

**Title:** CJK & Unihan Group Recommendations for UTC #168 Meeting

**Author:** Ken Lunde

**Date:** 2021-07-22

This document provides the CJK & Unihan Group recommendations for UTC #168 based on a meeting that took place from 6 to 9PM PDT on 2021-07-16, which was attended by Yi Bai, Eiso Chan, Lee Collins, Peter Constable, John Jenkins, Ken Lunde, William Nelson, 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 UTC #167. Comments are marked in **green**, and Copy&Paste-ready Recommendations to the UTC are marked in **red**.

## **1) UAX #38 / Unihan Database Public Feedback—PRI #421**

The six items of public feedback from [PRI #421](#) (*Proposed Update UAX #38*) that are in this section were discussed by the group.

Date/Time: **Tue May 25 20:07:51 CDT 2021**

Name: Eduardo Marín Silva

Report Type: Public Review Issue

Opt Subject: **A few improvements to the field descriptions of UAX#38**

1. Expand the description of kCCCII: Extra information is needed, like the meaning of the initials, the creators and the age. I propose for it to read:

Description | The mapping for this character in hexadecimal, in the "Chinese Character Code for Information Interchange" (CCCII). Created by the "Chinese Character Analysis Group" (CCAG), with its latest version coming out in 1987. Earlier versions of CCCII served as the base for the EACC (see kEACC) so many entries are identical between fields.

The terms between quotes indicate the use of italics. My source is the Wikipedia article

([https://en.wikipedia.org/wiki/Chinese\\_Character\\_Code\\_for\\_Information\\_Interchange](https://en.wikipedia.org/wiki/Chinese_Character_Code_for_Information_Interchange))

which in turn cites the book 'CJKV Information Processing'. I lack access to that book so I ignore the primary source, but if it can be found, it would be important to cite. The code scheme may have its origin in Taiwan (ROC).

Finally, in the field description of kEACC, the complete name (East Asian Character Code for Bibliographic Use) should be spelled with italics, so it is clear that is what the initials stand for.

2. Specify that the romanization used by kJapaneseKun is 'Hepburn'. I ignore if the same applies to kJapaneseOn.

3. Make all appearances of the word 'pinyin' be consistently and correctly spelled as 'pīnyīn' (except, of course, the names of the fields).
4. Include the number of entries for each field: This would add a new row between 'Syntax' and 'Description'. The purpose would be to estimate the size of the field, as well the relative coverage of Ideographs it has.

**Comments:** The group discussed this feedback, and concluded that John Jenkins and the Unicode Editorial Committee should decide which suggested editorial changes, if any, to accept for the text of UAX #38. In particular, the fourth suggested change represents a maintenance nightmare of extreme proportions, because it would require adjustments to a large number of UniHan database property tables in UAX #38 for the current and each subsequent version of the Unicode Standard. The feedback submitter can instead reference the three sheets of the [UniHan Property History](#) spreadsheet to get the desired information for each version of the UniHan database.

**Recommendations (UniHan-UTC168-R01):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Editorial Committee:** Update the text of UAX #38 as appropriate per feedback dated [Tue May 25 20:07:51 CDT 2021] in PRI #421 and UniHan-UTC168-R01 in document L2/21-129, for Unicode Version 14.0. **Action Item for Rick McGowan:** Convey to the feedback submitter the CJK & UniHan Group comments in document L2/21-129. [Tue May 25 20:07:51 CDT 2021]

Date/Time: **Tue May 25 21:13:11 CDT 2021**

Name: Eduardo Marín Silva

Report Type: Public Review Issue

Opt Subject: **Lack of documentation relating to other legacy East Asian encodings**

Using this article as my source

([https://en.wikipedia.org/wiki/Extended\\_Unix\\_Code](https://en.wikipedia.org/wiki/Extended_Unix_Code))

I list some encoding schemes and character sets, not mentioned in UAX#38. This is merely a fyi type of observation, so it should not affect the text of the annex for now.

Furthermore, I'm not sure if all elements of the list necessarily contain ideographs or can be considered identical to other fields. In the first case they can be dismissed, but in the latter, the equivalence must be clearly documented. If there is another document that clarifies the relation between different standards, it should be cited in the doc.

The list in question is:

EUC-JP (Extended Unix Japan), Shift JIS, DEC Kanji (by Digital Equipment Corporation), HP-15, HP-16, IKIS (by Data General), MacJapanese (MacOS), Windows-932 (IBM-943), KEIS (by Hitachi), EUC-KR (Extended Unix Korean/Wansung), UHC (Unified Hangul Code/Windows949/Extended Wansung), HangulTalk (MacOS) and EUC-TW (Extended Unix Taiwan)

Big5 is only mentioned but not properly explained, and the GB fields are not correctly attributed to Gubiao Standards.

**Comments:** The group discussed this feedback, and immediately concluded that this level of information does not belong in UAX #38, neither now nor in the future.

**Recommendations (Unihan-UTC168-R02):** The group recommends to the UTC the following disposition ➡ **Action Item for Rick McGowan:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Tue May 25 21:13:11 CDT 2021]

Date/Time: **Fri May 28 09:29:04 CDT 2021**

Name: Michel Mariani

Report Type: Error Report

Opt Subject: **Corrections for kTotalStrokes**

After reporting issues on the Unihan mailing list, I am submitting the following corrections for the kTotalStrokes property:

U+28668 𪛗 kTotalStrokes 7

U+2F9DD 𪛗 kTotalStrokes 9

- > It does not surprise me that there are some puzzlers lurking in the
- > kTotalStrokes property. The correction for U+28E0F 𪛗 was submitted
- > as public feedback by Jaemin Chung on 2020-09-03, and the correction
- > was applied to the Unihan database.
- > If you don't mind, please submit the following corrections via the
- > Contact Form so that we have a paper trail:
- > [...]
- > We should be able to get those corrections applied in time for
- > Unicode Version 14.0.

**Comments:** The group discussed this feedback, and agreed that the *kTotalStrokes* property values for these two ideographs, which are shown below, should be changed to the proposed values.

Current:

U+28668 kTotalStrokes 2

U+2F9DD kTotalStrokes 2

**Recommendations (Unihan-UTC168-R03):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Michel Suignard:** Update the *kTotalStrokes* property values in the Unihan database per feedback dated [Fri May 28 09:29:04 CDT 2021] in PRI #421 and Unihan-UTC168-R03 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Fri May 28 09:29:04 CDT 2021]

Change:

U+28668 kTotalStrokes 7

U+2F9DD kTotalStrokes 9

Date/Time: **Fri Jun 11 23:46:52 CDT 2021**

Name: Eduardo Marín Silva

Report Type: Public Review Issue

Opt Subject: **Anomalies in the spelling and format of Unihan field descriptions**

I have noticed some anomalies in the spelling or format in different field descriptions, particularly "kIRG\_GSource".

kIRG\_GSource:

Misspelled Book Name | Corrected book name

-----

ZhongHua ZiHai | Zhonghua Zihai

Chinese Encyclopedia | Encyclopedia of China (name also needs to be italicized)

Ci Hai | Cihai (if the "Ci Hai" spelling is preferred, then it should be used consistently everywhere)

Ci Yuan | Ciyuan

Hanyu Dacidian | Hanyu Da Cidian

Hanyu Dazidian | Hanyu Da Zidian

Hanyu Fangyan Dacidian | Hanyu Fangyan Da Cidian

Chinese Ancient Ethnic Research on Ancient (unless it doesn't refer to the book of the same name, it should be italicized)

Characters Research | Chinese Characters

Chinese book titles without pinyin or translations (if added, they should be italicized):

Chinese title | pinyin [translation]

-----

汉语大字典（第二版） | Hanyu Da Zidian (second edition)

漢文佛典疑難俗字彙釋與研究 | Hànwén Fódǎn Yínán Sù Zìhuì Shì Yǔ Yánjiū

[Explanation and Research on Difficult and Vulgar Words in Chinese Buddhist Classics]

龍龕手鑑 | Longkan Shoujian [The Handy Mirror in the Dragon Shrine or The Dragon Shrine/Niche Handbook]

Names of books not italicized:

Siku Quanshu, Yinzhou Jinwen Jicheng Yinde, Standard Telegraph Codebook (revised)

((the last one should also precede the Chinese name for consistency))

Also, the name of the publisher "Zhuang Liao Songs Research" appears without spaces between words

kHDZRadBreak:

