Title: CJK & Unihan Group Recommendations for UTC #174 Meeting

Author: Ken Lunde Date: 2023-01-11

This document provides the CJK & Unihan Group recommendations for the UTC #174 meeting, based on a meeting that took place on 2023-01-06 from 6 to 9PM PDT, which was attended by Yi Bai, Eiso Chan, Lee Collins, Peter Constable, John Jenkins, Kushim Jiang, Ken Lunde, William Nelson, CheonHyeong Sim, Stephan Hyeonjun Stiller, Michel Suignard, Xieyang Wang, and Yifán Wáng via Zoom. The Chair of the CJK & Unihan Group, Ken Lunde, chaired the meeting with John Jenkins as Vice-Chair.

The CJK & Unihan Group reviewed public feedback and documents that were received since the UTC #173 meeting took place. Comments are marked in **green**, and Recommendations to the UTC are marked in **red**. A "plain text" Copy&Paste-ready version of the Recommendations is also provided as the PDF attachment named *recommendations-20230106.txt*.

Public Feedback

The eight items of public feedback from L2/23-007 that are in this section were discussed by the group, and for convenience, the following data files, which are PDF attachments, include the recommended Unihan database and UAX #45 data file additions and changes based on them:

- Unihan-additions-20230106.txt
- Unihan-changes-20230106.txt
- USource-changes-20230106.txt

The recommended Unihan database and UAX #45 data file additions and changes are also shown inline, immediately following the **Recommendations**.

01) 2022-11-21 07:03:04 CST

```
Date/Time: Mon Nov 21 07:03:04 CST 2022
```

Name: Andrew West

Report Type: Other Document Submission
Opt Subject: Unihan_IRGSources.txt (15.0)

Comments: The group discussed this feedback, and agreed that **46.10** is a more intuitive *kRSUnicode* property value for U+3032E that will aid in discoverability, and should therefore be added as a second property value.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Accept the proposal to change the kRSUnicode property value of U+3032E, based on feedback [Mon Nov 21 07:03:04 CST 2022] in document L2/23-007 and Section 01 of document L2/23-011, for Unicode Version 15.1.

Action Item for Michel Suignard, CJK: Change the kRSUnicode property value of U+3032E to add 46.10 as a second property value, based on feedback [Mon Nov 21 07:03:04 CST 2022] in document L2/23-007 and Section 01 of document L2/23-011, and provide to John Jenkins an updated Unihan data file, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Update the Unihan database to add 46.10 as a second kRSUnicode property value for U+3032E, based on feedback [Mon Nov 21 07:03:04 CST 2022] in document L2/23-007 and Section 01 of document L2/23-011, for Unicode Version 15.1.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Group comments in Section 01 of document L2/23-011. See feedback [Mon Nov 21 07:03:04 CST 2022] in document L2/23-007.

Current:

U+3032E kRSUnicode 39.10

Adjusted:

U+3032E kRSUnicode 39.10 **46.10**

02) 2022-12-04 07:35:47 CST

Date/Time: Sun Dec 4 07:35:47 CST 2022

Name: Ken Lunde

Report Type: Error Report

Opt Subject: Unihan_OtherMappings.txt

In order to synchronize the provisional kKSC0 and kKSC1 properties with the corresponding source prefixes of the normative kIRG_KSource property, the following kKSC0 and kKSC1 property values should be removed:

U+F92C kKSC0 5011 U+F9B8 kKSC0 7170 U+8C6C kKSC1 7575

They should be replaced with the following new kKSC0 and kKSC1 property values:

U+FA2E kKSC0 5011 U+FA2F kKSC0 7170 U+27CEF kKSC1 7575

The two kKSC0 property values became out-of-sync with the corresponding K0 source prefix property values from Unicode Version 6.0 (2010), and the kKSC1 property value became out-of-sync with the corresponding K1 source prefix property value from Unicode Version 13.0 (2020).

Another alternative is to consider removing the two provisional properties altogether, given that they are intended to be synchronized with the kIRG_ KSource property values that are associated with the K0 and K1 source prefixes,

respectively. The only difference is that kKSCO and kKSC1 properties specify decimal Row-Cell (aka ku/ten) values, and the KO and K1 source prefixes specify the corresponding hexadecimal values.

Comments: The group discussed this feedback, and based on the intended 1:1 correspondence between these two provisional properties and the K0 and K1 source prefixes of the normative *kIRG_KSource* property, the group recommends the alternative that was mentioned in the feedback, specifically to remove these two provisional properties altogether.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Remove the provisional Unihan database properties, kKSC0 and kKSC1, based on feedback [Sun Dec 4 07:35:47 CST 2022] in document L2/23-007 and as amended in Section 02 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, EDC: Remove the provisional Unihan database properties, kKSC0 and kKSC1, from UAX #38, based on feedback [Sun Dec 4 07:35:47 CST 2022] in document L2/23-007 and as amended in Section 02 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Remove the provisional Unihan database properties, kKSC0 and kKSC1, from the Unihan database, based on feedback [Sun Dec 4 07:35:47 CST 2022] in document L2/23-007 and as amended in Section 02 of document L2/23-011, for Unicode Version 15.1.

Action Item for Rick McGowan, UTC: Extend PRI #467, Proposed Update UAX #38, with a new closing date of 2023-04-04.

03) 2022-12-04 08:06:30 CST

Date/Time: Sun Dec 4 08:06:30 CST 2022

Name: Ken Lunde

Report Type: Error Report

Opt Subject: Unihan_OtherMappings.txt

In order to synchronize the provisional kHKSCS property with the corresponding source prefix of the normative kIRG_HSource property, H, the following kHKSCS property values should be removed:

U+3D1D kHKSCS 91B5

U+4CA4 kHKSCS 9D73

U+2F9B2 kHKSCS 8FA8

They should be replaced with the following new kHKSCS property values (note the ordering difference):

U+9FD0 kHKSCS 9D73

U+270F0 kHKSCS 8FA8

U+2A3ED kHKSCS 91B5

The first and second kHKSCS property values became out-of-sync with the corresponding H source prefix property values from Unicode Version 8.0 (2015),

and the third kHKSCS property value became out-of-sync with the corresponding H source prefix property value from Unicode Version 11.0 (2018).

