which results in 6,485 unique ideographs, 58 of which are CJK Compatibility Ideographs. This figure is only 129 more ideographs than JIS X 0208 itself. The following five ideographs were grandfathered from the *kIICore* property and use the “J” source tag: U+5F34 弹 (J13), U+5F45 弩 (J13), U+974D 鬪 (J3), U+974F 鬪 (J13) & U+9AD9 高 (no *kIRG_JSource*). The total number of ideographs with the “J” source tag is therefore 6,490.

SPECIAL NOTES: One existing *kIICore* ideograph with the “J” source tag is excluded, because it is outside the scope of the four specified standards, but is included via other source tags. See the attached *excluded-j-1.txt* data file.

K—ROK

The scope of the “K” source tag is the union of the KS X 1001 (4,620) and 한문 교육용 기초 한자/漢文教育用基礎漢字/*Hannun Gyojug-yong Gicho Hanja* (1,800—see the [kKoreanEducationHanja](#) property) standards, which results in 4,632 unique ideographs, all of which are CJK Unified Ideographs. This figure is only 12 more ideographs than KS X 1001 itself. The following eight ideographs were grandfathered from the *kIICore* property and use the “K” source tag: U+3960 惇 (K3), U+4137 柏 (K3), U+4A12 霽 (K3), U+593D 奄 (K2), U+5D44 嵴 (K2), U+66A3 瞠 (K1), U+7807 破 (no *kIRG_KSource*) & U+7A66 穢 (no *kIRG_KSource*). The total number of ideographs with the “K” source tag is therefore 4,640.

SPECIAL NOTES: 126 existing *kIICore* ideographs with the “K” source tag are excluded, because they are outside the scope of the two specified standards, but are included via other source tags. See the attached *excluded-k-126.txt* data file.

M—Macao SAR

The scope of the “M” source tag is the union of the Big Five standard (13,060—see the [kBigFive](#) property) and the existing *kIICore* ideographs that have the “M” source tag (4,954), which results in 13,119 unique ideographs, all of which are CJK Unified Ideographs. This figure is only 59 more ideographs than Big Five itself.

SPECIAL NOTES: Only one existing *kIICore* ideograph with the “M” source tag, U+5F66 彥, is excluded for reasons explained in the [2018-02-15 CJK Type Blog article](#), but is covered by four of the other six source tags (G, J, K & P): *Only one ideograph, U+5F66 彥, stands out as odd in that its source references do not suggest Macao SAR use. Its related ideograph, U+5F65 彴, is also tagged “M” in kIICore (ATHM), and its source references, particularly T1-507D, more strongly suggest Macao SAR use.* See the attached *excluded-m-1.txt* data file.

P—DPRK

The scope of the “P” source tag is the KPS 9566 (4,653) standard, which means that this is unchanged from *kIICore*.

T—ROC

The scope of the “T” source tag is the union of the CNS 11643 Levels 1 & 2 (13,064) and Big Five (13,060—see the [kBigFive](#) property) standards, which results in 13,065 unique ideographs, all of which are CJK Unified Ideographs. The following three ideographs were grandfathered from the *kIICore* property and use the “T” source tag: U+4CB3 鳩 (T3), U+4D08 鷗 (T4) & U+713F 煥 (T3). The total number of ideographs with the “T” source tag is therefore 13,068.

SPECIAL NOTES: 90 existing *kIICore* ideographs with the “T” source tag are excluded, because they are outside the scope of the two specified standards, but are included via other source tags. See the attached *excluded-t-90.txt* data file.

No Priority Tags

Because the notion of priority is largely source-specific, the *kUnihanCore2020* property does not have a provision to specify priority tags. The author of the proposal felt that they are not necessary, and that the source tags are sufficient.