In the sentence "Indicates that 《漢語大字典》 Hanyu Da Zidian has a radical break beginning at this character's position." place the pinyin first, and

italicize it to be consistent. Similar suggestions apply to the descriptions of kHanyuPinlu and kHanyuPinyin

kIRG\_KPSource:

Reword the sentence: "... There may, therefore, be erroneous data in the values for this field." to say: "... Therefore, there may be erroneous data in the values for this field."

kIRG\_SSource:

Italicize or place quotes on the name "Taishō Shinshū Daizōkyō"

kIRG\_VSource:

Italicize or place quotes on the name "Kho Chũ Hán Nôm Mã Hoá"

**Comments:** The group discussed this feedback, and concluded that John Jenkins and the Unicode Editorial Committee should decide which suggested editorial changes, if any, to accept for the text of UAX #38. The group does recommend that the pinyin and English translation of the following book be changed as shown below:

漢文佛典疑難俗字彙釋與研究 | Hànwén Fódǎn Yínánsúzì Huìshì Yǔ Yánjiū [Explanation and Research on Difficult Vulgar Characters in Chinese Buddhist Classics]

**Recommendations (Unihan-UTC168-R04):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Editorial Committee:** Update the text of UAX #38 as appropriate per feedback dated [Fri Jun 11 23:46:52 CDT 2021] in PRI #421 as amended by the CJK & Unihan Group, and Unihan-UTC168-R04 in document L2/21-129, for Unicode Version 14.0. **Action Item for Rick McGowan:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Fri Jun 11 23:46:52 CDT 2021]

Date/Time: **Thu Jul 8 11:13:37 CDT 2021**

Name: Ken Lunde

Report Type: Public Review Issue

Opt Subject: **PRI #421 Feedback**

There are a small number of anomalies in earlier versions of the Unihan database, and it may be useful to document them in UAX #38, mainly in Section 5, "History":

1) The Version 2.0.0 Unihan database file, Unihan-1.txt, is truncated in the middle of the records for U+8BC1 証:

<https://www.unicode.org/Public/2.0-Update/Unihan-1.txt>

While this is already documented in Section 5 of UAX #38, it may be helpful to add that the CD that is included with the Unicode Version 2.0 book has the same issue, specifically that the files at {DOS,MAC,UNIX}/MAPPINGS/EASTASIA/UNIHAN.TXT are truncated at the same position. This would preclude those who have the Unicode Version 2.0 book from checking the CD on their own (like I did;-).

2) The Version 3.0.0 Unihan database file, Unihan-3.txt, includes 3,898 records for the undocumented *kJHJ* property:

<https://www.unicode.org/Public/3.0-Update/Unihan-3.txt>

I suggest that appropriate entries be added to the table in Section 4.2 of UAX #38, specifically the following:

Version 3.0 row: Add "*kJHJ*" to the "Fields Added" column

Version 3.1 row: Add "*kJHJ*" to the "Fields Dropped" column

It may also be useful to document this property in Section 5 for the benefit of those who parse older versions of the Unihan database.

3) The Version 3.1.1 Unihan database file, Unihan-3.1.1.txt, includes the following anomalous record at line 246,442:

U+64AC 297

See:

<https://www.unicode.org/Public/3.1-Update1/Unihan-3.1.1.txt>

It may be useful to document this in Section 5 for the benefit of those who parse older versions of the Unihan database.

4) The Versions 2.0.0, 2.1.2, 3.0.0, and 3.1.0 Unihan database files are not encoded in UTF-8:

<https://www.unicode.org/Public/2.0-Update/Unihan-1.txt>

<https://www.unicode.org/Public/2.1-Update/Unihan-2.txt>

<https://www.unicode.org/Public/3.0-Update/Unihan-3.txt>

<https://www.unicode.org/Public/3.1-Update/Unihan-3.1.txt>

It may be useful to document this in Section 5 for the benefit of those who parse older versions of the Unihan database.

That is all.

**Comments:** The group discussed this feedback, and agreed that the table in Section 4.2 of UAX #38 should be updated to reflect the *kJHJ* property, and that Section 5 of UAX #38 should be updated to document the other anomalies. The group also noted that while the *kAlternateJEF* and *kRSMerged* properties are covered in the table in Section 4.2, specifically that they were added in Version 3.0 then removed in Version 3.1, the header of the Unihan database file for Unicode Version 3.0, [Unihan-3.txt](#), provides no description for these two properties. The following is a suggested description for the *kAlternateJEF* property:

The code point in the 拡張漢字 (aka extended kanji) portion of the FACOM JEF (*Japanese processing Extended Feature*) character set expressed in hexadecimal

**Recommendations (Unihan-UTC168-R05):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Editorial Committee:** Update the text of UAX #38 to reflect the history of the *kJHJ* property in the table in Section 4.2, and to reflect the other anomalies in Section 5 per feedback dated [Thu Jul 8 11:13:37 CDT 2021] in PRI #421 as amended by the CJK & Unihan Group, and Unihan-UTC168-R05 in document L2/21-129, for Unicode Version 14.0.

Date/Time: **Thu Jul 15 13:24:42 CDT 2021**  
Name: Ben Scarborough  
Report Type: Public Review Issue  
Opt Subject: **Proposed change for UAX #38**

In the current proposed update for UAX #38, the syntax for the `kIRG_VSource` property is:

```
V[0-4N]-[023F]?[0-9A-F]{4}
```

To keep it in line with how regexes are laid out for other IRG source properties,

a more accurate regex would be:

```
V[0-4]-[0-9A-F]{4}
```

```
| VN-[023F][0-9A-F]{4}
```

because the `V[0-4]` sources are always 4 hex digits and the `VN` sources are always 5.

**Comments:** The group discussed this feedback, and agreed that the proposed change to the syntax of the `kIRG_VSource` property is appropriate.

**Recommendations (Unihan-UTC168-R06):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Editorial Committee:** Update the syntax for the `kIRG_VSource` property in UAX #38 per feedback dated [Thu Jul 15 13:24:42 CDT 2021] in PRI #421 and Unihan-UTC168-R06 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Thu Jul 15 13:24:42 CDT 2021]

## 2) UAX #42 Public Feedback—PRI #431

The two related items of public feedback from [PRI #431](#) (*Proposed Update UAX #42*) that are in this section were discussed by the group.

Date/Time: **Sat Jul 10 13:18:17 CDT 2021**  
Name: Ken Lunde  
Report Type: Public Review Issue  
Opt Subject: **PRI #431 (UAX #42) feedback**

I checked all known versions of the Unihan database, and I cannot find the `kWubi` property that is documented in Section 4.4.23, Unihan properties, of UAX #42 as follows:

```
code-point-attributes &= attribute kWubi  
    { text }?
```

This property is also not mentioned in UAX #38.

Unless someone can find a version of the Unihan database that includes this particular property, I recommend that its entry be dropped from UAX #42.

Date/Time: **Sun Jul 11 09:00:18 CDT 2021**

Name: Ken Lunde

Report Type: Public Review Issue

Opt Subject: **Additional information for PRI #431 (UAX #42) feedback**

The following information is relevant to my recommendation that the kWubi property be dropped from Section 4.4.23 of UAX #42:

1) The proposal that included the kWubi property was L2/06-208:

<https://www.unicode.org/L2/L2006/06208-unihan-new-fields.pdf>

2) This proposal was discussed during UTC #107 and reflected in the minutes:

<https://www.unicode.org/consortium/utc-minutes/UTC-107-200605.html>

The following are the relevant Consensus and Action Items:

>From 2006-05-18:

Properties - New Unihan Fields (B.14.6)

[107-C25] Consensus: Add kFourCornerCode and kWubi to Unihan.txt as provisional properties for Unicode 5.0, subject to clear title to the data. [L2/06-206]

[107-A57] Action Item for Deborah Goldsmith: Investigate the title and licensing questions about kWubi and kFourCornerCode data for Unihan.txt, and report to the Unicode officers by May 25th, 2006.

Note: see also 107-A68, 107-A69, and 107-A70 regarding kWubi data. Action 107-A57 was done, and reported back later in the meeting resulting in removal of kWubi data from the release.

>From 2006-05-19:

Properties - Unihan (B.14.6)

[107-A68] Action Item for John Jenkins: Remove kWubi from the Unihan database beta and generate a revision of Unihan.txt for the Unicode 5.0. [L2/06-208]

[107-A69] Action Item for John Jenkins, Richard Cook: Remove references to kWubi from Unicode Technical Report #38: A User's Guide to the Unihan Database.