Another alternative is to consider removing the provisional kHKSCS property altogether, given that it is intended to be synchronized with the kIRG_ HSource property values that are associated with the H source prefix. The only difference in their property values is the "H-" source prefix.

Comments: The group discussed this feedback, and based on the intended 1:1 correspondence between this provisional property and the H source prefixes of the normative *kIRG_HSource* property, the group recommends the alternative that was mentioned in the feedback, specifically to remove this provisional property altogether.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Remove the provisional Unihan database property, kHKSCS, based on feedback [Sun Dec 4 08:06:30 CST 2022] in document L2/23-007 and as amended in Section 03 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, EDC: Remove the provisional Unihan database property, kHKSCS, from UAX #38, based on feedback [Sun Dec 4 08:06:30 CST 2022] in document L2/23-007 and as amended in Section 03 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Remove the provisional Unihan database property, kHKSCS, from the Unihan database, based on feedback [Sun Dec 4 08:06:30 CST 2022] in document L2/23-007 and as amended in Section 03 of document L2/23-011, for Unicode Version 15.1.

04) 2022-12-04 12:00:37 CST

Date/Time: Sun Dec 4 12:00:37 CST 2022

Name: Ken Lunde

Report Type: Error Report

Opt Subject: Unihan_OtherMappings.txt

In order to synchronize the provisional kKPS1 property with the corresponding source prefix of the normative kIRG_KPSource property, KP1, the following 12 kKPS1 property values should be removed:

U+48D4 kKPS1 7E68 U+20ED4 kKPS1 3A79 U+23D3C kKPS1 5177 U+2435C kKPS1 5604 U+25864 kKPS1 628E U+25F0A kKPS1 662C U+27262 kKPS1 7164 U+2779A kKPS1 749F U+27AAC kKPS1 7755 U+28418 kKPS1 7D1B

```
U+28D4E kKPS1 8282
```

U+2A50C kKPS1 9186

They should be replaced with the following 188 new kKPS1 property values that include 12 that replace those listed above:

U+3E14 kKPS1 5DC7

U+431B kKPS1 7755

U+56D8 kKPS1 3749

U+6B76 kKPS1 61A9

U+7AD0 kKPS1 636D

U+9FC3 kKPS1 5E2B

U+F936 kKPS1 70DC

U+FA70 kKPS1 341D

U+FA71 kKPS1 347E

U+FA72 kKPS1 34D0

U+FA73 kKPS1 35DE

U+FA74 kKPS1 3714

U+FA75 kKPS1 3740

U+FA76 kKPS1 383E

U+FA77 kKPS1 3862

U+FA78 kKPS1 39E5

U+FA79 kKPS1 39EF

U+FA7A kKPS1 3A42

U+FA7B kKPS1 3A48

U+FA7C kKPS1 3BEE

U+FA7D kKPS1 3C51

U+FA7E kKPS1 3CAF

U+FA7F kKPS1 3CB2

U+FA80 kKPS1 3D45

U+FA81 kKPS1 3DFC

U+FA82 kKPS1 414E

U+FA83 kKPS1 416B

U+FA84 kKPS1 41FB

U+FA85 kKPS1 4244

U+FA86 kKPS1 4348

U+FA87 kKPS1 437E

U+FA88 kKPS1 4399

U+FA89 kKPS1 43FF

U+FA8A kKPS1 4410

U+FA8B kKPS1 4486

U+FA8C	kKPS1	4505
U+FA8D	kKPS1	469B
U+FA8E	kKPS1	46C4
U+FA8F	kKPS1	46F9
U+FA90	kKPS1	480E
U+FA91	kKPS1	4994
U+FA92	kKPS1	4A42
U+FA93	kKPS1	4A4B
U+FA94	kKPS1	4ABD
U+FA95	kKPS1	4F27
U+FA96	kKPS1	4FA9
U+FA97	kKPS1	5128
U+FA98	kKPS1	5281
U+FA99	kKPS1	52B4
U+FA9A	kKPS1	5336
U+FA9B	kKPS1	53FA
U+FA9C	kKPS1	5559
U+FA9D	kKPS1	561F
U+FA9E	kKPS1	5678
U+FA9F	kKPS1	5786
U+FAA0	kKPS1	57FE
U+FAA1	kKPS1	597A
U+FAA2	kKPS1	5A5D
U+FAA3	kKPS1	5ABB
U+FAA4	kKPS1	5BF5
U+FAA5	kKPS1	5BF9
U+FAA6	kKPS1	5D48
U+FAA7	kKPS1	5D5C
U+FAA8	kKPS1	5D91
U+FAA9	kKPS1	5DF3
U+FAAA	kKPS1	5E08
U+FAAB	kKPS1	6083
U+FAAC	kKPS1	6330
U+FAAD	kKPS1	6435
U+FAAE	kKPS1	659F
U+FAAF	kKPS1	66BA
U+FAB0	kKPS1	671B
U+FAB1	kKPS1	6873
U+FAB2	kKPS1	69B1