CJK Compatibility Ideographs

Although the *kUnihanCore2020* property specifies source tags for 68 CJK Compatibility Ideographs—11 with the “H” source tag, and 57 with the “J” source tag—it is expected that their corresponding SVSes (*Standardized Variation Sequences*) be used in actual implementations. In addition, the CJK Compatibility Ideographs that correspond to the Big Five (2) and KS X 1001 (268) standards have been intentionally excluded, because they represent genuine duplicate ideographs. See the attached *svs-68.txt* data file that provides a correspondence between these 68 CJK Compatibility Ideographs and their SVSes.

That is all.

CJK Type Blog

CJK Fonts, Character Sets & Encodings. All CJK. #AllOfTheTime.

[HOME](#) > Exploring IICore—Part 1

By Dr. Ken Lunde

 [Comments \(0\)](#)

Created

February 5, 2018

Exploring IICore—Part 1



Today's article is the very first one that references [IICore \(*International Ideographs Core*\)](#), which is best described as a region-agnostic subset that includes the most commonly used CJK Unified Ideographs in Unicode, and is intended for use in memory-challenged devices and environments. Included are 9,810 ideographs, the bulk of which are in the URO (9,706), with the remaining ones in Extensions A (42) and B (62).

IICore is instantiated as the [kIICore](#) property of the Unihan Database, and documented in [UAX #38](#). The kIICore property consists of an initial letter—A, B, or C—that indicates priority, followed by one or more letters that specify a source that more or less corresponds to a region: G, H, J, K, M, P (short for KP), and T.

In Part 1 of what may eventually become a multiple-part series about IICore, I will briefly explore the ideographs that are tagged “K” for Korean use, along with pointing out some that should have been tagged “K” after examining the mappings to the KS X 1001 standard.

A total of 4,744 ideographs are tagged “K” in their kIICore property values. Of these, 138 are outside of KS X 1001. We'll come back to them at the end of this article.

It is very curious that only 14 of the 4,620 ideographs that are included in the KS X 1001 standard are not tagged “K” in their kIICore property values, yet are included in kIICore. The table below lists them and their kIICore property values, along with a related ideograph, if any:

Ideograph	kIICore	Related Ideograph	kIICore
U+585E	AGTJHMP	塞 U+F96C	n/a
U+596C	AP	U+734E	ATHKM
U+5FA9	ATJHMP	復 U+F966	n/a
U+6144	ATJHMP	慄 U+F9D9	n/a
U+6200	ATHMP	戀 U+F990	n/a
U+649A	ATJHMP	撲 U+F991	n/a
U+6817	AGTJHMP	栗 U+F9DA	n/a
U+6E17	AG	U+6EF2	ATJHKMP
U+8009	AP	U+8008	CK
U+80C4	AGTJP	U+5191	ATJHK
U+8A70	ATJHMP	NONE	n/a
U+8AFE	ATJHMP	諾 U+F95D	n/a
U+8F26	ATJHP	輦 U+F998	n/a
U+9ED8	AGTHMP	U+9ED9	AJK

Eight of the ideographs can be explained by guessing that an initial version of IIICore may have included the corresponding CJK Compatibility Ideographs that were subsequently stripped out. Another five—U+734E, U+6EF2, U+8008, U+5191 & U+9ED9—can be explained because they were apparently the preferred code points for the very popular HWP (*Hangul Word Processor*) app (according to Jaemin Chung), which was likely used to enter the ideographs by those who compiled the list for Korea (ROK). The only possible explanation for U+8A70 seems to be because it happens to be the very last hanja (aka ideograph) in the KS X 1001 standard, and may have felt victim to an inadvertent off-by-one error.

The obvious fix here is to simply tag the 14 characters on the left column of the table with “K” in their kIICore property values, which will make KS X 1001 support complete, and the best part is that it will not change the number of ideographs in IIICore.

