Title: Unihan Ad Hoc Recommendations for UTC #163 Meeting
Author: Ken Lunde & Unihan Ad Hoc
Date: 2020-04-15

This document provides the Unihan Ad Hoc recommendations for UTC #163 based on a meeting that took place from 6 to 9PM PDT on 2020-04-10, which was attended by Eiso Chan, Lee Collins, John Jenkins, Ken Lunde, William Nelson, and Ben Yang via Zoom. John Jenkins and Ken Lunde co-chaired the meeting.

The Unihan Ad Hoc reviewed public feedback and documents that were received since UTC #162 or were otherwise unable to be reviewed during that meeting. Comments are marked in **green**, and Recommendations to the UTC are marked in **red**.

1) UAX #45 / UTC-Source Public Feedback

The *USourceData-changes-20200410.txt* (PDF attachment) file includes the changes that were proposed in the following three items of public feedback that were discussed by the ad hoc, along with additional proposed changes that have accumulated since Unicode Version 13.0 was released:

> Date/Time: **Wed Feb 19 02:51:51 CST 2020**
> Name: Eiso Chan
> Report Type: Error Report
> Opt Subject: **Corrections for UAX #45**
>
> In the current USourceData file, the data for UTC-00550 is shown as below.
> UTC-00550;UTC-02637;;30.11;0206.201;⿰口筌;kCheungBauerIndex 375.06;
> However, UTC-02637 has been updated to UK-02637, so Field 1 should be updated to UK-02637 correspondingly as below.
> UTC-00550;UK-02637;;30.11;0206.201;⿰口筌;kCheungBauerIndex 375.06;

Comments: This is clearly a clerical error, and should be fixed.

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins to update *USourceData.txt* to change the second field of UTC-00550 from **UTC-02637** to **UK-02637** per Eiso Chan's feedback, for Unicode Version 14.0. See *USourceData-changes-20200410.txt* (PDF attachment).

> Date/Time: **Wed Feb 19 05:38:12 CST 2020**
> Name: Eiso Chan
> Report Type: Error Report
> Opt Subject: **Corrections for UAX #45**
>
> The data for UTC-00749 should be updated as below.
> UTC-00749;UK-02870;;157.9;1230.291;⿰𧾷某;kCheungBauerIndex 463.06;
> UTC-00749 and UK-02870 are duplicates.

Comments: This is clearly an error, and should be fixed. UK-02870 is encoded in Extension G as U+30E1C, which is already reflected in its *USourceData.txt* entry. In addition, the first IDS component is currently U+2ECA ⻊ CJK RADICAL FOOT, and should be changed to U+27FB7 ⾷ . The following additional UTC-Source ideographs require the same change to the same IDS component: UTC-00051, UTC-00109, UTC-00248, UTC-00383, UTC-00748, UTC-00749, UTC-00750, UTC-00751, UTC-00752, UTC-00753, UTC-00754, and UTC-00810.

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins to update *USourceData.txt* to change the second field of UTC-00749 from **W** to **UK-02870** per Eiso Chan's feedback, and to change the IDS component U+2ECA ⻊ CJK RADICAL FOOT to U+27FB7 ⾷ in the IDSes for UTC-00051, UTC-00109, UTC-00248, UTC-00383, UTC-00748, UTC-00749, UTC-00750, UTC-00751, UTC-00752, UTC-00753, UTC-00754, and UTC-00810, for Unicode Version 14.0. See *USourceData-changes-20200410.txt* (PDF attachment).

> Date/Time: **Fri Mar 6 05:52:34 CST 2020**
> Name: Eiso Chan
> Report Type: Error Report
> Opt Subject: **IDSes for UTC-00475 and UTC-00476 in UAX #45**
>
> The IDSes for UTC-00475 and UTC-00476 should be updated as below,
> UTC-00475;V;U+947D;167.15;1326.121;□金□□夬夬貝;kLau 1580;
> UTC-00476;U;U+947D;167.17;1327.051;□金□□失失貝;kLau 1581;

Comments: These are clearly errors, and should be fixed. For reference, the current IDSes for UTC-00475 and UTC-00476 are □醒□□夬夬貝 and □醒□□失失貝, respectively, and do not match their representative glyphs as shown in *USourceGlyphs.pdf*.