U+FAB3	kKPS1	6DBB
U+FAB4	kKPS1	6E6C
U+FAB5	kKPS1	7250
U+FAB6	kKPS1	7451
U+FAB7	kKPS1	74C8
U+FAB8	kKPS1	74D4
U+FAB9	kKPS1	769A
U+FABA	kKPS1	769E
U+FABB	kKPS1	76A4
U+FABC	kKPS1	76B3
U+FABD	kKPS1	76C2
U+FABE	kKPS1	76F1
U+FABF	kKPS1	76FD
U+FAC0	kKPS1	77CB
U+FAC1	kKPS1	7976
U+FAC2	kKPS1	7CAE
U+FAC3	kKPS1	7DF1
U+FAC4	kKPS1	7F3E
U+FAC5	kKPS1	80CE
U+FAC6	kKPS1	8340
U+FAC7	kKPS1	8404
U+FAC8	kKPS1	84E4
U+FAC9	kKPS1	862D
U+FACA	kKPS1	866A
U+FACB	kKPS1	869A
U+FACC	kKPS1	8705
U+FACD	kKPS1	8BAB
U+FACE	kKPS1	927F
U+FACF	kKPS1	441C
U+FAD0	kKPS1	443E
U+FAD1	kKPS1	4B26
U+FAD2	kKPS1	4CE2
U+FAD3	kKPS1	5DF6
U+FAD5	kKPS1	5EBB
U+FAD6	kKPS1	654F
U+FAD7	kKPS1	7A0A
U+FAD8	kKPS1	91E5
U+FAD9	kKPS1	9273
U+20617	kKPS1	49E2

U+20F25	kKPS1	3A79
U+20F96	kKPS1	3AA3
U+214A1	kKPS1	9186
U+2258D	kKPS1	7352
U+23D1F	kKPS1	5177
U+243DF	kKPS1	5604
U+24454	kKPS1	5642
U+24662	kKPS1	5717
U+24AF4	kKPS1	59FC
U+2564A	kKPS1	6134
U+25871	kKPS1	628E
U+25F45	kKPS1	6620
U+27261	kKPS1	7164
U+27762	kKPS1	749F
U+28416	kKPS1	7D1B
U+286CD	kKPS1	7E68
U+28D64	kKPS1	8282
U+2ABAB	kKPS1	4731
U+2ABE9	kKPS1	487E
U+2B0DC	kKPS1	6651
U+2B0EF	kKPS1	672B
U+2B0F5	kKPS1	6752
U+2B29B	kKPS1	70BE
U+2B2AB	kKPS1	712E
U+2B2FE	kKPS1	73E1
U+2F804	kKPS1	34EE
U+2F805	kKPS1	3534
U+2F833	kKPS1	38CF
U+2F835	kKPS1	54B7
U+2F84C	kKPS1	3A92
U+2F84F	kKPS1	3AD2
U+2F852	kKPS1	3BAF
U+2F855	kKPS1	3BD5
U+2F887	kKPS1	40D3
U+2F88B	kKPS1	4120
U+2F899	kKPS1	41F8
U+2F8A0	kKPS1	42C8
U+2F8A6	kKPS1	43C1
U+2F8A7	kKPS1	43D2

U+2F8AD	kKPS1	4462
U+2F8B1	kKPS1	44A4
U+2F8B4	kKPS1	4539
U+2F8B7	kKPS1	45D5
U+2F8BA	kKPS1	462A
U+2F8D0	kKPS1	49C3
U+2F8E0	kKPS1	4B46
U+2F8E1	kKPS1	4B5D
U+2F8E2	kKPS1	4B57
U+2F8E5	kKPS1	4BF7
U+2F8E6	kKPS1	4C6C
U+2F8FE	kKPS1	511E
U+2F900	kKPS1	5145
U+2F901	kKPS1	5150
U+2F907	kKPS1	51C4
U+2F912	kKPS1	53B8
U+2F922	kKPS1	56A8
U+2F926	kKPS1	57D0
U+2F938	kKPS1	5AF0
U+2F94E	kKPS1	6043
U+2F959	kKPS1	6246
U+2F95F	kKPS1	639A
U+2F96C	kKPS1	6725
U+2F99F	kKPS1	6E4B
U+2F9B8	kKPS1	7168
U+2F9BA	kKPS1	71CD
U+2F9D3	kKPS1	78AC
U+2F9DB	kKPS1	7AAD
U+2F9DC	kKPS1	7B04
U+2F9E8	kKPS1	8054
U+2F9EA	kKPS1	80AD
U+2F9EE	kKPS1	8255
U+2FA00	kKPS1	8703
U+2FA0D	kKPS1	8E7E
U+2FA1B	kKPS1	9190
U+31C2D	kKPS1	5AC6

In terms of the 12 kKPS1 property value mapping changes, the first one became out-of-sync with the corresponding KP1 source prefix property values from Unicode Version 3.2 (2002), and the remaining 11 became out-of-sync with

the corresponding KP1 source prefix property values from Unicode Version 5.2 (2009).

Another alternative is to consider removing the two provisional properties, kKPS0 and kKPS1, altogether, given that they are intended to be synchronized with the kIRG_KPSource property values that are associated with the KP0 and KP1 source prefixes, respectively. The only difference in their property values is the "KP0-" or "KP1-" source prefixes.

Comments: The group discussed this feedback, and based on the intended 1:1 correspondence between these two provisional properties and the KP0 and KP1 source prefixes of the normative *kIRG_KPSource* property, the group recommends the alternative that was mentioned in the feedback, specifically to remove these two provisional properties altogether.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Remove the provisional Unihan database properties, kKPS0 and kKPS1, based on feedback [Sun Dec 4 12:00:37 CST 2022] in document L2/23-007 and as amended in Section 04 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, EDC: Remove the provisional Unihan database properties, kKPS0 and kKPS1, from UAX #38, based on feedback [Sun Dec 4 12:00:37 CST 2022] in document L2/23-007 and as amended in Section 04 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Remove the provisional Unihan database properties, kKPS0 and kKPS1, from the Unihan database, based on feedback [Sun Dec 4 12:00:37 CST 2022] in document L2/23-007 and as amended in Section 04 of document L2/23-011, for Unicode Version 15.1.