[107-A70] Action Item for Ken Whistler: Make relevant updates to the Unicode 5.0 text to remove references to kWubi data.

In other words, the kWubi property was never added to a released version of the Unihan database, and should therefore be removed from Section 4.4.23 of UAX #42.

**Comments:** The editor of UAX #42 responded to this feedback by confirming that no released version of the Unihan database included the *kWubi* property, and agreed that it is appropriate to remove its entry from Section 4.4.23.

**Recommendations (Unihan-UTC168-R07):** The group recommends to the UTC the following disposition ➡ **Action Item for Eric Muller, Editorial Committee:** Adjust Section 4.4.23 of UAX #42 by removing the entry for the *kWubi* property per PRI #431 feedback dated [Sat Jul 10 13:18:17 CDT 2021] and [Sun Jul 11 09:00:18 CDT 2021], and Unihan-UTC168-R07 in document L2/21-129, for Unicode Version 14.0.

### 3) Unicode Version 14.0.0 Beta Public Feedback—PRI #433

The 12 items of public feedback from [PRI #433](#) (*Unicode Version 14.0.0 Beta*) that are in this section were discussed by the group, and for convenience, the *Unihan-additions-20210716.txt* and *Unihan-changes-20210716.txt* (PDF attachments) files include all of the recommended Unihan database additions and changes based on these items of public feedback.

Date/Time: **Tue Jun 8 11:08:50 CDT 2021**

Name: Mike FABIAN

Report Type: Error Report

Opt Subject: **Bugs in Unihan\_Variants.txt (Unicode 14 draft version)**

I think Unihan\_Variants.txt from

<https://www.unicode.org/Public/14.0.0/ucd/Unihan-14.0.0d4.zip>

still has bugs for the following characters:

乾 U+4E7E, 著 U+8457, 覆 U+8986

are used not only use traditional Chinese but also in simplified Chinese,

杰 U+6770, 系 U+7CFB, 只 U+53EA

are used not only in simplified Chinese but also in traditional Chinese.

The Chinese input method ibus-table (which I maintain) has options to show only simplified, only traditional, simplified first, traditional first, or both in no particular order.

To figure out which characters are simplified, traditional, or both, ibus-table uses Unihan\_Variants.txt.

Therefore, I still apply these two patches to Unihan\_Variants.txt:

<https://github.com/mike-fabian/ibus-table/commit/c37452a7bf49ccfd-0f2062a3e90085022cb3735b>

<https://github.com/mike-fabian/ibus-table/commit/6384e217f7283d12b1bdabbc10da4d2cf2e4f94a>

```
$ git show c37452a7bf49ccfd0f2062a3e90085022cb3735b
```

```
commit c37452a7bf49ccfd0f2062a3e90085022cb3735b
```

```
Author: Mike FABIAN <mfabian@redhat.com>
```

```
Date: Tue Jun 8 13:58:22 2021 +0200
```

```
Keep our fixes to Unihan_Variants which are not yet included upstream
```

```
diff --git a/tools/Unihan_Variants.txt b/tools/Unihan_Variants.txt
```

```
index 72be50f..a29e627 100644
```

```
--- a/tools/Unihan_Variants.txt
```

```

+++ b/tools/Unihan_Variants.txt
@@ -1015,7 +1015,7 @@ U+4E66      kTraditionalVariant U+66F8
    U+4E70 kTraditionalVariant U+8CB7
    U+4E71 kSemanticVariant      U+4E82<kMatthews,kMeyerWempe
    U+4E71 kTraditionalVariant U+4E82
-U+4E7E kSimplifiedVariant      U+5E72
+U+4E7E kSimplifiedVariant      U+4E7E U+5E72
    U+4E7E kSpecializedSemanticVariant U+4E81<kFenn
    U+4E81 kSpecializedSemanticVariant U+4E7E<kFenn
    U+4E82 kSemanticVariant      U+4E71<kMatthews,kMeyerWempe
@@ -3814,7 +3814,7 @@ U+6769      kTraditionalVariant U+69AA
    U+676E kSpoofingVariant      U+67FF
    U+676F kSemanticVariant      U+76C3<kLau,kMatthews,kMeyerWempe
    U+6770 kSemanticVariant      U+5091<kMatthews
-U+6770 kTraditionalVariant U+5091
+U+6770 kTraditionalVariant U+6770 U+5091
    U+6771 kSemanticVariant      U+4E1C<kFenn
    U+6771 kSimplifiedVariant      U+4E1C
    U+6777 kSemanticVariant      U+6733<kMatthews
@@ -5912,7 +5912,7 @@ U+7CF9      kSemanticVariant      U+7CF8<kMatthews
    U+7CF9 kSimplifiedVariant      U+7E9F
    U+7CFA kSemanticVariant      U+7CFE<kMatthews,kMeyerWempe
    U+7CFA kSimplifiedVariant      U+2B119
-U+7CFB kTraditionalVariant U+4FC2 U+7E6B
+U+7CFB kTraditionalVariant U+7CFB U+4FC2 U+7E6B
    U+7CFD kSimplifiedVariant      U+30AFC
    U+7CFE kSemanticVariant      U+7CFA<kMatthews,kMeyerWempe
    U+7CFE kSimplifiedVariant      U+7EA0
@@ -6938,7 +6938,7 @@ U+8441      kSemanticVariant      U+8591<kMatthews
    U+8449 kSimplifiedVariant      U+53F6
    U+8452 kSimplifiedVariant      U+836D
    U+8457 kSemanticVariant      U+7740
-U+8457 kSimplifiedVariant      U+7740
+U+8457 kSimplifiedVariant      U+8457 U+7740
    U+8457 kSpecializedSemanticVariant U+7740<kFenn U+87AB<kFenn
    U+845A kSemanticVariant      U+6939<kFenn
    U+845D kSimplifiedVariant      U+2B20E
@@ -7475,7 +7475,7 @@ U+8975      kSimplifiedVariant      U+2B307
    U+8978 kSimplifiedVariant      U+2C877

```

```

U+8979 kSimplifiedVariant U+30CFC
U+897C kSimplifiedVariant U+30CF5
-U+8986 kSimplifiedVariant U+590D
+U+8986 kSimplifiedVariant U+590D U+8986
U+8987 kSemanticVariant U+9738<kMeyerWempe
U+898A kSemanticVariant U+7F88<kMatthews
U+898B kSimplifiedVariant U+89C1
$
$ git show 6384e217f7283d12b1bdabbc10da4d2cf2e4f94a
commit 6384e217f7283d12b1bdabbc10da4d2cf2e4f94a
Author: Mike FABIAN <mfabian@redhat.com>
Date: Tue Jun 8 13:59:59 2021 +0200
Fix bug in Unihan_Variants.txt, 只 U+53EA is both simplified *and* traditional
Chinese
Resolves: https://github.com/kaio/ibus-table/issues/74
diff --git a/engine/chinese_variants.py b/engine/chinese_variants.py
index e53e6d6..0b79eea 100644
--- a/engine/chinese_variants.py
+++ b/engine/chinese_variants.py
@@ -1116,7 +1116,7 @@ VARIANTS_TABLE = {
     u'叙': 1,
     u'叠': 1,
     u'叢': 2,
-    u'只': 1,
+    u'只': 3,
     u'台': 3,
     u'叶': 1,
     u'号': 1,
diff --git a/tools/Unihan_Variants.txt b/tools/Unihan_Variants.txt
index a29e627..ab84494 100644
--- a/tools/Unihan_Variants.txt
+++ b/tools/Unihan_Variants.txt
@@ -1695,7 +1695,7 @@ U+53E2 kSemanticVariant U+6A37<kFenn
U+53E2 kSimplifiedVariant U+4E1B
U+53E8 kSemanticVariant U+9955<kMeyerWempe
U+53EA kSemanticVariant U+5B50<kLau
-U+53EA kTraditionalVariant U+96BB
+U+53EA kTraditionalVariant U+53EA U+96BB
U+53EB kSemanticVariant U+544C<kLau,kMatthews,kMeyerWempe

```

```

U+53F0 kSemanticVariant      U+81FA<kHKGLyph,kLau
U+53F0 kSimplifiedVariant    U+53F0
diff --git a/tools/generate-chinese-variants.py b/tools/generate-chinese-variants.py
index 391b5ef..310d6ae 100755
--- a/tools/generate-chinese-variants.py
+++ b/tools/generate-chinese-variants.py
@@ -273,6 +273,7 @@ TEST_DATA = {
    u'系': 3, # U+7CFB
    u'乾': 3, # U+4E7E
    u'著': 3, # U+8457 Patch by Heiher <r@hev.cc>
+   u'只': 3, # U+53EA, see: https://github.com/kaio/ibus-table/issues/74
    }
def test_detection(generated_script) -> int:
$

```

**Comments:** The group reviewed this feedback, and agreed that the suggested *kSimplifiedVariant* and *kTraditionalVariant* property value changes are appropriate, though multiple property values should be in increasing code point order. The current property values are shown below.