Going back to the 138 ideographs outside of KS X 1001 that are tagged “K” in their kIICore property values, it turns out that the following seven do not have a kIRG_KSource property value, which raises the proverbial red flag :

Ideograph	kIICore—Claimed K-Source (in IRG N1025)	Source References
U+5AB4	CK—K3	G5-4047, HB2-DD43, T2-4249
U+741F	CK—K3	G3-3F59, H-98CA, KP1-5945, T3-3D35

U+74A4	CK—K3	GE-3354, H-FC71, T3-6567
U+74B8	BTK—K3	G3-3F71, HB2-F040, KP1-59CB, T2-622D
U+7807	CK—K3	G5-577A, KP1-5FAC, T3-2E3B
U+7A66	CK—K3	GE-3642, KP1-62B1, T3-5A65
U+9ED9	AJK—K0	GE-4874, J0-4C5B, T4-5560

Unfortunately, the people who compiled the “K” portion of IICore either passed away or are no longer participating in the Korean National Body, compounded by the fact that there is no document nor report explaining how the “K” portion of IICore was prepared, so we may never know exactly why these seven ideographs were tagged “K” in their kIICore property values, as well as the other 131 that are outside the scope of KS X 1001. Only U+9ED9 , which makes an appearance in both tables, can be explained by being the preferred code point for the HWP app.

About the 131 K-tagged IICore ideographs that are outside the scope of KS X 1001, 79 have K1 (aka KS X 1002) source prefixes, 48 have K2 (aka KS X 1027-1), and only four have K3 (aka KS X 1027-2).



[Careers](#) [Permissions & Trademarks](#) [EULAs](#) [Report Piracy](#) [Contact Adobe](#) [Security](#)

Copyright © 2019 Adobe Systems Incorporated. All rights reserved.

[Terms of Use](#) | [Privacy Policy and Cookies](#) ([Updated](#))

AdChoices

CJK Type Blog

CJK Fonts, Character Sets & Encodings. All CJK. #AllOfTheTime.

[HOME](#) > Exploring IICore—Part 2

By Dr. Ken Lunde

 [Comments \(0\)](#)

Created

February 7, 2018

Exploring IICore—Part 2



In Part 1 of this series, which is intended to scrutinize the 9,810 CJK Unified Ideographs that comprise IICore, we explored some of the oddities that related to ROK (aka South Korea). In Part 2 of this series, we will explore the ideographs that are tagged “P” and “J” for DPRK (aka North Korea) and Japan use, respectively.

DPRK—North Korea

To my surprise, there was nothing at all odd to be found here. The number of ideographs in IICore that are tagged “P” for DPRK use is 4,653. Those code points perfectly matched the 4,653 ideographs that correspond to the KPS 9566 standard, and whose kIRG_KPSource property values use the “KP0” source prefix.

Short and sweet.

Japan

Japan, on the other hand, is a bit more complex, but thankfully not nearly as complex as the ROK situation that was described in Part 1 of this series.

The number of ideographs in IICore that are tagged “J” for Japan use is exactly 4,600. 4,567 of these 4,600 ideographs correspond to JIS X 0208, and if we look only at JIS Level 1, 2,950 of its 2,965 ideographs are included in IICore. Only 15 of them—U+5147

, U+5283 , U+540B , U+54E9 , U+5678 , U+5C61 , U+6994 , U+6D6C , U+79A6 , U+7BAA , U+7CCE , U+86CE , U+91C6 , U+91E6 & U+976D —have been excluded.

Of the 33 J-tagged IICore ideographs that are outside of JIS X 0208, 31 are in JIS X 0213. Of these 31 ideographs, 28—U+52DB , U+53B2 , U+5733 , U+5861 , U+5DB8 , U+5F34 , U+5F45 , U+6673 , U+6A94 , U+6D31 , U+7006 , U+7028 , U+752F , U+76CC , U+7C1E , U+7D53 , U+7FDF , U+82B7 , U+8A79 , U+8D1B , U+8EC0 , U+9127 , U+95A9 , U+974D , U+974F , U+9DD7 , U+9EB4 & U+9F94 —are in JIS Level 3, and only three—U+5E2E , U+60F2 & U+7AD1 —are in JIS Level 4.

So far, so good.