05) 2022-12-18 08:34:02 CST

```
Date/Time: Sun Dec 18 08:34:02 CST 2022
Name: Ken Lunde
Report Type: Error Report
Opt Subject: CJKRadicals.txt
The following entry is missing from the UCD's CJKRadicals.txt data file:
148'; 2EC6; 89D2
It is the Chinese simplified form of Radical 148.
```

Comments: The group discussed this feedback, and instead of adding a new entry for the Chinese simplified form of Radical #148, 角, to the UCD's CJKRadicals.txt data file, the group instead recommends that the entry for 162', 162'; 2ECC; 8FB6, which corresponds to the Chinese simplified form of Radical #162, 辶, be removed from the UCD's CJKRadicals.txt data file, and that the corresponding kRSUnicode property value, 162'.0, be removed from U+8FB6 辶. The entry for the Chinese simplified form of Radical #162 in the UCD's CJKRadicals.txt data file served as the precedent for proposing a comparable entry for the Chinese simplified form of Radical #148. Furthermore, U+8FB6 辶 is the only ideograph that includes 162' as the first part of its kRSUnicode property value, meaning that it was likely added in error in Unicode Version 6.0 (2010). The group also noted that similar single-apostrophe treatment could be given to

Radicals #140 (** versus **) and #188 (骨 versus 骨), because their Chinese simplified forms are separately encoded in the CJK Radicals Supplement block as U+2EBE ** CJK RADICAL GRASS ONE and U+2EE3 骨 CJK RADICAL BONE, respectively, but doing so is not useful for searching or other operations.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Accept the proposal to remove the entry for 162' in the UCD's CJKRadicals.txt data file and to change the kRSUnicode property value of U+8FB6, based on Section 05 of document L2/23-011, for Unicode Version 15.1.

Action Item for Michel Suignard, CJK: Change the kRSUnicode property value of U+8FB6 by removing 162'.0, based on Section 05 of document L2/23-011, and provide to John Jenkins an updated Unihan data file, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Update the Unihan database to remove 162'.0 from the kRSUnicode property value of U+8FB6, based on Section 05 of document L2/23-011, for Unicode Version 15.1.

Action Item for Markus Scherer, PAG: Update the UCD's CJKRadicals.txt data file by removing the entry "162'; 2ECC; 8FB6," based on Section 05 of document L2/23-011, for Unicode Version 15.1.

Current:

U+8FB6 kRSUnicode 162.0 **162'.0**

Adjusted:

U+8FB6 kRSUnicode 162.0

06) 2022-12-25 19:10:34 CST

Date/Time: Sun Dec 25 19:10:34 CST 2022

Name: Paul Masson

Report Type: Error Report

Opt Subject: kPhonetic for U+645E

This character appears on p.350 of Casey but is not in a phonetic group.

It appears that the appropriate one is 842, if you want to add this to the database.

Thank you.

Comments: The group discussed this feedback, and agreed that **842*** would be an appropriate *kPhonetic* property value for U+645E 摞. Because this is a provisional property, John Jenkins already updated the Unihan database to reflect this addition.

Recommendations: The group recommends to the UTC the following disposition ⇒

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Group comments in Section 06 of document L2/23-011. See feedback [Sun Dec 25 19:10:34 CST 2022] in document L2/23-007.

07) 2022-12-27 19:01:03 CST

Date/Time: Tue Dec 27 19:01:03 CST 2022

Name: Eiso Chan

Report Type: Other Document Submission

Opt Subject: Change the fields for UTC-02994 in UAX #45 and kTotalStrokes for

U+44D5

The Status field for this character in UAX #45 should be changed to ExtA as below for the possible horizontal extension in future. On the other hand, we should also add 162.8 as the second RS of 选 (U+44D5), because the original RS is also reasonable for UTC-02994 based on the new evidence. For the General comments field, the kMandarin property value for 选 (U+44D5) has been assigned and useful for the comment Chinese use, and the Putonghua value as yuǎn is just only used for the use in Guangdong, Guangxi, Hong Kong and Macao, so it is better to remove the first letter k. The current kCantonese property is suitable for UTC-02994 as well.

UTC-02994;ExtA;U+44D5;140.8 162.8;;Ш辶芫;UTCDoc L2/17-062 14;Mandarin: yuǎn;12 10;1

Comments: The group discussed this feedback, and agreed that **162.8** should be added as a second *kRSUnicode* property value for U+44D5 远, that U+44D5 远 should be horizontally extended to add **UTC-02994** as its *kIRG_USource* property value, and that the record for UTC-02994 in the UAX #45 data file, *USourceData.txt*, should be adjusted per the feedback, but with further elaboration about its readings in the comment field. The group noted that the *kTotalStrokes* property value for U+44D5 远 was corrected during UTC #173, and currently reflects **12** as its property value, so no action is necessary. The table below constitutes a proposal to horizontally-extend U+44D5 by adding UTC-02994 as a new U-Source source reference:

		U-Source		
Code Point	Code Chart Excerpt—Version 15.0	Source Reference	Representative Glyph	
U+44D5	44D5 艸 140.8	UTC-02994	选	

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Accept the proposal to change the kRSUnicode property value of U+44D5 and add UTC-02994 as a new kIRG_USource property value for U+44D5, based on feedback [Tue Dec 27

19:01:03 CST 2022] in document L2/23-007 and as amended in Section 07 of document L2/23-011, for Unicode Version 15.1.

Action Item for Michel Suignard, CJK: Change the kRSUnicode property value of U+44D5 by adding 162.8 and add UTC-02994 as a new kIRG_USource property value for U+44D5, based on feedback [Tue Dec 27 19:01:03 CST 2022] in document L2/23-007 and as amended in Section 07 of document L2/23-011, and provide to John Jenkins an updated Unihan data file, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Update the Unihan database to add 162.8 to the kRSUnicode property value of U+44D5 and add UTC-02994 as a new kIRG_USource property value for U+44D5, based on feedback [Tue Dec 27 19:01:03 CST 2022] in document L2/23-007 and as amended in Section 07 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Update the record for UTC-02994 in the UAX #45 data file, USourceData.txt, based on feedback [Tue Dec 27 19:01:03 CST 2022] in document L2/23-007 and as amended in Section 07 of document L2/23-011, for Unicode Version 15.1.

