Title: CJK & Unihan Group Recommendations for UTC #166 Meeting
Author: Ken Lunde and CJK & Unihan Group
Date: 2021-01-15

This document provides the CJK & Unihan Group recommendations for UTC #166 based on a meeting that took place from 6 to 9PM PST on 2021-01-08, which was attended by Eiso Chan, Lee Collins, John Jenkins, Ken Lunde, William Nelson, Stephan Hyeonjun Stiller, Yifán Wáng, and Selena Wei via Zoom. John Jenkins and Ken Lunde co-chaired the meeting.

The CJK & Unihan Group reviewed public feedback and documents that were received since UTC #165. Comments are marked in **green**, and Copy&Paste-ready Recommendations to the UTC are marked in **red**.

1) UAX #45 / U-Source Documents

The documents in this section were discussed by the group, and for convenience, the *USourceData-additions-20210108.txt* (PDF attachment) file includes all of the recommended UAX #45 *USourceData.txt* data file additions based on those documents. These recommended data file additions are also shown inline, following the Recommendations. In addition, the *UTCHan2021u166-Regular.ttf* TrueType font (PDF attachment) includes the representative glyphs for adding to the UAX #45 *UTCHan.ttf* TrueType font.

L2/20-277: Proposal to Add One UTC-Source Ideograph to UAX #45

Comments: This document proposes that one ideograph be added to UAX #45 as a new U-Source ideograph, and provides sufficient metadata and evidence. As the proposal demonstrates, this ideograph is a well-attested variant of U+5B78 學, and is therefore also a variant of U+5B66 学 and U+6588 孝. Whether this ideograph is unifiable with an existing ideograph is not a barrier to adding it to UAX #45.

Recommendations (**Unihan-UTC166-R01**): The group recommends to the UTC the following disposition ⇨ Accept one new U-Source ideograph as **UTC-03234** with a UAX #45 status value of V, based on document L2/20-277 and Unihan-UTC166-R01 in document L2/21-015, for Unicode Version 14.0. **Action Item for John Jenkins**: Add one new record to *USourceData.txt* and its representative glyph to *USourceGlyphs.pdf*, based on document L2/20-277 and Unihan-UTC166-R01 in document L2/21-015, for Unicode Version 14.0.

```
UTC-03234;V;U+5B78;39.13;;⿳⿲⿱臼冖文子;UTCDoc L2/20-277;;16;3
```

L2/21-025: Proposal to Add to UAX #45 Four Ideographs from Japan

Comments: This document proposes that four ideographs be added to UAX #45 as new U-Source ideographs, and provides sufficient metadata and evidence. The group noted that the IDS for the fourth ideograph includes as a component an Extension G ideograph that is not

yet broadly supported in fonts, but concluded that its use is appropriate. An expert from TCA noted that the first two ideographs are included in the CNS 11643 standard as 12-227E and 12-2578, respectively, meaning that TCA is likely to submit them for the next IRG working set.

Recommendations (**Unihan-UTC166-R02**): The group recommends to the UTC the following disposition ⇨ Accept four new U-Source ideographs as **UTC-03235** through **UTC-03238** with a UAX #45 status value of N, based on document L2/21-025 and Unihan-UTC166-R02 in document L2/21-015, for Unicode Version 14.0. **Action Item for John Jenkins**: Add four new records to *USourceData.txt* and their representative glyphs to *USourceGlyphs.pdf*, based on document L2/21-025 and Unihan-UTC166-R02 in document L2/21-015, for Unicode Version 14.0.

```
UTC-03235;N;;15.3;;⿰冫夕;UTCDoc L2/21-025 1;kSpoofingVariant U+6C50;5;3
UTC-03236;N;;32.3;;⿰土夕;UTCDoc L2/21-025 2;;6;3
UTC-03237;N;;86.12;;⿰晶灬;UTCDoc L2/21-025 3;kJapaneseOn SHOU;16;2
UTC-03238;N;;162.27;;⿰辶 ;UTCDoc L2/21-025 4;;31;1
```

Note: The last component of the IDS for UTC-03238 is not displayed as intended in this document due to the lack of a font with a glyph for Extension G ideograph U+30182 whose own IDS is ⿲南⿱南南.