The first oddity is that there are two ideographs that are tagged “J” yet do not have a kIRG_JSource property value. Interestingly, I pointed these out in [an article](#) from last month. The table below provides the details:

Ideograph	kIICore	Source References
U+9592	AGTJHKMP	HB1-B6A2, KP0-F2D8, K0-7959, T1-6267, V2-907C
U+9AD9	CJ	GE-464C, KP1-8B29, T4-362D

The second and final oddity—easily explained because IICore was developed before Japan’s Jōyō Kanji (常用漢字) list was revised at the end of 2010 to expand from 1,945 to 2,136 ideographs—is that there are three ideographs in that set that are not tagged “J” in IICore. All three of these ideographs do have kIRG_JSource property values that correspond to JIS X 0213, specifically JIS Level 3. The details are in the table below:

Ideograph	kIICore	Source References
U+525D	ATHKMP	GE-233B, HB1-ADE9, J3A-2F7E , KP0-DCD6, K0-5A4E, T1-544C, V1-4D2A
U+9830	ATHKMP	G1-3C55, HB1-C055, J13-7D7A , KP0-F3DF, K0-7A7A, T1-727E
U+20B9F	n/a	GKX-0173.01, H-8D40, J3A-4F54

Of course, U+525D and U+9830 could easily be tagged “J” in IICore without increasing its repertoire. And, U+20B9F is a candidate to be added to IICore.

Stay tuned for Part 3 of this series...



[Careers](#) [Permissions & Trademarks](#) [EULAs](#) [Report Piracy](#) [Contact Adobe](#) [Security](#)

Copyright © 2019 Adobe Systems Incorporated. All rights reserved.
[Terms of Use](#) | [Privacy Policy and Cookies](#) ([Updated](#))



CJK Type Blog

CJK Fonts, Character Sets & Encodings. All CJK. #AllOfTheTime.

[HOME](#) > Exploring IICore—Part 3

By Dr. Ken Lunde

 [Comments \(0\)](#)

Created

February 10, 2018

Exploring IICore—Part 3



In [Part 1](#) and [Part 2](#) of this series, we examined and scrutinized the ideographs that are tagged “K” (for ROK or South Korea), “P” (for DPRK or North Korea), and “J” (for Japan) in the kIICore property. In Part 3, which is today’s article, we will explore the 5,825 ideographs that are tagged “G” (for PRC or China).

The good news is that all of the ideographs that are included in the most common sets for China—the first 3,500 ideographs in [通用规范汉字表 \(Tōngyòng Guīfàn Hánzìbiǎo or TGH 2013\)](#) and the 3,755 ideographs of GB 2312 Level 1—are tagged “G” in IICore. When I merged these two sets, which resulted in 3,874 unique ideographs, 1,951 are not accounted for.

When I explored the next most important sets of ideographs for China, I found that 1,787 of the remaining 1,951 ideographs are in the second set of ideographs of [通用规范汉字表 \(3,000\)](#), and 1,771 of them are among the 3,008 ideographs of GB 2312 Level 2. When merged, these two sets resulted in accounting for 1,847 ideographs of the remaining 1,951 ones, meaning that 104 are still not accounted for.

Finally, I found that 75 of the remaining 104 ideographs are in the third set of ideographs of 通用规范汉字表 (1,605), which a mere 29 unaccounted for. The tables below lists these 29 remaining ideographs, separated by kIRG_GSource source prefix:

Ideograph	kIICore	kIRG_GSource—GB/T 12345
濛 U+6FDB	AGTHM	G1-7855
矇 U+77C7	AGTHM	G1-7857
硃 U+7843	AGTHM	G1-7927
穀 U+7A40	AGTJHKMP	G1-7836

Ideograph	kIICore	kIRG_GSource—GB 7589 unsimplified forms
鄜 U+48C5	CG	G3-6F29
郿 U+48D3	CG	G3-7B67
勑 U+52BB	AGT	G3-333F
屨 U+5C4C	BGT	G3-3B53
𢂔 U+6793	AGTKP	G3-4066
𠂔 U+808F	CG	G3-305B
蹠 U+8E53	BGT	G3-7045
僚 U+9BC8	AGT	G3-3233