Current:

U+44D5 kRSUnicode 140.8

Adjusted:

U+44D5 kRSUnicode 140.8 **162.8**

New:

U+44D5 kIRG_USource UTC-02994

Current (*USourceData.txt*):

UTC-02994;WS-2017;U+44D5;162.8;; 恒之芫;UTCDoc L2/17-062 14; kMandarin: yuǎn;12 10;1

Adjusted (*USourceData.txt*):

UTC-02994; ExtA; U+44D5; 140.8 162.8; 三上芫; UTCDoc L2/17-062 14; Mandarin: yuǎn; 12 10; 1

08) 2022-12-06 09:54:11 CST

Date/Time: Tue Dec 6 09:54:11 CST 2022

Name: Jaycee Carter

Report Type: Error Report

Opt Subject: Unihan_DictionaryIndices.txt

The following additions and changes to the Unihan kKangXi values appear to be possible, based on the scanned text of the Kangxi dictionary available at

https://www.kangxizidian.com/:

U+34D8: 0134.290 (compare source glyphs T5-3129 & UTC-00639)

U+385B: 0335.010 (compare source glyph T4-3D29)

U+3CC7: 0612.120 (compare source glyph T6-284A; appears to be unifiable by UCV

#235)

U+4945: 0735.130 1318.220

U+4A36: 1377.190 U+4DB9: 0304.130 U+4DBA: 0504.310

U+4DBB: 0505.010 U+4DBC: 0506.140 U+5173: 0127.070 0879.050 (compare source glyph T4-2234) U+537C: 0159.290 (compare source glyphs HB2-CDF5 & T2-2937) U+67D0: 0519.010 0528.090 U+67F5: 0521.070 U+6F5B: 0649.090 U+7D01: 0915.140 (compare source glyphs HB2-D04C & T2-2C6D; appears to be unifiable by UCV #46) U+8275: 1014.110 (appears to be unifiable by UCV #181) U+8286: 1018.030 (compare source glyph T3-2728; appears to be unifiable by UCV #152 & #235) U+833E: 1029.160 (compare source glyph T3-2A5F; appears to be unifiable by UCV U+84F1: 1052.220 (compare source glyphs HB2-DFA9 & T2-464E and compatibility ideograph U+2F9A8; appears to be unifiable by UCV #181) U+20020: 0887.260 (appears to be unifiable similarly to UCV #303) U+200CA: 1362.170 U+200CB: 0976.090 U+201DD: 0141.130 U+2028B: 1536.370 U+209C8: 0473.100 U+209E7: 1128.110 U+20A1D: 0382.030 U+20A81: 0794.220 U+20AA5: 0162.040 U+20B46: 0101.070 U+2174F: 0272.390 U+21D13: 0381.070 U+21D4F: 0295.040 U+21DA2: 0615.020 U+2215B: 1536.360 (appears to be unifiable by UCV #62) U+221CF: 0342.180 0362.280 (compare source glyph T6-5C4E) U+22597: 0872.030 U+22FF9: 0585.140 (compare source glyph T7-2351) U+23184: 1355.180 U+231F6: 0611.130 U+23353: 0331.140 U+2363B: 0547.100 (appears to be unifiable by UCV #152) U+23C9D: 0131.250

```
U+23E27: 0632.020 (appears to be unifiable by UCV #222)
U+24548: 0409.100 (appears to be unifiable by UCV #36)
U+24A61: 0744.190
U+253EC: 0192.140 0825.160
U+257A9: 0855.240 (appears to be unifiable by UCV #181)
U+257DF: 0859.080
U+2627D: 0946.220
U+26B12: 1018.050 (appears to be unifiable by UCV #152)
U+26FB1: 1063.110 1070.150 (compare source glyph T4-5D25; appears to be
unifiable by UCV #174)
U+2845F: 0456.100 (appears to be unifiable by UCV #245)
U+2865E: 1265.060
U+28EEA: 1354.080 (compare source glyph T7-3D33)
U+29272: 1526.020
U+2A594: 1530.100 (final stroke difficult to make out in copy)
U+2AEC5: 0714.190
U+2E4D6: 1072.400 (no Chinese source given in Unihan, but appears to be an
exact match)
U+31C2D: 0760.040 (compare compatibility ideograph U+2F936)
Additionally, please consider the following changes:
U+23551: 0539.240 0542.170 (appears potentially unifiable, as no semantic
difference)
U+26400: 1008.170 (presumably unifiable; compare 羽 U+7FBD, source glyph K0-
6962)
U+29692: 1416.220 (final stroke appears to be missing in copy, appearing
identical to 飣 (U+4B22; kKangXi 1416.050); however, context suggests this
character is intended; if so, left radical appears to be unifiable by UCV #358)
Many thanks.
```

Comments: The group discussed this feedback, and agreed that while these appear to be worthwhile additions and changes to the provisional *kKangXi* property, these proposed additions and changes would first need to be vetted.

Recommendations: The group recommends to the UTC the following disposition ⇒

Action Item for John Jenkins, CJK: Check the proposed kKangXi property value additions and changes, then make the appropriate additions and changes to the Unihan database, based on feedback [Tue Dec 6 09:54:11 CST 2022] in document L2/23-007 and Section 08 of document L2/23-011, for Unicode Version 15.1.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Group comments in Section 08 of document L2/23-011. See feedback [Tue Dec 6 09:54:11 CST 2022] in document L2/23-007.

UAX #38 / Unihan Database Documents

The documents in this section were discussed by the group, and for convenience, the following data files, all of which are PDF attachments, include all of the recommended UCD and Unihan database additions, changes, and removals based on them:

- CJKRadicals-20230106.txt
- Unihan-additions-20230106.txt
- Unihan-changes-20230106.txt
- Unihan-removals-20230106.txt

The recommended UCD and Unihan database additions, changes, and removals are also shown inline, immediately following the **Recommendations**.

09) L2/22-288: Proposal to enhance the provisional kBigFive property

Comments: The group discussed this proposal, which is for an enhancement to the provisional *kBigFive* property that flags two ideographs as being CNS 11643 equivalents, and agreed that it is helpful for font developers. A representative from TCA (*Taiwan Computer Association*) also provided positive feedback on this proposal via email.