2) UAX #38 / Unihan Database Public Feedback

The single item of public feedback from L2/21-011 that is in this section was discussed by the group.

```
Date/Time: Thu Jan 7 15:18:08 CST 2021
Name: William T. Nelson
Report Type: Error Report
Opt Subject: UAX38 kMandarin readings for two ideographs

U+7B7D 筽 has kMandarin value o which is its Korean reading. Please change the
value to wú as per 两万汉字中日韩越英俄读音释义字典 page 846 entry 17174.

U+9730 霰 has kMandarin value sǎn, but the correct reading is xiàn according
to character dictionaries. (The PRI #297 feedback page has an error report
from Markus Scherer regarding this value.)
```

Comments: The group reviewed this feedback, and agreed that the suggested *kMandarin* property value changes are appropriate. The group noted that the existing *kMandarin* property value for U+9730 霰, *sǎn*, is completely valid, but because this property can specify only a single reading for zh-Hans (CN), the more common reading, *xiàn*, should be specified.

Recommendations (**Unihan-UTC166-R03**): The group recommends to the UTC the following disposition ⇨ **Action Item for Peter Edberg**: Ask the CLDR-TC to check the proposed *kMandarin* property value changes for U+7B7D 筽 and U+9730 霰 then report back to the UTC. See document L2/21-011 and Unihan-UTC166-R03 in document L2/21-015.

3) UAX #38 / Unihan Database Documents

The documents in this section were discussed by the group.

L2/21-022: Status on the Update to the Unihan kCantonese Field

Comments: The group discussed this document, which conveys the current status of the update to the *kCantonese* property as proposed in L2/20-153 and per Action Item 164-A11 to which John Jenkins was assigned. John indicated that the 26,698 ideographs in the *kCantonese-Values.txt* data file, which is a PDF attachment of L2/21-022, are stable and can be accepted. The remaining 3,470 ideographs that are listed in the *kCantonese-Lookup.txt* data file, which is also a PDF attachment of L2/21-022, have uncertain *kCantonese* property values, and therefore need to be checked by Cantonese experts. John will post the 3,470 uncertain *kCantonese* property values to the *#cantonese* channel in the Unicode Slack workspace for verification by Cantonese experts.

Recommendations (**Unihan-UTC166-R04**): The group recommends to the UTC the following disposition ⇨ Accept the 26,698 *kCantonese* property values as specified in the *kCantonese-Values.txt* data file that is attached to L2/21-022, based on Unihan-UTC166-R04 in document L2/21-015, for Unicode Version 14.0. **Action Item for John Jenkins**: Update the Unihan database by replacing the existing *kCantonese* property values with the 26,698 *kCantonese* property values, based on document L2/21-022 and Unihan-UTC166-R04 in document L2/21-015, for Unicode Version 14.0.

L2/21-032: Proposal to remove and improve provisional Unihan database properties (draft)

Comments: The group discussed this draft proposal for removing the provisional *kIRGDaiKanwaZiten* property and enhancing the provisional *kMorohashi* property. This represents a considerable expansion that brings this property to nearly full coverage of the 大漢和辞典 kanji dictionary to which it corresponds. CITPC (*Character Information Technology Promotion Council*) has tentatively granted to the UTC permission to use the Moji Jōhō Kiban (文字情報基盤) database to 1) establish the provisional *kMojiJoho* property as proposed in L2/20-146; and 2) improve and expand the existing the *kMorohashi* property as proposed in this draft proposal. Formal permission, which will effectively bypass the CC BY-SA 2.1 JP license, is expected to be received shortly. After formal permision has been received, Ken Lunde will submit a revised version of L2/20-146, along with a final version of this proposal.

Recommendations (**Unihan-UTC166-R05**): The group recommends to the UTC the following disposition ⇨ **Action Item for Ken Lunde**: Follow up with CITPC's liaison contact, Taro YAMAMOTO, about the status of the request for permission to use the Moji Jōhō Kiban Database to 1) establish the provisional *kMojiJoho* property as proposed in L2/20-146; and 2) improve and expand the existing the *kMorohashi* property as proposed in L2/21-032 (draft).

4) Other Unihan-related Documents

The documents in this section were discussed by the group, and for convenience, the *Unihan-additions-20210108.txt*, *Unihan-removals-20210108.txt*, *Unihan-changes-20210108.txt*, and *USource-changes-20210108.txt* (PDF attachments) files include all of the recommended Unihan database and *USourceData.txt* additions, removals, and changes based on those documents. These recommended Unihan database and *USourceData.txt* additions, removals, and changes are also shown inline, following the Recommendations.