Ideograph	kIICore	kIRG_GSource—GB 7590 unsimplified forms
𠂔 U+48B5	CG	G5-6F4F
俠 U+4F15	AGTHM	G5-314F
睆 U+6665	AGKP	G5-496D
珮 U+73EE	AGTJHM	G5-4231
𠂊 U+753D	AGT	G5-5A23
𠂊 U+793D	BGT	G5-574C

Ideograph	kIICore	kIRG_GSource—GB 8565.2
晳 U+6673	AGJKP	G8-2D72 *
洩 U+6D29	AGTJKMP	G8-2F6B
濬 U+6FEC	AGTHKMP	G8-2D59 *
釤 U+9964	CG	G8-2D43
餹 U+997E	CG	G8-2D48

* = There is an issue with U+6673 晳 and U+6FEC 濬 in that the actual GB 8565.2 standard does not include characters at code points 0x2D72 (13-82) or 0x2D59 (13-57). These ideographs are actually present in [ISO-IR-165](#) at those code points. See Jaemin Chung's [IRG N2276](#) for more details.

Ideograph	kIICore	kIRG_GSource—GB/T 16500
郤 U+537B	AGTHM	GE-237B
埶 U+5775	AGTKP	GE-2554
罾 U+776A	AGT	GE-3471
蹠 U+8E60	AGTJKP	GE-3F43
閒 U+9592	AGTJHKMP	GE-4361

Ideograph	kIICore	kIRG_GSource—
陮 U+49D1	CG	GKX-1352.16

Below is a modified version of the fifth table, which includes the five ideographs whose source references use the “GE” prefix, and which adds other source references from other properties. GB/T 16500 is interesting in a couple of ways. First and foremost, its 3,778 ideographs are simply meant to “fill in” URO (*Unified Repertoire & Ordering*) code points that otherwise lacked a kIRG_GSource property value, so they are effectively GBK characters. Second, as [this tweet](#) reports, the first two hexadecimal digits of **all** 3,778 source references are low by exactly 0x0F, and the source references in the table below reflect the corr

Ideograph	kIRG_GSource	Other Source References
郤 U+537B	GE-327B	HB1-AB6F, J0-524A, KP1-38C9, K1-5730, T1-5033, V1-4D7A
埶 U+5775	GE-3454	HB2-CBFA, J14-2468, KP0-D0EB, K0-4F26, T2-257A, V0-3438
罾 U+776A	GE-4371	HB1-B841, J14-7227, KP1-5E72, K2-4B4C, T1-6548

蹠 U+8E60	GE-4E43	HB2-F0F9, J0-6D28, KP0-EDA4, K0-7432, T2-6364
閒 U+9592	GE-5261	HB1-B6A2, KP0-F2D8, K0-7959, T1-6267, V2-907C

The fact that these five ideographs are tagged “G” in IICore is interesting, because on one hand their presence in the GB/T 16500 standard may suggest that they are not actually used in China, but on the other hand, they may actually be used in some specific contexts. At least, they are tagged with not only “G,” but with at least one or more additional tags.

Stay tuned for Part 4 of this series...



[Careers](#) [Permissions & Trademarks](#) [EULAs](#) [Report Piracy](#) [Contact Adobe](#) [Security](#)

Copyright © 2019 Adobe Systems Incorporated. All rights reserved.

[Terms of Use](#) | [Privacy Policy and Cookies](#) ([Updated](#))

AdChoices

CJK Type Blog

CJK Fonts, Character Sets & Encodings. All CJK. #AllOfTheTime.