鑽　鑽

UTC-00475　　UTC-00476

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins to update *USourceData.txt* to change the IDSes for **UTC-00457** and **UTC-00476** per Eiso Chan's feedback, for Unicode Version 14.0. See *USourceData-changes-20200410.txt* (PDF attachment).

2) UAX #45 / UTC-Source Documents

L2/20-073: Evidence for four U-source ideographs

Comments: This document provides evidence for the following UTC-Source ideographs: UTC-00258, UTC-00281, UTC-00536, and UTC-00609. This evidence makes it possible for the UTC to consider including these four UTC-Source ideographs in its submission to the next IRG working set. The evidence images have already been uploaded to the Google Drive folder for IRG Working Set 2020 evidence images.

Recommendations: The ad hoc recommends that the UTC thank William Nelson for gathering the evidence for these four existing UTC-Source ideographs, and for uploading them to the Google Drive folder for IRG Working Set 2020 evidence images.

L2/20-074: Proposal to add four characters to the U-source ideographs

Comments: This document proposes that four ideographs be added to UAX #45 as new UTC-Source ideographs, and provides sufficient metadata and evidence. Alexander Zapryagaev, who provided feedback to the ad hoc prior to the meeting, noted that misc-amanohashidate-itaiji-001 and misc-amanohashidate are variants of the fourth ideograph in the proposal. There was general agreement that the former is a non-unifiable variant, and that the latter is a unifiable variant.

Recommendations: The ad hoc recommends that the UTC accept these four new UTC-Source ideographs as **UTC-03211**, **UTC-03212**, **UTC-03213**, and **UTC-03214** with a UAX #45 status value of **N**, and assign an Action Item to John Jenkins to add the appropriate entries to *USourceData.txt* and their representative glyphs to *USourceGlyphs.pdf*, for Unicode Version 14.0. See *USourceData-additions-20200410.txt* (PDF attachment). The ad hoc further recommends that the UTC convey to Alexander Zapryagaev a suggestion to prepare and submit a proposal to add the ideographs that correspond to misc-amanohashidate-itaiji-001 and misc-amanohashidate to UAX #45.

L2/20-093: Proposal to Add Four UTC-Source Ideographs to UAX #45

Comments: This document proposes that four ideographs be added to UAX #45 as new UTC-Source ideographs, and provides sufficient metadata, though additional evidence may be necessary to submit them to the IRG in a future working set.

Recommendations: The ad hoc recommends that the UTC accept these four new UTC-Source ideographs as **UTC-03215**, **UTC-03216**, **UTC-03217**, and **UTC-03218** with a UAX #45 status value of **N**, and assign an Action Item to John Jenkins to add the appropriate entries to *USourceData.txt* and their representative glyphs to *USourceGlyphs.pdf*, for Unicode Version 14.0. See *USourceData-additions-20200410.txt* (PDF attachment).

L2/20-096: Proposed Addition to UAX #45

Comments: This document proposes that an ideograph be added to UAX #45 as a new UTC-Source ideograph, and provides sufficient metadata and evidence.

Recommendations: The ad hoc recommends that the UTC accept this new UTC-Source ideograph as **UTC-03219** with a UAX #45 status value of **N**, and assign an Action Item to John Jenkins to add the appropriate entry to *USourceData.txt* and its representative glyph to *USourceGlyphs.pdf*, for Unicode Version 14.0. See *USourceData-additions-20200410.txt* (PDF attachment).

3) UAX #38 / Unihan Database Public Feedback

The *Unihan-changes-20200410.txt* (PDF attachment) file includes the changes that were proposed in the following three items of public feedback were discussed by the ad hoc, along with additional proposed changes that have accumulated since Unicode Version 13.0 was released:

> Date/Time: **Mon Feb 17 07:38:13 CST 2020**
> Name: Eiso Chan
> Report Type: Error Report
> Opt Subject: **Radical for U+80E7 胧**
>
> The current radical for U+80E7 胧 is #130, but it's G-Source reference value is G0-6B4A, that means it's the simplified variant form of U+6727 朧 which the G-Source reference value is G1-6B4A and the radical is #74, not U+268AB. Is it better to update the radical for U+80E7 胧 to #74?