Current:

U+4E7E	kSimplifiedVariant	U+5E72
U+53EA	kTraditionalVariant	U+96BB
U+6770	kTraditionalVariant	U+5091
U+7CFB	kTraditionalVariant	U+4FC2 U+7E6B
U+8457	kSimplifiedVariant	U+7740
U+8986	kSimplifiedVariant	U+590D

**Recommendations (Unihan-UTC168-R08):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins:** Change the *kSimplifiedVariant* and *kTraditionalVariant* property values in the Unihan database per feedback dated [Tue Jun 8 11:08:50 CDT 2021] in PRI #433 as amended by the CJK & Unihan Group, and Unihan-UTC168-R08 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Tue Jun 8 11:08:50 CDT 2021]

Change:

U+4E7E	kSimplifiedVariant	U+4E7E U+5E72
U+53EA	kTraditionalVariant	U+53EA U+96BB
U+6770	kTraditionalVariant	U+5091 U+6770
U+7CFB	kTraditionalVariant	U+4FC2 U+7CFB U+7E6B
U+8457	kSimplifiedVariant	U+7740 U+8457
U+8986	kSimplifiedVariant	U+590D U+8986

Date/Time: **Tue Jun 8 23:34:03 CDT 2021**

Name: Yi Bai

Report Type: Error Report

Opt Subject: **Glyph missing in code chart of CJK Unified Ideographs Extension A**

**Note: This glyph error will be corrected by the chart editors and will be posted.**

In code chart of CJK Unified Ideographs Extension A in Version 14.0 Beta, glyph of 4DB9 with UTC source UTC-00120 is missing. The glyph can be found in current 13.0 code chart.

Please update the code chart accordingly, thank you.

**Comments:** This feedback has already been addressed, and the feedback submitter attended the CJK & Unihan Group meeting and is therefore aware of its disposition.

Date/Time: **Thu Jun 10 09:18:03 CDT 2021**

Name: Andrew Christopher West

Report Type: Public Review Issue

Opt Subject: **PRI #433 Unicode 14.0.0 Beta**

U+2B8D9 and U+2B8DA have the wrong radical and stroke count since the glyph changes for Unicode 13.0 simplified 來 to 来 in both cases (in retrospect this was a destabilizing glyph change, and encoding two new characters would have been a better solution). The code charts for Unicode 14.0 still show the old radical stroke count of 9.12, but as the new glyph forms no longer include the 'person' radical (9) they must be assigned to a new radical. I suggest 'wood' (75) as this is the radical for 来, which would give 75.9 for both characters.

**Comments:** The group discussed this feedback, and while the group agreed that these are appropriate changes, the group also noted that the *kTotalStrokes* property values also need adjustment, and that 30.10 should be added as a secondary *kRSUnicode* property value for both ideographs. The current property values are shown below.

Current:

U+2B8D9	kRSUnicode	9.12
U+2B8D9	kTotalStrokes	14
U+2B8DA	kRSUnicode	9.12
U+2B8DA	kTotalStrokes	14

**Recommendations (Unihan-UTC168-R09):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Michel Suignard:** Change the *kRSUnicode* and *kTotalStrokes* property values in the Unihan database per feedback dated [Thu Jun 10 09:18:03 CDT 2021] in PRI #433 as amended by the CJK & Unihan Group, and Unihan-UTC168-R09 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Thu Jun 10 09:18:03 CDT 2021]

Change:

U+2B8D9	kRSUnicode	75.9 30.10
U+2B8D9	kTotalStrokes	13
U+2B8DA	kRSUnicode	75.9 30.10

Date/Time: **Thu Jun 10 09:48:33 CDT 2021**  
 Name: Andrew Christopher West  
 Report Type: Public Review Issue  
 Opt Subject: **PRI #433 Unicode 14.0.0 Beta**

Following glyph changes in Unicode 11.0, the following characters have the wrong radical and stroke counts:

U+2B1CD: 132.20 -- should be 132.19

U+2B584: 176.12 -- should be 176.11

U+2B8DE: 9.12 -- should be 9.11

U+2C7C3: 140.15 -- should be 140.14

Following glyph changes in Unicode 13.0, the following characters have the wrong radical and stroke counts:

U+2BD61: 44.10 -- should be 44.9

U+2BE4A: 59.14 -- should be 59.15

U+2BF9D: 64.18 -- should be 64.19

U+2C0B8: 75.7 -- should be 75.8

U+2C142: 75.15 -- should be 75.16

U+2C316: 91.17 -- should be 91.19

U+2C83A: 142.17 -- should be 142.18

U+2CC88: 182'.13 -- should be 182'.16

**Comments:** The group discussed this feedback, and while the group agreed that these are appropriate changes, the group also noted that most of the *kTotalStrokes* property values also need adjustment. The current property values are shown below.

#### Current:

U+2B1CD	kRSUnicode	132.20
U+2B1CD	kTotalStrokes	26
U+2B584	kRSUnicode	176.12
U+2B584	kTotalStrokes	21
U+2B8DE	kRSUnicode	9.12
U+2B8DE	kTotalStrokes	14
U+2C7C3	kRSUnicode	140.15
U+2C7C3	kTotalStrokes	17

U+2BD61	kRSUnicode	44.10
U+2BE4A	kRSUnicode	59.14
U+2BE4A	kTotalStrokes	17
U+2BF9D	kRSUnicode	64.18
U+2BF9D	kTotalStrokes	22
U+2C0B8	kRSUnicode	75.7
U+2C0B8	kTotalStrokes	11
U+2C142	kRSUnicode	75.15
U+2C142	kTotalStrokes	19
U+2C316	kRSUnicode	91.17

U+2C316	kTotalStrokes	20
U+2C83A	kRSUnicode	142.17
U+2C83A	kTotalStrokes	23
U+2CC88	kRSUnicode	182'.13
U+2CC88	kTotalStrokes	17

**Recommendations (Unihan-UTC168-R10):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Michel Suignard:** Change the *kRSUnicode* and *kTotalStrokes* property values in the Unihan database per feedback dated [Thu Jun 10 09:48:33 CDT 2021] in PRI #433 as amended by the CJK & Unihan Group, and Unihan-UTC168-R10 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Thu Jun 10 09:48:33 CDT 2021]

#### Change:

U+2B1CD	kRSUnicode	132.19
U+2B1CD	kTotalStrokes	25
U+2B584	kRSUnicode	176.11
U+2B584	kTotalStrokes	20
U+2B8DE	kRSUnicode	9.11
U+2B8DE	kTotalStrokes	13
U+2C7C3	kRSUnicode	140.14
U+2C7C3	kTotalStrokes	18
U+2BD61	kRSUnicode	44.9
U+2BE4A	kRSUnicode	59.15
U+2BE4A	kTotalStrokes	18
U+2BF9D	kRSUnicode	64.19
U+2BF9D	kTotalStrokes	23
U+2C0B8	kRSUnicode	75.8
U+2C0B8	kTotalStrokes	12
U+2C142	kRSUnicode	75.16
U+2C142	kTotalStrokes	20
U+2C316	kRSUnicode	91.19
U+2C316	kTotalStrokes	23
U+2C83A	kRSUnicode	142.18
U+2C83A	kTotalStrokes	24
U+2CC88	kRSUnicode	182'.16
U+2CC88	kTotalStrokes	20

Date/Time: **Mon Jun 14 09:03:19 CDT 2021**

Name: Mike FABIAN

Report Type: Error Report

Opt Subject: **Some more possible bugs in Unihan\_Variants.txt**

Recently I reported some problems in Unihan\_Variants.txt from

<https://www.unicode.org/Public/14.0.0/ucd/Unihan-14.0.0d4.zip>

I received this bug report about ibus-table classifying some characters wrongly as simplified only or traditional only:

<https://github.com/ibus/ibus/issues/2323>

According to this bug report there maybe a few more bugs concerning *kTraditionalVariant* and *kSimplifiedVariant*.