[HOME](#) > Exploring IICore—Part 4

By Dr. Ken Lunde

[Comments \(0\)](#)

Created

February 15, 2018

Exploring IICore—Part 4



In [Part 1](#), [Part 2](#), and [Part 3](#) of this series, we examined and scrutinized the ideographs that are tagged “K” (for ROK or South Korea), “P” (for DPRK or North Korea), “J” (for Japan), and “G” (for PRC or China) in the kIICore property. In Part 4, which is today’s article, we will explore the ideographs that are tagged “T” (for ROC or Taiwan), “H” (for Hong Kong SAR), and “M” (for Macao SAR).

ROC—Taiwan

A total of 6,566 ideographs are tagged “T” in IICore. When I compared these against the two most basic ideograph sets from Taiwan—the 5,401 ideographs in CNS 11643 Plane 1 and the 4,808 ideographs in 常用國字標準字體表 (*chángyòng guózì biāozhǔn zìtǐ bìǎo*)—I discovered that only one, U+5F5E 羿, is neither tagged “T” nor present in IICore, though its related ideograph that is included in Big Five Level 1, U+5F5D 翌, is tagged “T” in IICore. (This ideograph pair represents the only difference between CNS 11643 Plane 1 and Big Five Level 1, both of which include 5,401 ideographs.)

Other than the one omission pointed out in the previous paragraph, 1,156 ideographs remain outside the scope of what is a reasonably minimal set. Predictably, most of them—1,063 to be exact—map to CNS 11643 Plane 2, which is equivalent to Big Five Level 2, and another 81, two of which—U+3577 咏 and U+4CB3 鶴—are in Extension A, map to CNS 11643 Plane 3.

That leaves a mere 12 T-tagged IICore ideographs outside the scope of the first three planes of CNS 11643. Six of them map to CNS 11643 Plane 4 (with half being in Extension A), one maps to Plane 5, and two map to Plane 15. The three tables below provide their details:

Ideograph	kIICore	KIRG_TSource—CNS 11643 Plane 4
鰩 U+4C81	CT	T4-697C
鯤 U+4C85	CT	T4-697B
鷄 U+4D08	CT	T4-6C52
獮 U+7374	CT	T4-566C
鶠 U+8025	CT	T4-462C
鯇 U+9BDD	CT	T4-625C

Ideograph	kIICore	KIRG_TSource—CNS 11643 Plane 5
鱸 U+9C72	BTH	T5-7A53

Ideograph	kIICore	KIRG_TSource—CNS 11643 Plane 15
畊 U+7551	ATJKP	TF-2B7A
鰈 U+9C47	ATJKP	TF-6A3E

The three remaining ideographs are the only somewhat suspicious ones in that they do not have a KIRG_TSource property value, but are related to ideographs that are tagged “T” in IICore and are in CNS 11643 Plane 1 or 2, per the table below:

Ideograph	kIICore	Other Source References	Related Ideograph
𠙴 U+55EC	BGTH	G0-6040, H-8F52	𠈌 U+5475
礡 U+7934	BGT	G0-6D67, H-FEE8, J13-7932, KP1-6109, K2-4D65	礡 U+7921
纖 U+7E4A	ATJ	GE-3858, J0-4121, KP1-67CC, K2-5330	纖 U+7E96

The only actions that I can suggest are to tag U+5F5E “T” in IICore, and for Taiwan to consider a horizontal extension for U+55EC 嘴, U+7934 磡, and U+7E4A 纖.

Hong Kong SAR

A total of 5,224 ideographs are tagged “H” in IICore. When I compared these against the 5,401 ideographs in Big Five Level 1, I

discovered that 577 are not included. This leaves 400 ideographs, 171 of which map to Big Five Level 2, and the remaining 229 map to Hong Kong SCS proper (24 are in Extension A, 61 are in Extension B, and the remaining 144 are in the URO).