Comments: It is apparent that the radical for U+80E7 胧 should be #74 based on the fact that the radical for its traditional form, U+6727 朧, is #74.

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins and Michel Suignard to update the Unihan Database to change the *kRSUnicode* property value for U+80E7 胧 from **130.5** to **74.5** per Eiso Chan's feedback, for Unicode version 14.0. See *Unihan-changes-20200410.txt* (PDF attachment).

> Date/Time: **Mon Feb 17 07:38:54 CST 2020**
> Name: Eiso Chan
> Report Type: Error Report
> Opt Subject: **Radical for U+4E80 亀**
>
> The current radical for U+4E80 亀 is #5, and in ISO/IEC DIS 10646-1.2:1992, the references are T:E-396C and J:0-3535, and there is also the EACC source, 2D632D, in Unicode 1.0. It's the variant of U+9F9C 龜. Could we add the second radical as #213 under U+4E80 亀? Cf.
>
> https://mojikiban.ipa.go.jp/search/detail/MJ006424
>
> Notice that the radical for U+2A6C9 is #213, U+2A6C9 and U+4E80 亀 are similar.

Comments: It is apparent that the radical for U+4E80 亀 should be #213 based on the fact that it is Japan's simplified form of U+9F9C 龜, which represents Radical #213 itself. In this particular case, it is better to add a new *kRSUnicode* property value to U+4E80 亀.

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins and Michel Suignard to update the Unihan Database to add **213.0** as a second *kRSUnicode* property value for U+4E80 龜 per Eiso Chan's feedback, for Unicode Version 14.0. See *Unihan-changes-20200410.txt* (PDF attachment).

> Date/Time: **Mon Feb 17 11:34:50 CST 2020**
> Name: Lee Collins
> Report Type: Public Review Issue
> Opt Subject: **Wrong Japanese reading in Unihan**
>
> U+5807 shows the Kun reading "Sumire". This is no doubt a mistake for the similar but different character U+83EB. The respective DKWZ codes are 05212 and 31207. U+5807 should have the Kun readings NEBATUTI, NURU, WAZUKA, etc.
>
> The On readings are KIN and GON

Comments: Per Japan's *Moji Jōhō Kiban Database* entry for U+5807 堇, MJ009128, its *kJapaneseKun* property value should be changed to NURU NEBATSUCHI WAZUKA, and KON and GON should be added to its *kJapaneseOn* property value. The ad hoc also noted that this particular ideograph does not have a *kIRG_JSource* property value, but is included in Japan's *Moji Jōhō Kiban Database*.

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins to update the Unihan Database to change the *kJapaneseKun* property value of U+5807 堇 from **SUMIRE** to **NURU NEBATSUCHI WAZUKA**, and to add **KON** and **GON** to its *kJapaneseOn* property value, per Lee Collins' feedback and Japan's *Moji Jōhō Kiban Database*, for Unicode Version 14.0. See *Unihan-changes-20200410.txt* (PDF attachment).

4) UAX #38 / Unihan Database Documents

L2/20-059: Proposal for new provisional Unihan Database property: kStrange

Comments: The ad hoc was completely in favor of this new new provisional Unihan property, *kStrange*, with the understand that it is expected to grow over time as new CJK Unified Ideograph blocks are standardized, to include adding new categories to the current set of 12 and discovering new *kStrange* candidates that are still lurking in the existing CJK Unified Ideographs blocks.

Recommendations: The ad hoc recommends that the UTC accept *kStrange* as a new provisional Unihan property, and assign an Action Item to John Jenkins to add *kStrange* as a new provisional Unihan property, and to update the Unihan Database and UAX #38 per L2/20-059, for Unicode Version 14.0. The property data, along with the information that is necessary to update UAX #38, are provided in L2/20-059 as a PDF attachment. If accepted, there are additional Action Items due to the fact that this is a new property. The ad hoc further recommends that the UTC assign an Action Item to Ken Lunde to produce a draft of a new UTN (*Unicode Technical Note*) that documents this property, if accepted.