In the above bug report, the user said that

着 U+7740 is used in Hong Kong, 云 U+4E91 is used both in Hong Kong and Taiwan, 裡 U+88E1 and 復 U+5FA9 are used everywhere (i.e. also in simplified Chinese), 采 U+91C7 is used in Hong Kong (The user wrote he was nor sure about Taiwan, but probably it iis used in Taiwan as well as it is listed in

<http://dict.revised.moe.edu.tw/cgi-bin/cbdic/gswweb.cgi>),

吓 U+5413 is used in Cantonese, 搵 U+63FE is used in Hong Kong.

The user wrote he doesn't know where 尸 is used but it is one of the radicals used on a Cangjie keyboard, so it seems to be used in traditonal Chinese, at least as a radical.

I fixed it like this:

<https://github.com/mike-fabian/ibus-table/commit/5ed1cc16b398e0161e63ef35421d94166caf56c0>

I.e. I applied the following changes to *Unihan\_Variants.txt*:

-U+4E91 <i>kTraditionalVariant</i>	U+96F2
+U+4E91 <i>kTraditionalVariant</i>	U+4E91 U+96F2
-U+5413 <i>kTraditionalVariant</i>	U+5687
+U+5413 <i>kTraditionalVariant</i>	U+5413 U+5687
-U+5C38 <i>kTraditionalVariant</i>	U+5C4D
+U+5C38 <i>kTraditionalVariant</i>	U+5C38 U+5C4D
-U+5FA9 <i>kSimplifiedVariant</i>	U+590D
+U+5FA9 <i>kSimplifiedVariant</i>	U+590D U+5FA9
-U+63FE <i>kTraditionalVariant</i>	U+6435
+U+63FE <i>kTraditionalVariant</i>	U+63FE U+6435
-U+7740 <i>kTraditionalVariant</i>	U+8457
+U+7740 <i>kTraditionalVariant</i>	U+7740 U+8457
-U+88E1 <i>kSimplifiedVariant</i>	U+91CC
+U+88E1 <i>kSimplifiedVariant</i>	U+88E1 U+91CC
-U+91C7 <i>kTraditionalVariant</i>	U+57F0 U+63A1
+U+91C7 <i>kTraditionalVariant</i>	U+57F0 U+63A1 U+91C7

**Comments:** The group reviewed this feedback, and agreed that the suggested *kSimplifiedVariant* and *kTraditionalVariant* property value changes are appropriate. The current property values are shown below.

Current:

U+4E91	<i>kTraditionalVariant</i>	U+96F2
U+5413	<i>kTraditionalVariant</i>	U+5687
U+5C38	<i>kTraditionalVariant</i>	U+5C4D
U+5FA9	<i>kSimplifiedVariant</i>	U+590D
U+63FE	<i>kTraditionalVariant</i>	U+6435
U+7740	<i>kTraditionalVariant</i>	U+8457

U+88E1	kSimplifiedVariant	U+91CC
U+91C7	kTraditionalVariant	U+57F0 U+63A1

**Recommendations (Unihan-UTC168-R11):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins:** Change the *kSimplifiedVariant* and *kTraditionalVariant* property values in the Unihan database per feedback dated [Mon Jun 14 09:03:19 CDT 2021] in PRI #433 and Unihan-UTC168-R11 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Mon Jun 14 09:03:19 CDT 2021]

Change:

U+4E91	kTraditionalVariant	U+4E91	U+96F2
U+5413	kTraditionalVariant	U+5413	U+5687
U+5C38	kTraditionalVariant	U+5C38	U+5C4D
U+5FA9	kSimplifiedVariant	U+590D	U+5FA9
U+63FE	kTraditionalVariant	U+63FE	U+6435
U+7740	kTraditionalVariant	U+7740	U+8457
U+88E1	kSimplifiedVariant	U+88E1	U+91CC
U+91C7	kTraditionalVariant	U+57F0	U+63A1 U+91C7

Date/Time: **Tue Jun 15 06:29:07 CDT 2021**

Name: M

Report Type: Other Question, Problem, or Feedback

Opt Subject: **FEEDBACK ABOUT UNIHAN DATABASE**

1. feedback:

[𪛗][U+4E7E] is a variant of [𪛗][U+4E7E], [𪛗][U+4E81] according to the variant list

(第一批异体字整理表) <https://upload.wikimedia.org/wikipedia/commons/2/29/%E7%AC%A%E4%B8%80%E6%89%B9%E5%BC%82%E4%BD%93%E5%AD%97%E6%95%B4%E7%90%86%E8%A1%A8.pdf>

page 4

2. question:

[卿][U+537F]

kRSUnicode 26.9

kTotalStrokes 10

This doesn't make sense.

SHOULD BE

kRSUnicode 26.8

kTotalStrokes 10

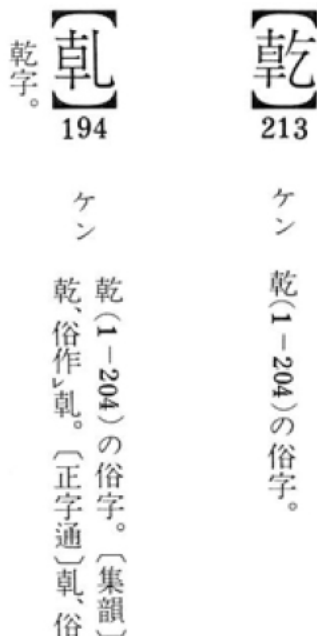
OR

kRSUnicode 26.9

kTotalStrokes 11

**Comments:** The group discussed this feedback. For #1, the first code point in the feedback, U+4E7E, is incorrect, and should have been U+4E79. The ideograph shown is that of U+4E79. While the group agreed that adding appropriate *kSemanticVariant* property values for these

three ideographs is appropriate, there were concerns about being able to cite a source, preferably one that is associated with a Unihan database property. Volume One of 大漢和辞典 (aka the *kMorohashi* property) includes entries for U+4E79 𤎆 (194) and U+4E81 𤎇 (213) that indicate that they are vulgar variants (俗字) of U+4E7E 乾, which is considered the standard form, and whose entry, 204, spans pp 396 through 403 of Volume One. Excerpts from pp 396 and 403 of Volume One of 大漢和辞典 are shown below:



For #2, the group agreed that the *kRSUnicode* and *kTotalStrokes* property values for U+537F 卿 do not make sense, and concluded that this ideograph should have two *kRSUnicode* and two *kTotalStrokes* property values. The Unicode Version 13.0 code chart excerpt below makes this need clear:



Current:

U+537F	<i>kRSUnicode</i>	26.9
U+537F	<i>kTotalStrokes</i>	10

**Recommendations (Unihan-UTC168-R12):** The group recommends to the UTC the following disposition ⇒ **Action Item for John Jenkins:** Add the *kSemanticVariant* property values per feedback dated [Tue Jun 15 06:29:07 CDT 2021] in PRI #433 as amended by the CJK & Unihan Group, and Unihan-UTC168-R12 in document L2/21-129, for Unicode Version 14.0. **Action Item for John Jenkins, Michel Suignard:** Change the *kRSUnicode* and *kTotalStrokes* property values for U+537F 卿 in the Unihan database per feedback dated [Tue Jun 15 06:29:07 CDT 2021] in PRI #433 as amended by the CJK & Unihan Group, and Unihan-UTC168-R12 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Tue Jun 15 06:29:07 CDT 2021]

Add:

U+4E79	<i>kSemanticVariant</i>	U+4E7E< <i>kMorohashi</i> :Z
--------	-------------------------	------------------------------

U+4E7E	kSemanticVariant	U+4E79<kMorohashi	U+4E81<kMorohashi
U+4E81	kSemanticVariant	U+4E7E<kMorohashi	Z

## Change:

U+537F	kRSUnicode	26.8	26.10
U+537F	kTotalStrokes	10	12

Date/Time: **Thu Jun 17 07:59:42 CDT 2021**

Name: Ken Lunde

Report Type: Error Report

Opt Subject: **21 CJK Unified Ideographs are missing kTotalStrokes property values**

The following 21 CJK Unified Ideographs, in the range U+9FD6 through U+9FEA that were added in Unicode Version 10.0, are missing kTotalStrokes property values, and the suggested property values are provided below:

U+9FD6	kTotalStrokes	7
U+9FD7	kTotalStrokes	9
U+9FD8	kTotalStrokes	12
U+9FD9	kTotalStrokes	13
U+9FDA	kTotalStrokes	16
U+9FDB	kTotalStrokes	24
U+9FDC	kTotalStrokes	22
U+9FDD	kTotalStrokes	27
U+9FDE	kTotalStrokes	20
U+9FDF	kTotalStrokes	12
U+9FE0	kTotalStrokes	21
U+9FE1	kTotalStrokes	33
U+9FE2	kTotalStrokes	14
U+9FE3	kTotalStrokes	15
U+9FE4	kTotalStrokes	18
U+9FE5	kTotalStrokes	22
U+9FE6	kTotalStrokes	23
U+9FE7	kTotalStrokes	25
U+9FE8	kTotalStrokes	27
U+9FE9	kTotalStrokes	29
U+9FEA	kTotalStrokes	17

**Comments:** The group discussed this feedback, and agreed that these 21 *kTotalStrokes* property values should be added to the Unihan database.