Recommendations: The group recommends to the UTC the following disposition ⇒

Action Item for John Jenkins, EDC: Update the Syntax and Description of the provisional kBigFive property in UAX #38, based on document L2/22-288 and Section 09 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Update the Unihan database to add kBigFive property values to U+5F5E and U+7B9A, based on document L2/22-288 and Section 09 of document L2/23-011, for Unicode Version 15.1.

New:

U+5F5E kBigFive C255' U+7B9A kBigFive E35A'

10) L2/22-200: Proposal to enhance the kRSUnicode property for non-Chinese simplified radicals

Comments: The group discussed this proposal during the last meeting, and while implementing the recommended changes that the UTC accepted during the UTC #173 meeting, the Properties & Algorithms Group (aka PAG) identified an issue with the *CJKRadicals.txt* data file, specifically that treatment for the following two of the six non-Chinese simplified radicals were not covered by our recommendations in Section 14 of L2/22-247:

```
182''
208''
```

PAG asked our group to discuss this issue, and make further recommendations to the UTC. Like the four previous non-Chinese simplified radicals, the group recommends that entries for

these two non-Chinese simplified radicals be added to the UCD's *CJKRadicals.txt* data file with an empty second field (code point for the radical), because their corresponding characters are not included in the CJK Radicals Supplement block, and that they be included in the Unicode Version 15.1 Alpha. Other options were considered, such as A) repeating the third field (corresponding CJK Unified Ideograph code point) in the second field; or B) use the non-simplified radical code point in the second field. The group agreed that an empty second field is the cleaner approach, and makes it explicit that their corresponding characters are not included in the CJK Radicals Supplement block.

```
Option A:
```

```
182''; 322C4; 322C4
208''; 9F21; 9F21
Option B:
182''; 2FB5; 322C4
208''; 2FCF; 9F21
```

The group also does not recommend that these non-Chinese simplified radicals be added to the CJK Radicals Supplement block.

In terms of documenting an empty second field in the UCD's *CJKRadicals.txt* data file, the following is the relevant paragraph:

```
# There is one line per CJK radical number. Each line contains three # fields, separated by a semicolon (';'). The first field is the # CJK radical number. The second field is the CJK radical character. # The third field is the CJK unified ideograph.
```

The following reflects the recommended changes, which are additions shown in red:

```
# There is one line per CJK radical number. Each line contains three # fields, separated by a semicolon (';'). The first field is the # CJK radical number. The second field is the CJK radical character, # which may be empty if the CJK radical character is not included in # the Kangxi Radicals block or the CJK Radicals Supplement block. The # third field is the CJK unified ideograph.
```

Lastly, Ken Lunde should be given an action to draft documentation for the Unicode Version 15.1 Alpha, Beta, and landing pages that specifies that parsers need to be aware of these changes, specifically that 1) the notion of "simplified radical" is no longer a boolean; 2) a small number of *kRSUnicode* property records may specify two-apostrophe radicals numbers as the primary or only property value; 3) the second field of the UCD's *CJKRadicals.txt* data file can be empty, which breaks any assumption that the second field is not empty, and that the second field is a character from the CJK Radicals Supplement or Kangxi Radicals blocks; and 4) the third field of 182" in the UCD's *CJKRadicals.txt* data file is code point outside the URO, which breaks any assumption that the code point is in the URO.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Accept the two additions to the UCD's CJKRadicals.txt data file, based on Section 10 of document L2/23-011, for Unicode Version 15.1.

Action Item for Markus Scherer, PAG: Update the UCD's CJKRadicals.txt data file, based Section 10 of document L2/23-011 and its attached CJKRadicals-20230106.txt data file, for Unicode Version 15.1.

Action Item for Ken Lunde, EDC: Draft documentation for the Unicode Version 15.1 Alpha, Beta, and landing pages that specifies that parsers need to be aware of the presence of non-Chinese simplified radicals in kRSUnicode property values and in the UCD's CJKRadicals.txt data file, based on Section 10 of document L2/23-011, for Unicode Version 15.1.

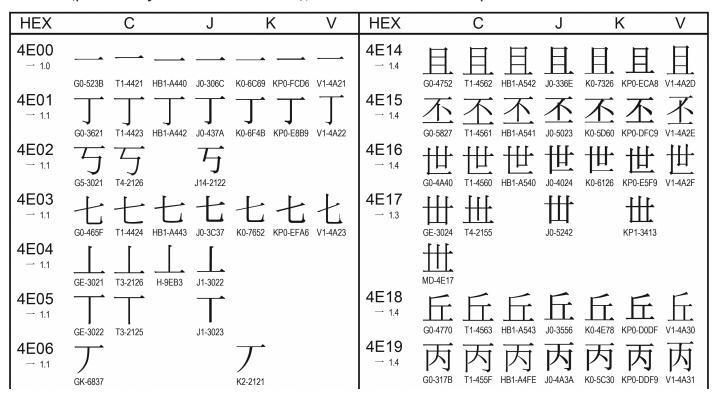
New (CJKRadicals.txt):

182''; ; 322C4 208''; ; 9F21

11) L2/22-238: Proposal to consider adding CodeCharts support for kIRG_KPSource representative glyphs

Comments: This proposal was discussed during the UTC #173 meeting, and Michel Suignard was assigned Action Item 173-A80: Add the remaining kIRG_KPSource representative glyphs to the code charts for the CJK Unified Ideographs (aka URO), Extension A, and Extension B blocks, based on document L2/22-238 and as amended in Section 35 of document L2/22-247, for a future version of the Unicode Standard.

Michel already did this work, and produced preliminary code charts for the URO, Extension A, and Extension B blocks that include representative glyphs for KP-Source ideographs. In particular, the format of the code charts for the URO has been enhanced to accommodate seven columns (previously it was six columns), and below is an excerpt:



This work resulted in the need to adjust the *kIRG_KPSource* property values for only eight CJK Unified Ideographs in those three blocks. The group, along with the authors of L2/22-238, agreed that these property value adjustments, which are shown in the table below, are appropriate.