L2/20-276: Suggestion concerning tentative roadmap placement of Kanbun Extended block

Comments: The group discussed this proposal to encode the proposed Kanbun Extended block in the BMP instead of in the SMP. The group didn't see a valid reason for moving the not-yet-accepted block to the BMP. The group further felt that it is prudent to keep that unassigned block open for characters that do have a valid reason for being encoded in the BMP. In addition, a recently-revised version of the original proposal that has not yet been reviewed increases the number of proposed characters beyond 16, which effectively renders the proposal to move the 16-character block to the BMP moot.

Recommendations (**Unihan-UTC166-R06**): The group recommends to the UTC the following disposition ⇨ **Action Item for Ken Lunde**: Convey to the proposal author the CJK & Unihan Group feedback.

L2/20-291: Request for consideration to disunify U+3B3F (IRG N2443)

Comments: The group discussed this proposal to disunifiy U+3B3F 㬿, which is a de facto UNC (*Urgently Needed Character*) proposal. The group agreed that the justification for disunification is sound based on semantics and the different form of Radical #130 for some regions. The next available code point is U+2B736 (Extension C). The representative glyph for the disunified U-Source ideograph, UTC-00732, was also discussed, because it is a Cantonese ideograph whose form should adhere to the regional conventions of Hong Kong SAR. The group agreed that the representative glyph for this U-Source ideograph should be modified accordingly. It was also determined that eight additional Cantonese U-Source ideographs require the same treatment, and one of the experts prepared the updated glyphs (see the table below). The *UTCHan2021u166-Regular.ttf* TrueType font (PDF attachment) includes the representative glyphs for modifying the UAX #45 *UTCHan.ttf* TrueType font. One of the experts noted that UTC-00443 should no longer be flagged as a duplicate of UTC-00345, because their representative glyphs do not share the same left-side radical. For future horizontal extension purposes, the same expert also noted that UTC-00731 is included in IRG Working Set 2017 as serial number 01658 with UK-Source source reference UK-10851.

U-Source	Current Glyph	Proposed Glyph
UTC-00345	朒	朒
UTC-00414	朤	朤
UTC-00728	朓	朓
UTC-00729	𦜗	𦜗
UTC-00730	腊	腊
UTC-00731	腤	腤
UTC-00732	朡	朡
UTC-00734	樸	樸
UTC-00735	臁	臁

Recommendations (**Unihan-UTC166-R07**): The group recommends to the UTC the following disposition ⇨ Accept the disunification proposed in L2/20-291, and encode the disunified ideograph at code point U+2B736 with Unihan database changes, based on Unihan-UTC166-R07 in document L2/21-015, for Unicode Version 14.0. **Action Item for Ken Whistler**: Update the pipeline to add the disunified ideograph with code point U+2B736. See Unihan-UTC166-R07 in document L2/21-015. **Action Item for John Jenkins and Michel Suignard**: Update the Unihan database by adding, removing, and changing property values, based on document L2/20-291 and Unihan-UTC166-R07 in document L2/21-015, and as amended by the group, for Unicode Version 14.0.

```
Add:
U+3B3F      kSpoofingVariant      U+2B736
U+3B3F      kVietnamese           thôn
U+2B736     kIRG_USource          UTC-00732
```

```
U+2B736        kIRG_VSource              V2-8166
U+2B736        kRSUnicode                130.12
U+2B736        kTotalStrokes             16
U+2B736        kDefinition               cheeks
U+2B736        kCheungBauerIndex         453.10
U+2B736        kVietnamese               trôn
U+2B736        kSpoofingVariant          U+3B3F
```
Remove:
```
U+3B3F         kIRG_USource              UTC-00732
```
Change:
```
U+3B3F         kIRG_VSource              VU-03B3F
```

Action Item for John Jenkins: Update the records for UTC-00443 and UTC-00732 in *USourceData.txt*, based on document L2/20-291 and Unihan-UTC166-R07 in document L2/21-015, for Unicode Version 14.0.

```
UTC-00443;N;;74.8;;☐月枕;kCowles 129;;12;1
UTC-00732;C;U+2B736;130.12;;☐月敦;kCheungBauerIndex 453.10;;16;4
```

Action Item for John Jenkins: Update the representative glyphs for UTC-00345, UTC-00414, UTC-00728 through UTC-00732, UTC-00734, and UTC-00735 to reflect a left-side form of Radical #130 that adheres to Hong Kong SAR regional conventions, based on document L2/20-291 and Unihan-UTC166-R07 in document L2/21-015, for Unicode Version 14.0. **Action Item for John Jenkins and Lee Collins**: Provide updated U- and V-Source fonts to Michel Suignard.