**Recommendations (Unihan-UTC168-R13):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins, Michel Suignard:** Add 21 *kTotalStrokes* property values in the Unihan database per feedback dated [Thu Jun 17 07:59:42 CDT 2021] in PRI #433 and Unihan-UTC168-R13 in document L2/21-129, for Unicode Version 14.0.

#### Add:

U+9FD6	kTotalStrokes	7
U+9FD7	kTotalStrokes	9
U+9FD8	kTotalStrokes	12
U+9FD9	kTotalStrokes	13
U+9FDA	kTotalStrokes	16
U+9FDB	kTotalStrokes	24
U+9FDC	kTotalStrokes	22
U+9FDD	kTotalStrokes	27
U+9FDE	kTotalStrokes	20
U+9FDF	kTotalStrokes	12
U+9FE0	kTotalStrokes	21
U+9FE1	kTotalStrokes	33
U+9FE2	kTotalStrokes	14
U+9FE3	kTotalStrokes	15
U+9FE4	kTotalStrokes	18
U+9FE5	kTotalStrokes	22
U+9FE6	kTotalStrokes	23
U+9FE7	kTotalStrokes	25
U+9FE8	kTotalStrokes	27
U+9FE9	kTotalStrokes	29
U+9FEA	kTotalStrokes	17

Date/Time: **Fri Jun 18 15:49:09 CDT 2021**

Name: Paul Masson

Report Type: Error Report

Opt Subject: **kPhonetic for U+96B1**

This character appears in group 1483 on p.152 of Casey. The field is missing in the database and needs to be added.

**Comments:** The group discussed this feedback, and agreed that the suggested *kPhonetic* property value should be added to the Unihan database. Because this is a provisional property, John Jenkins added this property value during the meeting.

**Recommendations (Unihan-UTC168-R14):** The group recommends to the UTC the following disposition ➡ **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Fri Jun 18 15:49:09 CDT 2021]

#### Add:

U+96B1	kPhonetic	1483
--------	-----------	------

Date/Time: **Sat Jun 26 09:07:52 CDT 2021**

Name: Ken Lunde

Report Type: Public Review Issue

Opt Subject: **PRI #433 (Unicode Version 14.0.0 Beta) feedback**

The kIRG\_VSource property value for U+20307 御 should be changed from V4-4131 to VN-20307. IRG Working Set 2017 (aka Extension H) serial number 00138 will use V4-4131 as its kIRG\_VSource property value, and Vietnam confirmed that it

is correct, and that the `kIRG_VSource` property value for U+20307 儋 should be changed to VN-20307. See:

<https://hc.jsecs.org/irg/ws2017/app/index.php?id=00138>

**Comments:** The group discussed this feedback, and agreed that the `kIRG_VSource` property value for U+20307 儋 should be changed from V4-4131 to VN-20307. The current property value is shown below.

Current:

U+20307      `kIRG_VSource`      V4-4131

**Recommendations (Unihan-UTC168-R15):** The group recommends to the UTC the following disposition ⇒ **Action Item for John Jenkins, Michel Suignard:** Change the `kIRG_VSource` property value in the Unihan database per feedback dated [Sat Jun 26 09:07:52 CDT 2021] in PRI #433 and Unihan-UTC168-R15 in document L2/21-129, for Unicode Version 14.0.

Change:

U+20307      `kIRG_VSource`      VN-20307

Date/Time: **Sun Jun 27 05:59:21 CDT 2021**

Name: M

Report Type: Other Question, Problem, or Feedback

Opt Subject: **Suggestion for Unihan Database**

You can add "stroke order" information to Unihan Database, e.g. `kStroke`. Represented by numbers 1-5, 5 basic strokes in Chinese. It is called 笔顺号码 (stroke order numbers) in Chinese.

一: | : 丿 : ㇏ : ㇚

横: 竖: 撇: 捺: 折

héng: shù: piě: nà: zhé

1: 2: 3: 4: 5

1: 横 >> 一 丿 (from left to right, from bottom left to top right)

2: 竖 >> | 丨 (from top to bottom)

3: 撇 >> 丿 ㇏ (from top right to bottom left)

4: 捺 >> ㇏ ㇚ (from top left to bottom right)

5: 折 >> ㇚ ㇔ ㇕ ㇖ ㇗ ㇘ ㇙ ㇚ (anything else that folds except 丨 which is included in the 2nd type of stroke)

Let's pick U+7B14 笔 [bǐ] as an example:

It's made of 丿 一 ㇏ ㇚ ㇔ ㇕ ㇖ ㇗ ㇘ ㇙ ㇚ [撇横捺撇横捺撇横横折] in stroke order.

So the `kStroke` value of this character would be 3143143115.

Probably you must know all this.

This would help find characters by stroke order quickly.

Thanks

**Comments:** The group discussed this feedback, and while the group would favor *kStrokeOrder* as the name of the new Unihan property, someone first needs to submit a formal proposal for adding a new provisional Unihan property and supply the data in a form that is not encumbered by a license that is incompatible with the [Unicode License](#).

**Recommendations (Unihan-UTC168-R16):** The group recommends to the UTC the following disposition ➡ **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Sun Jun 27 05:59:21 CDT 2021]

Date/Time: **Tue Jun 29 21:21:24 CDT 2021**

Name: Ryusei Yamaguchi

Report Type: Public Review Issue

Opt Subject: **PRI #433 Unicode 14.0.0 Beta**

Unihan\_IRGSources.txt from <https://www.unicode.org/Public/14.0.0/ucd/Unihan-14.0.0d4.zip>

has some bugs: the *kTotalStrokes* for following characters don't have exact stroke counts.

UCS, char, current, correct

U+21FE8, 𠄐, 3, 1

U+248E5, 𠄐, 5, 4

U+2634D, 𠄐, 6, 5

U+264D0, 𠄐, 6, 5

U+26612, 𠄐, 6, 5

U+27607, 𠄐, 5, 4

U+2795B, 𠄐, 7, 6

U+2795C, 𠄐, 7, 6

U+27C27, 𠄐, 7, 6

U+27C28, 𠄐, 7, 6

U+28210, 𠄐, 7, 5

U+28211, 𠄐, 7, 6

**Comments:** The group discussed this feedback, and agreed that these *kTotalStrokes* property value changes, whose current values are shown below, are appropriate.