Code Point	Ideograph	Current KP-Source	Proposed KP-Source	Notes
U+3F94	痦	(none)	KP1-5B5D	
U+48EE	酦	(none)	KP1-7EF4	
U+9166	酉发	KP1-7EF4	(none)	
U+23CC0	泙	(none)	KP1-50FB	
U+24D6A	溶	KP1-5B5D	(none)	
U+23CD9	涯	KP1-50FB	(none)	
U+249D6	環	KP1-594E	KP1-5961	
U+249E8	琊	KP1-596F	KP1-594E	KP1-596F is a duplicate of KP0-D7D7 (U+746F)

Now that the code charts for the URO can accommodate seven columns, Michel has a strong preference to target this work for Unicode Version 15.1, and the group agrees that this is reasonable. The group, which included the authors of L2/22-238, were provided preliminary code charts for the URO, Extension A, and Extension B blocks that incorporate the KP-Source ideographs, and the authors of L2/22-238 have been tasked to check them for possible issues.

Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Accept the kIRG_KPSource property value additions, changes, and removals, based on Section 11 of document L2/23-011, for Unicode Version 15.1.

Action Item for Michel Suignard, CJK: Apply the kIRG_KPSource property value additions, changes, and removals, based on Section 11 of document L2/23-011, and provide to John Jenkins an updated Unihan data file, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Update the Unihan database to reflect the kIRG_KPSource property value additions, changes, and removals, based on Section 11 of document L2/23-011, for Unicode Version 15.1.

Action Item for Ken Lunde, EDC: Draft documentation for the Unicode Version 15.1 Alpha, Beta, and landing pages that describes the addition of the KP-Source ideographs in the code charts for the URO, Extension A, and Extension B, based on Section 11 of document L2/23-011, for Unicode Version 15.1.

Current:

U+249D6 kIRG_KPSource KP1-594E U+249E8 kIRG_KPSource KP1-596F

Adjusted: U+249D6 U+249E8	kIRG_KPSource kIRG_KPSource	KP1-5961 KP1-594E
Remove: U+9166 U+23CD9 U+24D6A	kIRG_KPSource kIRG_KPSource kIRG_KPSource	KP1-7EF4 KP1-50FB KP1-5B5D
New: U+3F94 U+48EE U+23CC0	kIRG_KPSource kIRG_KPSource kIRG_KPSource	KP1-5B5D KP1-7EF4 KP1-50FB

12) L2/23-002: Proposal to add kFanqie to Unihan Database

Comments: The group discussed this proposal, which is a revision of L2/22-135 that served as a preliminary proposal for the same provisional Unihan database property that was discussed during the UTC #172 meeting. Overall, the group agreed that the proposal is in good shape, but requires some adjustments to the proposed Syntax, specifically to exclude the CJK Compatibility Ideographs Supplement block (2F800..2FA1F) and limit Plane 3 code points to those for existing blocks (Extensions G and H: 30000..323AF). In addition, a statement to the effect that *kFanqie* property data would serve as a good source of data for phonologists would help bolster the proposal. Lastly, the ideographs that serve as the property values have a close relationship with the *kSBGY* (*Song Ben Guang Yun*) property, which needs to be investigated before this new property could be accepted. In particular, some *kSBGY* property values may need to be adjusted in order to better align the two properties.

The proposal author will work with Lee Collins to prepare a revised proposal that the group will review during a future meeting.

Recommendations: The group recommends to the UTC the following disposition ⇒

No UTC action is required.

UAX #45 / U-Source Documents

The document in this section was discussed by the group, and for convenience, the following data file, which is a PDF attachment, includes all of the recommended UAX #45 *USourceData.txt* data file additions based on it:

USource-additions-20230106.txt

These recommended UAX #45 *USourceData.txt* data file additions are also shown inline, following the **Recommendations**. In addition, the *UTCHan2023u174-Regular.ttf* TrueType font (PDF attachment) includes the representative glyphs for adding to the UAX #45 *UTCHan.ttf* TrueType font.

13) L2/22-282: Proposal to add one common character used for Cantonese cuisine to UAX #45

Comments: The group reviewed this proposal, and agreed that adding this new ideograph to UAX #45 with a status value of **FutureWS** is reasonable. This ideograph, whose representative glyph is shown below, can therefore be considered as a candidate for inclusion in the UTC's submission for the next IRG working set:



Recommendations: The group recommends to the UTC the following disposition ⇒

Consensus: Accept one new U-Source ideograph as UTC-03336 with a UAX #45 status value of FutureWS, based on document L2/22-282 and Section 13 of document L2/23-011, for Unicode Version 15.1.

Action Item for John Jenkins, CJK: Add one new record to the UAX #45 USourceData.txt data file and its representative glyph to USourceGlyphs.pdf, based on document L2/22-282 and Section 13 of document L2/23-011, for Unicode Version 15.1.

New:

UTC-03336; FutureWS;;195'.8;; 三鱼孟; UTCDoc L2/22-282; kTraditionalVariant 艋;16;5

Other Documents

14) L2/22-274R: Disruptive Changes in GB 18030-2022 L2/23-003: Industry Recommendations for GB 18030 Testing & Certification (DRAFT)

Comments: The UTC discussed L2/22-274 during the UTC #173 meeting (L2/22-247R, an editorial revision, was issued after the meeting), which resulted in Action Item 173-A97: *Prepare a liaison communication from the Unicode Consortium to CESI*.

Ken Lunde, the Unicode Consortium's liaison to CESI (*China Electronics Standardization Institute*) and a representative from Apple, worked with representatives from Adobe, Google, and Microsoft to draft a two-page liaison communication that relays industry recommendations for improving the GB 18030 testing and certification process, which was posted to the UTC document register as L2/23-003. Ken Lunde will incorporate feedback from those representatives and from the group into L2/23-003R, which should be reviewed in plenary during the UTC #174 meeting.

Recommendations: The group recommends to the UTC the following disposition ⇒

No UTC action is required.