All looks okay until we consider Hong Kong SCS-2016 that added 24 new characters, 22 of which are best described as the preferred Hong Kong SAR forms of existing Big Five ideographs. Of these 22 ideographs, 14 have corresponding Big Five versions that are tagged “H” in IICore, which strongly suggests that they should be tagged “H” if already present in IICore, or added to IICore and tagged “H.” The following table provides the details:

HKSCS-2016	kIICore	Big Five Level 1	kIICore
兑 U+5151	AG	兌 U+514C	ATJHKMP
告 U+543F	n/a	告 U+544A	AGTJHKMP
媯 U+5AAA	CG	媯 U+5ABC	ATJHKM
悅 U+60A6	AGJ	悅 U+6085	ATHKMP
慍 U+6120	CG	慍 U+614D	ATHM
氳 U+6C32	n/a	氳 U+6C33	ATH
稅 U+7A0E	AGJ	稅 U+7A05	ATHKMP
脫 U+8131	AGJ	脫 U+812B	ATHKMP
蘊 U+85F4	n/a	蘊 U+860A	ATJHKMP
蛻 U+8715	AG	蛻 U+86FB	ATHM
說 U+8AAC	AJ	說 U+8AAA	ATHKMP
醞 U+9196	n/a	醞 U+919E	ATHM
銳 U+92ED	AJ	銳 U+92B3	ATHKMP
閱 U+95B2	AJ	閱 U+95B1	ATHKMP

Macao SAR

A total of 4,955 ideographs are tagged “M” in IICore. When I compared these against the 5,401 ideographs in Big Five Level 1, I discovered that 739 are not included. This leaves 283 ideographs, 223 of which map to Big Five Level 2, and 59 of which map to HKSCS (two are in Extension A, eight are in Extension B, and the remaining 49 are in the URO). Only one ideograph, U+5F66 彦, stands out as odd in that its source references do not suggest Macao SAR use. Its related ideograph, U+5F65 彦, is also tagged “M” in IICore (ATHM), and its source references, particularly T1-507D, more strongly suggest Macao SAR use. The table below provides more details about these two ideographs:

kIICore—AGJKMP	Source References	kIICore—ATHM	Source References

彦 U+5F66

G0-5165, J0-4927, KP0-
F8BA, K0-6569, T3-2C50

彥 U+5F65

GE-2955, HB1-ABDB, KP1-
41F9, T1-507D

In addition, 13 of the 14 ideographs—meaning all except for U+6C32 彙—in the first column of the table in the “Hong Kong \$ section above should probably be tagged “M” in IICore, because Macao SAR has similar regional conventions, and because the ideographs in the third column are already tagged “M” in IICore.

Interestingly, I never mentioned anything about the kIRG_MSource property in the previous paragraph, because none of the M-tagged ideographs in IICore have such source references. Given that there is a fairly close relationship with Big Five and HKSCS, comparing against those sets seemed to be appropriate, and as it turned out, was completely appropriate.



[Careers](#) [Permissions & Trademarks](#) [EULAs](#) [Report Piracy](#) [Contact Adobe](#) [Security](#)

Copyright © 2019 Adobe Systems Incorporated. All rights reserved.

[Terms of Use](#) | [Privacy Policy and Cookies](#) ([Updated](#))

 AdChoices



CJK Type Blog

CJK Fonts, Character Sets & Encodings. All CJK. #AllOfTheTime.