L2/21-018: Mismatched T-Source Identifiers

Comments: The group discussed this proposal to correct three T-Source source references. Based on the information that is available at the CNS 11643 website for U+55B9, U+841C, and U+2B74E, the corrections appear to be justified.

Recommendations (**Unihan-UTC166-R08**): The group recommends to the UTC the following disposition ⇨ Make changes to the Unihan database, based on document L2/21-018 and Unihan-UTC166-R08 in document L2/21-015, for Unicode Version 14.0. **Action Item for John Jenkins and Michel Suignard**: Update the Unihan database by changing three *kIRG_TSource* property values, based on document L2/21-018 and Unihan-UTC166-R08 in document L2/21-015, for Unicode Version 14.0.

```
U+55B9         kIRG_TSource              TE-5F2F
U+841C         kIRG_TSource              TE-5D48
U+2B74E        kIRG_TSource              TB-7334
```

L2/21-029: Proposal to Correct Four Vietnamese Glyphs (IRG N2445)

Comments: The group discussed this proposal to correct the representative glyphs for four V-Source ideopgraphs, one of which requires Unihan database changes. These corrections seem justified per the supplied evidence, as do the proposed Unihan database property value changes.

Recommendations (**Unihan-UTC166-R09**): The group recommends to the UTC the following disposition ⇨ Make changes to the Unihan database, based on document L2/21-029 and Unihan-UTC166-R09 in document L2/21-015, for Unicode Version 14.0. **Action Item for Lee Collins**: Provide an updated V-Source font to Michel Suignard. **Action Item for John Jenkins and Michel Suignard**: Update the Unihan database by changing property values, based on document L2/21-029 and Unihan-UTC166-R09 in document L2/21-015, for Unicode Version 14.0.

```
U+26F28      kRSUnicode            140.13
U+26F28      kTotalStrokes         17
U+26F28      kRSKangXi             140.13
```

L2/21-044: UNC Proposal for One G-Source Ideograph (IRG N2446)

Comments: The group discussed this UNC (*Urgently Needed Character*) proposal for one G-Source ideograph, which is the simplified form of U+287AA 䡪, and the consensus was that there is sufficient justification to treat this ideograph as a UNC. The implicit request to append this ideograph to the URO conflicts with Macao SAR's UNC proposal that was accepted during UTC #165, which consumes the last remaining code points at the end of the URO (3) and Extension B (2) blocks. The group recommends using the next available code point, which is U+2B737 (Extension C).

Recommendations (**Unihan-UTC166-R10**): The group recommends to the UTC the following disposition ⇨ Accept the urgently needed character proposed in L2/21-044 with code point U+2B737, based on Unihan-UTC166-R10 in L2/21-015, for Unicode Version 14.0. **Action Item for Ken Whistler**: Update the pipeline to add the urgently needed character with code point U+2B737. See Unihan-UTC166-R10 in document L2/21-015. **Action Item for China**: Provide a font to Michel Suignard. **Action Item for John Jenkins and Michel Suignard**: Update the Unihan database by adding property values, based on document L2/21-044 and Unihan-UTC166-R10 in document L2/21-015, for Unicode Version 14.0.

```
U+287AA      kSimplifiedVariant    U+2B737
U+2B737      kIRG_GSource          GDM-00221
U+2B737      kRSUnicode            163.7
U+2B737      kTotalStrokes         10
U+2B737      kMandarin             shòu
U+2B737      kTraditionalVariant   U+287AA
```

5) Other Unihan-related Topics

The UTC has established a liaison relationship with CITPC (*Character Information Technology Promotion Council* or 文字情報技術促進協議会 in Japanese). Ken Lunde serves as the UTC's liaison, and Taro YAMAMOTO (山本太郎) serves as CITPC's liaison.

6) IRG-related Topics

The IRG issued IRG Working Set 2017 Version 5.2 as IRG N2444, which was posted to the UTC document register as L2/20-023. Note that its attributes file (Microsoft Excel) is a PDF attachment. A review schedule has not yet been announced.

That is all.