15) L2/22-277: Proposal to add Kangxi Radical annotations to the CJK Radicals Supplement block

Comments: The group discussed this proposal, which completes Action Item 173-A59: *Prepare a proposal for adding numbered annotations for the radicals in the CJK Radicals Supplement block, based on Section 14 of document L2/22-247.*

The group noted that the following four patterns are used in the proposed code chart annotations:

```
variant of Kangxi Radical [nnn] — 84 instances simplified variant of Kangxi Radical [nnn] — 5 instances simplified Chinese variant of Kangxi Radical [nnn] — 24 instances simplified Japanese variant of Kangxi Radical [nnn] — 4 instances
```

The group also noted that the following two characters include two annotations:

U+2EAB:

```
variant of Kangxi Radical 109
variant of Kangxi Radical 122
U+2EB2:
variant of Kangxi Radical 109
variant of Kangxi Radical 122
```

The group was overwhelmingly supportive of this proposal, and recommends that they be added to the UCD's *NamesList.txt* data file.

Recommendations: The group recommends to the UTC the following disposition ⇒

Action Item for Ken Whistler, EDC: Add the annotations for the 115 characters in the CJK Radicals Supplement block to the UCD's NamesList.txt data file, based on document L2/22-287 and Section 15 of document L2/23-011 and its attached NamesList-20230106.txt data file, for Unicode Version 15.1.

16) L2/22-279: UCS Seal Script Source Mapping Data

Comments: This document is intended as a resource for those who are involved in the encoding of Seal Script, and is therefore an FYI document for the UTC.

Recommendations: The group recommends to the UTC the following disposition ⇒

No UTC action is required.

17) L2/22-287: Request to change glyphs of 63 K-source Hanja chars in UCS [WG2 N5186 & IRG N2585]

Comments: This is a follow up document to L2/22-258 that provides to WG2 more details about the representative glyph changes for 63 K-Source ideographs in the URO (13), Extension A (20), Extension C (3), Extension F (11), and Extension G (16) blocks. This is an FYI document for the UTC, because the UTC already accepted these 63 K-Source representative glyph changes, along

with three additional ones in Extension H, as Consensus 173-C16: Accept the proposal to change the representative glyphs of 66 K-Source ideographs, based on document L2/22-258 and Section 25 of document L2/22-247, for Unicode Version 15.1.

Recommendations: The group recommends to the UTC the following disposition ⇒

No UTC action is required.

Other Topics

18) GB 18030-2022 Amendment (Working Draft)

Henry Chan, who was unable to attend the meeting, kindly informed the group that China issued late last year a Working Draft of a GB 18030-2022 amendment whose highlights can be described as follows:

- Adds the ideographs that have been subsequently appended to the URO (16:9FF0..9FF), Extension A (10:4DB6..4DBF), Extension B (9:2A6D7..2A6DF), and Extension C (5:2B735..2B739), along with Extensions G and H in their entirety. This will synchronize GB 18030-2022 with Unicode Version 15.0, in terms of CJK Unified Ideograph coverage. (For those who are not aware, GB 18030-2022 proper is synchronized with Unicode Version 11.0.) Amending GB 18030 to synchronize with Unicode Version 15.0 is a good thing. At the same time, the Working Draft specifies the same implementation deadline of 2023-08-01 to support Extension G (4,939 ideographs) and Extension H (4,192 ideographs) at Implementation Level 3. This would likely be problematic for developers who are already burdened with supporting CJK Unified Ideographs through Extension F by the 2023-08-01 deadline. A longer lead time for development should be provided.
- Adds the 26 ideographs that have been appended to the URO (16: 9FF0..9FF) and Extension A (10: 4DB6..4DBF) to the scope of Implementation Level 1, meaning that they will be required. Ken Lunde predicted that this would happen, though Ken didn't expect it to happen so quickly. **This is a good thing.**
- Adds 897 ideographs, described as 公安人口信息专用字库补充汉字 (supplementary Chinese characters of the special font for population information used by public security departments), to the beginning of Plane 10 (A0000..A0380) whose code points are currently unallocated. **This is an extraordinarily bad thing.**

As indicated above, the first two highlights are non-problematic and expected, at least for developers that need to support Implementation Levels 1 or 2, but the third is extraordinarily problematic, because hijacking Plane 10 completely circumvents the standardization process, and defeats the purpose of synchronization with the Unicode Standard (aka ISO/IEC 10646 and GB/T 13000). Below is an excerpt from page 50 of the Working Draft.

	30	31	32	33	34	35	36	37	38	39
8F					叆	穽	罧	生	绳	搲
					A0000	A0001	A0002	A0003	A0004	A0005
90	壑	觜	뵇	卿	嫭	禒	崔	圮	彧	跃
	A0006	A0007	A0008	A0009	A000A	A000B	A000C	A000D	A000E	A000F
91	汉	洩	滧	治	漗	濃	灯	炮	勋	尪
	A0010	A0011	A0012	A0013	A0014	A0015	A0016	A0017	A0018	A0019
92	玦	琺	亚	琌	瑄	瓘	盽	砖	硇	磔
	A001A	A001B	A001C	A001D	A001E	A001F	A0020	A0021	A0022	A0023
93	学	宏	盐	学	莣	荻	药	蕉	蕲	射
	A0024	A0025	A0026	A0027	A0028	A0029	A002A	A002B	A002C	A002D
94	那	部	钗	铸	蝾	镗	铠	润	售	霊
	A002E	A002F	A0030	A0031	A0032	A0033	A0034	A0035	A0036	A0037

The group recommends that the UTC prepare a contribution that relays these concerns to China, which may include recommending a later deadline for Implementation Level 3 when considering the addition of Extension G and Extension H within its scope.

Recommendations: The group recommends to the UTC the following disposition ⇒

Action Item for Ken Lunde, UTC: Prepare a contribution on behalf of the Unicode Consortium that relays concerns about encoding 897 ideographs in Plane 10 in the Working Draft of the GB 18030-2022 amendment.

That is all.