Current:

U+21FE8	<i>kTotalStrokes</i>	3
U+248E5	<i>kTotalStrokes</i>	5
U+2634D	<i>kTotalStrokes</i>	6
U+264D0	<i>kTotalStrokes</i>	6
U+26612	<i>kTotalStrokes</i>	6
U+27607	<i>kTotalStrokes</i>	5
U+2795B	<i>kTotalStrokes</i>	7
U+2795C	<i>kTotalStrokes</i>	7
U+27C27	<i>kTotalStrokes</i>	7
U+27C28	<i>kTotalStrokes</i>	7
U+28210	<i>kTotalStrokes</i>	7

**Recommendations (Unihan-UTC168-R17):** The group recommends to the UTC the following disposition ⇒ **Action Item for John Jenkins, Michel Suignard:** Change the *kTotalStrokes* property values in the Unihan database per feedback dated [Tue Jun 29 21:21:24 CDT 2021] in PRI #433 and Unihan-UTC168-R17 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Tue Jun 29 21:21:24 CDT 2021]

Change:

U+21FE8	kTotalStrokes	1
U+248E5	kTotalStrokes	4
U+2634D	kTotalStrokes	5
U+264D0	kTotalStrokes	5
U+26612	kTotalStrokes	5
U+27607	kTotalStrokes	4
U+2795B	kTotalStrokes	6
U+2795C	kTotalStrokes	6
U+27C27	kTotalStrokes	6
U+27C28	kTotalStrokes	6
U+28210	kTotalStrokes	5
U+28211	kTotalStrokes	6

Date/Time: **Fri Jul 2 03:17:10 CDT 2021**

Name: M

Report Type: Other Question, Problem, or Feedback

Opt Subject: **Missing Data in Unihan Database**

Found some missing simplification data in Unihan\_Variants

U+44D6 kTraditionalVariant U+85ED

U+4E86 kTraditionalVariant U+77AD

U+4F19 kTraditionalVariant U+5925

U+501F kTraditionalVariant U+85C9

U+51AC kTraditionalVariant U+9F15

U+5343 kTraditionalVariant U+97C6

U+535C kTraditionalVariant U+8514

U+5377 kTraditionalVariant U+6372

U+5401 kTraditionalVariant U+7C72

U+5408 kTraditionalVariant U+95A4

U+56DE kTraditionalVariant U+8FF4

U+59DC kTraditionalVariant U+8591

U+5BB6 kTraditionalVariant U+50A2

U+5CC3 kTraditionalVariant U+5DA8

U+5EBC kTraditionalVariant U+5ECE

U+624D kTraditionalVariant U+7E94

U+6298 kTraditionalVariant U+647A

U+65CB kTraditionalVariant U+93C7  
U+6731 kTraditionalVariant U+7843  
U+7076 kTraditionalVariant U+7AC8  
U+79CB kTraditionalVariant U+97A6  
U+8499 kTraditionalVariant U+61DE U+6FDB U+77C7  
U+8511 kTraditionalVariant U+884A  
U+9709 kTraditionalVariant U+9EF4

U+85ED kSimplifiedVariant U+44D6  
U+77AD kSimplifiedVariant U+4E86  
U+5925 kSimplifiedVariant U+4F19  
U+85C9 kSimplifiedVariant U+501F  
U+9F15 kSimplifiedVariant U+51AC  
U+97C6 kSimplifiedVariant U+5343  
U+8514 kSimplifiedVariant U+535C  
U+6372 kSimplifiedVariant U+5377  
U+7C72 kSimplifiedVariant U+5401  
U+95A4 kSimplifiedVariant U+5408  
U+8FF4 kSimplifiedVariant U+56DE  
U+8591 kSimplifiedVariant U+59DC  
U+50A2 kSimplifiedVariant U+5BB6  
U+5DA8 kSimplifiedVariant U+5CC3  
U+5ECE kSimplifiedVariant U+5EBC  
U+7E94 kSimplifiedVariant U+624D  
U+647A kSimplifiedVariant U+6298  
U+93C7 kSimplifiedVariant U+65CB  
U+7843 kSimplifiedVariant U+6731  
U+7AC8 kSimplifiedVariant U+7076  
U+97A6 kSimplifiedVariant U+79CB  
U+61DE kSimplifiedVariant U+8499  
U+6FDB kSimplifiedVariant U+8499  
U+77C7 kSimplifiedVariant U+8499  
U+884A kSimplifiedVariant U+8511  
U+9EF4 kSimplifiedVariant U+9709

**Comments:** The group discussed this feedback, and agreed that these are appropriate additions, though according to Henry Chan, U+85C9 藉 and U+77AD 瞭 are also simplified forms themselves, and should therefore be added as such. All can be found in [this table](#).

**Recommendations (Unihan-UTC168-R18):** The group recommends to the UTC the following disposition ➡ **Action Item for John Jenkins:** Add the *kSimplifiedVariant* and *kTraditionalVariant* property values in the Unihan database per feedback dated [Fri Jul 2 03:17:10 CDT 2021] in PRI #433 as amended by the CJK & Unihan Group, and Unihan-UTC168-R18 in document L2/21-129, for Unicode Version 14.0. **Action Item for Ken Lunde:** Convey to the feedback submitter the CJK & Unihan Group comments in document L2/21-129. [Fri Jul 2 03:17:10 CDT 2021]

Add:

U+44D6	kTraditionalVariant	U+85ED
U+4E86	kTraditionalVariant	U+77AD
U+4F19	kTraditionalVariant	U+5925
U+501F	kTraditionalVariant	U+85C9
U+50A2	kSimplifiedVariant	U+5BB6
U+51AC	kTraditionalVariant	U+9F15
U+5343	kTraditionalVariant	U+97C6
U+535C	kTraditionalVariant	U+8514
U+5377	kTraditionalVariant	U+6372
U+5401	kTraditionalVariant	U+7C72
U+5408	kTraditionalVariant	U+95A4
U+56DE	kTraditionalVariant	U+8FF4
U+5925	kSimplifiedVariant	U+4F19
U+59DC	kTraditionalVariant	U+8591
U+5BB6	kTraditionalVariant	U+50A2
U+5CC3	kTraditionalVariant	U+5DA8
U+5DA8	kSimplifiedVariant	U+5CC3
U+5EBC	kTraditionalVariant	U+5ECE
U+5ECE	kSimplifiedVariant	U+5EBC
U+61DE	kSimplifiedVariant	U+8499
U+624D	kTraditionalVariant	U+7E94
U+6298	kTraditionalVariant	U+647A
U+6372	kSimplifiedVariant	U+5377
U+647A	kSimplifiedVariant	U+6298
U+65CB	kTraditionalVariant	U+93C7
U+6731	kTraditionalVariant	U+7843
U+6FDB	kSimplifiedVariant	U+8499
U+7076	kTraditionalVariant	U+7AC8
U+77AD	kSimplifiedVariant	U+4E86 U+77AD
U+77C7	kSimplifiedVariant	U+8499
U+7843	kSimplifiedVariant	U+6731
U+79CB	kTraditionalVariant	U+97A6
U+7AC8	kSimplifiedVariant	U+7076
U+7C72	kSimplifiedVariant	U+5401
U+7E94	kSimplifiedVariant	U+624D
U+8499	kTraditionalVariant	U+61DE U+6FDB U+77C7
U+8511	kTraditionalVariant	U+884A
U+8514	kSimplifiedVariant	U+535C
U+8591	kSimplifiedVariant	U+59DC
U+85C9	kSimplifiedVariant	U+501F U+85C9
U+85ED	kSimplifiedVariant	U+44D6
U+884A	kSimplifiedVariant	U+8511
U+8FF4	kSimplifiedVariant	U+56DE
U+93C7	kSimplifiedVariant	U+65CB
U+95A4	kSimplifiedVariant	U+5408
U+9709	kTraditionalVariant	U+9EF4

U+97A6	kSimplifiedVariant	U+79CB
U+97C6	kSimplifiedVariant	U+5343
U+9EF4	kSimplifiedVariant	U+9709
U+9F15	kSimplifiedVariant	U+51AC

#### 4) UAX #38 / Unihan Database Document for Unicode Version 14.0

The document in this section was discussed by the group, and for convenience, the *Unihan-additions-20210716.txt* and *Unihan-changes-20210716.txt* (PDF attachments) files include all of the recommended Unihan database additions and changes based on this document.

#### L2/21-119: Feedback on 2020-2021 proposed kMandarin changes for 45 characters (per several UTC actions)

**Comments:** The group discussed this document, which is the result of the CLDR Committee vetting suggested changes and additions to the *kMandarin* property. No one in the group objected to the recommendations in this document. Any current *kMandarin* property values are shown below.

Current:

U+7B7D	kMandarin	o
U+9730	kMandarin	sǎn
U+2B413	kMandarin	bù

**Recommendations (Unihan-UTC168-R19):** The group recommends to the UTC the following disposition ⇒ Accept the *kMandarin* property value changes and additions as specified in L2/21-119, based on Unihan-UTC168-R19 in document L2/21-129, for Unicode Version 14.0.

**Action Item for John Jenkins:** Update the Unihan database, based on document L2/21-119 and Unihan-UTC168-R19 in document L2/21-129, for Unicode Version 14.0.