[HOME](#) > Exploring IICore—Part 5

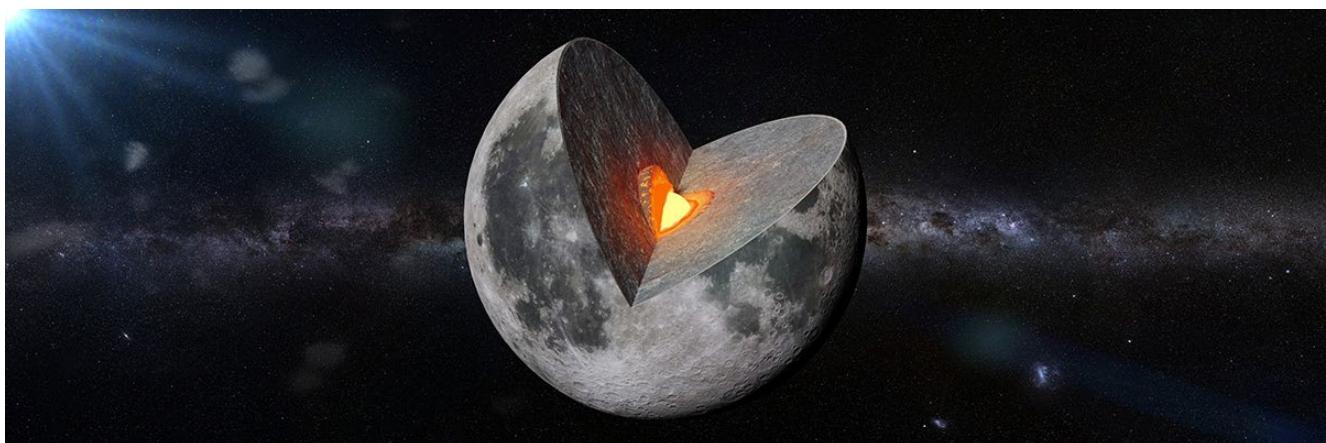
By Dr. Ken Lunde

 [Comments \(0\)](#)

Created

February 16, 2018

Exploring IICore—Part 5



[Part 1](#), [Part 2](#), [Part 3](#), and [Part 4](#) of this series scrutinized the ideographs that are associated with each of the seven region tags of the `kIICore` property. In this fifth and final article of this series, I will provide some details about the earlier versions of IICore, and what changed between them.

Version 1.1

IICore Version 1.1 was made available as [IRG N1025](#) during [IRG #21](#), which took place in November of 2003 in [Guilin](#) (桂林 *Guīlín*), China (PRC). It included 11,039 ideographs. 7,772 were Category A, 1,077 were Category B, and 2,190 were Category C.

Version 2.1

IICore Version 2.1 was made available as [IRG N1052](#) during [IRG #22](#), which took place in May of 2004 in [Chengdu](#) (成都 *Chéngdū*), China (PRC). It included 9,811 ideographs. 7,772 were Category A, 417 were Category B, and 1,622 were Category C. Compared to Version 1.1, 1,231 ideographs were removed (1,122 mapped to the URO, 100 mapped to Extension A, and the remaining nine mapped to Extension B), but the three ideographs shown in the table below were added:

Ideograph	kIICore	Source References
𩫚 U+4C85	CT	G3-7831, H-8759, KP1-8D7B, K3-3660, T4-697B
厍 U+5399	CM	G1-5847, HB2-CDF8, J1-3449, KP1-38DD, K2-2579, T2-293A, V2-8A5E
龍 U+8622	CM	G1-5C57, HB2-F4B8, J0-6944, KP1-705E, K1-5F25, T2-6964, V0-426B

Version 2.2

IICore Version 2.2, which is the current version, is available as [IRG N1067](#) on the [IRG website](#), and its mapping files are also available in the [IRG #23 document register](#). It is also a normative property of the Unihan Database. It includes 9,810 ideographs. 7,772 are Category A, 417 are Category B, and 1,621 are Category C. Compared to Version 2.1, only the following Category C ideograph was removed:

Ideograph	kIICore	Source References
氵 U+6C3D	CH	G0-595B, H-8ADB, KP1-5077, K2-3F48, V2-8D45

Its variant form, U+5C3F 尸, is included in IICore as a Category A ideograph with the following tags: GTJHKMP.

Special Note: The links that are associated with the three IRG documents for the three versions of IICore—[IRG N1025](#), [IRG N1052](#) & [IRG N1067](#)—are single PDFs that I prepared for convenience, and which also include the machine-readable mapping files as PDF attachments. What is provided on the IRG website are multiple files that need to be individually downloaded.



[Careers](#) [Permissions & Trademarks](#) [EULAs](#) [Report Piracy](#) [Contact Adobe](#) [Security](#)

Copyright © 2019 Adobe Systems Incorporated. All rights reserved.

[Terms of Use](#) | [Privacy Policy and Cookies](#) ([Updated](#))