Change:

U+7B7D	kMandarin	ōu wú
U+9730	kMandarin	xiàn
U+2B413	kMandarin	róu

Add:

U+348B	kMandarin	sī
U+34EE	kMandarin	diāo
U+37D7	kMandarin	bīn
U+38FA	kMandarin	xīn
U+3A85	kMandarin	gèng gēng
U+3BC3	kMandarin	qī
U+3BFD	kMandarin	bīn
U+4103	kMandarin	yāng
U+4674	kMandarin	qiān
U+4C17	kMandarin	jiū
U+4C7D	kMandarin	chāng
U+FA18	kMandarin	lǐ
U+20041	kMandarin	dòu
U+2009D	kMandarin	zā

U+2353C	kMandarin	gān
U+24360	kMandarin	rè
U+24514	kMandarin	luàn
U+25081	kMandarin	yán
U+255FD	kMandarin	lán
U+27039	kMandarin	zhè
U+27DB6	kMandarin	fàn
U+289C0	kMandarin	dù
U+28A0F	kMandarin	bō
U+28B4E	kMandarin	xǐ
U+2B737	kMandarin	shòu
U+2B738	kMandarin	huà
U+2B825	kMandarin	lòu
U+2BBAC	kMandarin	jiàng
U+2BF1B	kMandarin	diào
U+2BF6E	kMandarin	dǎng
U+2C037	kMandarin	shāng
U+2C19B	kMandarin	nǎo
U+2C41A	kMandarin	hǎi
U+2C454	kMandarin	zhòng
U+2C461	kMandarin	chǒu
U+2C498	kMandarin	xìn
U+2C6F8	kMandarin	niē
U+2CC03	kMandarin	mù
U+2CC9C	kMandarin	chēng
U+2CCF2	kMandarin	chǎn
U+2D016	kMandarin	lí
U+2F835	kMandarin	huī

## 5) UAX #45 / U-Source Document

The document in this section was discussed by the group, and for convenience, the *U-Source-additions-20210716.txt* (PDF attachment) file includes all of the recommended UAX #45 *U-SourceData.txt* data file additions based on this document. These recommended data file additions are also shown inline, following the Recommendations. In addition, the *UTCHan2021u168-Regular.ttf* TrueType font (PDF attachment) includes the representative glyphs for adding to the UAX #45 *UTCHan.ttf* TrueType font.

## L2/21-101: Proposal to Encode the Characters Occurring in the Man'yōshū

**Comments:** The group reviewed these nine ideographs, and while they are all acceptable additions to UAX #45, the group noted that the first and second ideographs in this proposal, 𣏟 and 𣞵, were submitted by the UK for IRG Working Set 2021 as UK-20901 and UK-20900, respectively, with Serial Numbers 01785 and 02549, respectively. It is reasonable to include them in UAX #45 in case they are removed or withdrawn from that IRG working set. The group also improved the IDS for the eighth ideograph.

**Recommendations (Unihan-UTC168-R20):** The group recommends to the UTC the following disposition ➡ Accept nine new U-Source ideographs as **UTC-03239** through **UTC-03247** with a UAX #45 status value of N, based on document L2/20-277 and Unihan-UTC168-R20 in document

L2/21-129, for Unicode Version 15.0. **Action Item for John Jenkins:** Add nine new records to *USourceData.txt* and their representative glyphs to *USourceGlyphs.pdf*, based on document L2/21-101 and Unihan-UTC168-R20 in document L2/21-129, for Unicode Version 15.0.

Add:

UTC-03239;N;U+2BB71;75.7;;𠩺木皿;UTCDoc L2/21-101 1;;11;1  
UTC-03240;N;U+7F3B;98.6;;𠩺瓦缶;UTCDoc L2/21-101 2;;11;3  
UTC-03241;N;U+2E99B;173.10;;𠩺雨流;UTCDoc L2/21-101 3;;18;4  
UTC-03242;N;U+4FDD;113.7;;𠩺𠩺呆;UTCDoc L2/21-101 4;;11;2  
UTC-03243;N;U+973A;173.13;;𠩺雨徽;UTCDoc L2/21-101 5;;21;3  
UTC-03244;N;U+5E79 U+2D16A;37.11;;𠩺卓夸;UTCDoc L2/21-101 6;;14;1  
UTC-03245;N;U+7E20;120.9;;𠩺𠩺士一糸爻;UTCDoc L2/21-101 7;;15;1  
UTC-03246;N;U+7E3F;120.13;;𠩺糸𠩺;UTCDoc L2/21-101 8;;19;5  
UTC-03247;N;U+84AD;140.6;;𠩺𠩺𠩺𠩺;UTCDoc L2/21-101 9;;10;5

## 6) Other UAX #38 / Unihan Database Documents

The documents in this section were discussed by the group.

**L2/21-118:** Preliminary proposal to add a new provisional *kIDS* property (Unihan)

**L2/21-134:** Collections of ideograph components for use in IDSes

**IRG N2273R Response:** Feedback on Proposals of 4 IDCs (IRG N2273R)

**Comments:** The group discussed L2/21-118, which is a preliminary proposal that aims to 1) encode four or five new Ideographic Description Characters (IDCs); 2) encode 100 or more ideograph components to be used in Ideographic Description Sequences (IDSes); and 3) add a new provisional *kIDS* property that provides at least one valid IDS per CJK Unified Ideograph. The target for #1 and #2 is Unicode Version 15.0, and the target for #3 is Unicode Version 16.0. The group agreed that there is precedence to treat ideograph components as CJK Unified Ideographs, but that a new block, tentatively named *CJK Unified Ideographs Components*, be allocated after the formal proposal for #2 has been prepared and submitted. The range U+2FA20 through U+2FAFF (224 code points) was suggested in the preliminary proposal, which provides growing room for the encoding of additional ideograph components. Although L2/21-134, which includes a candidate set of 119 ideograph components, was referenced during the meeting, it will be reviewed offline during the coming weeks.

**Recommendations (Unihan-UTC168-R21):** The group recommends to the UTC the following disposition ⇒ Discuss the possibility of encoding four or five new Ideographic Description Characters (IDCs), along with 100+ ideograph components as CJK Unified Ideographs for the purpose of adding a new provisional *kIDS* property to the Unihan database with a timeline as discussed in L2/21-118. **Action Item for Ken Lunde:** Update L2/21-118 to indicate that the ideograph components will be added to UAX #45, and will therefore have *kIRG\_USource* property values. The update should also reflect any feedback from the UTC, then be submitted to the IRG for discussion during IRG #57 in September.

## 7) IRG & IRG-related Documents

The documents in this section were discussed by the group.

### IRG N2475: CJK Unified Ideographs Extension H 2021-07-08 DRAFT

**Comments:** This document is the current draft of the CJK Unified Ideographs Extension H (aka IRG Working Set 2017) code chart that includes 4,192 ideographs in the range U+31350 through U+323AF. The group confirmed that the issues that were present in the previous draft have been fixed, and that this new draft reflects changes based on the disunification of two CJK Compatibility Ideographs, U+2F8A4 𠂔 and U+2F936 𠂔, whose draft code points are U+317AB and U+31C2E, respectively. The group further recommends that Extension H be included in the repertoire of Unicode Version 15.0, though that action should be considered during the UTC #169 meeting, which takes place after the IRG #57 meeting (2021-09-13 through 2021-09-17).

### L2/21-097: UTC Submission for IRG Working Set 2021 (IRG N2488)

**Comments:** Although this is an FYI document, the group would like to inform the UTC that the IRG Working Set 2021 [Online Review Tool](#) is now available, and reflects Version 1.0 of this IRG working set. The total number of ideographs in this submission is 153.

## 8) Other Unihan-related Topics

The Unicode Consortium is still in the process of establishing a formal liaison relationship with the SAT (*SAṃgaṇikīkṛtaṃ Taiśotripiṭakaṃ*) Daizōkyō Text Database Committee (aka [SAT](#)). Ken Lunde has agreed to serve as the Unicode Consortium's liaison contact.

In terms of the status of licensing the [Moji Jōhō Kiban Database](#) from CITPC (*Character Information Technology Promotion Council*) to 1) establish the new provisional *kMojiJoho* property as proposed in [L2/20-146](#); 2) improve and expand the existing provisional *kMorohashi* property as proposed in [L2/21-032](#); and 3) establish a new provisional *kJapanese* property that provides Japanese readings expressed in hiragana and katakana, the Unicode Consortium received a draft license agreement from CITPC on 2021-01-23. After consulting with Legal, feedback was provided to CITPC on 2021-03-07. The last communication with CITPC on this matter was on 2021-07-18 (PDT). While these Unihan database additions and enhancements are still targeted for Unicode Version 15.0 (2022), there is now risk that they may need to be targeted for Unicode Version 16.0 (2023) instead.

That is all.