L2/20-077: Syntax regular expression issues in UAX #38

Comments: This Syntax (aka regular expression errors for the *kIRG_GSource*, *kIRG_HSource*, and *kIRG_TSource* properties as noted in this proposal were confirmed, and should be fixed.

Recommendations: The ad hoc recommends that the UTC accept this feedback, and assign an Action Item to John Jenkins to update the Syntax fields of the *kIRG_GSource*, *kIRG_HSource*, and *kIRG_TSource* properties in UAX #38 per L2/20-077, for Unicode Version 14.0.

L2/20-082: Proposal to update kHangul and kKoreanName

Comments: The ad hoc agreed with this proposal to adjust the *kHangul* and *kKoreanName* properties, but a minor adjustment is required to account for the fact that U+39B3 戝 (K3-2623) is a known and genuine duplicate of U+363D 戝 (K3-2321). Given that the KS standard in which these two ideographs appear, KS X 1027-2:2011, provides no reading information, and instead sorts the two ideographs according to two different radicals—Radical #32 土 for U+363D 戝 (K3-2321) and Radical #62 戈 for U+39B3 戝 (K3-2623)—it is not possible to objectively prefer one over the other, other than to select the one that appears first in the CJK Unified Ideographs Extension A block. Annotated excerpts from the KS X 1027-2:2011 standard, which show how U+363D 戝 (K3-2321 and U+39B3 戝 (K3-2623) are sorted according to radical/strokes, are shown below:

The ad hoc therefore concluded that both ideographs should be covered by these two properties.

Recommendations: The ad hoc recommends that the UTC respond to the proposal author to recommend adjusting the proposal to include U+39B3 戝 per the ad hoc comments, along with updating the Syntax and Description fields for the *kKoreanName* property, then resubmit.

L2/20-094: □魚夲 and □魚本 should be separated

Comments: The ad hoc noted that 夲 and 本 are known unifiable components that are documented by the IRG as UCV (*Unifiable Component Variations*) #126 (see IRG N2387), which therefore means that U+4C41 鱂 should not be disunified unless the two forms are non-cognate, which they are not per L2/20-094. Furthermore, U+4C41 鱂 serves as an example unified ideograph in UCV #126 that is shown below:

Recommendations: The ad hoc recommends that the UTC repond to the proposal author to point out that 夲 and 本 are considered unified components per IRG's UCV #126, and that U+4C41 鱂 shall therefore not be disunified.

L2/20-097: Proposed kMandarin Values

Comments: This document proposes to add *kMandarin* property values to one CJK Compatibility Ideograph and 35 CJK Unified Ideographs. No issues were found in this proposal.

Recommendations: The ad hoc recommends that the UTC assign an Action Item to CLDR to check the proposed *kMandarin* property values then report back to the UTC.

L2/20-111 (aka WG2 N5129): Proposal of China's horizontal extension for technical used characters

Comments: The WG2 Convenor, Michel Suignard, requested that this WG2 document be discussed during the ad hoc meeting. The ad hoc didn't find any issues with this particular proposal, other than the declared *kRSUnicode* property values for U+943D 鐽 and U+28B4E 𨭎, 167.12 and 167.13, respectively, being incorrect. Their current *kRSUnicode* values, 167.13 and 167.12, respectively, as reflected in the Unihan Database of Unicode Version 13.0 are correct. This minor deficiency doesn't affect the acceptance of this proposal, because providing property values for existing ideographs was not necessary.

Recommendations: The ad hoc recommends that the UTC accept this horizontal extension, and assign an Action Item to John Jenkins and Michel Suignard to update the Unihan Database to change three existing *kIRG_GSource* property values and add 20 new *kIRG_GSource* property values per L2/20-111, for Unicode Version 14.0. See *Unihan-additions-20200410.txt* and *Unihan-changes-20200410.txt* (PDF attachments).
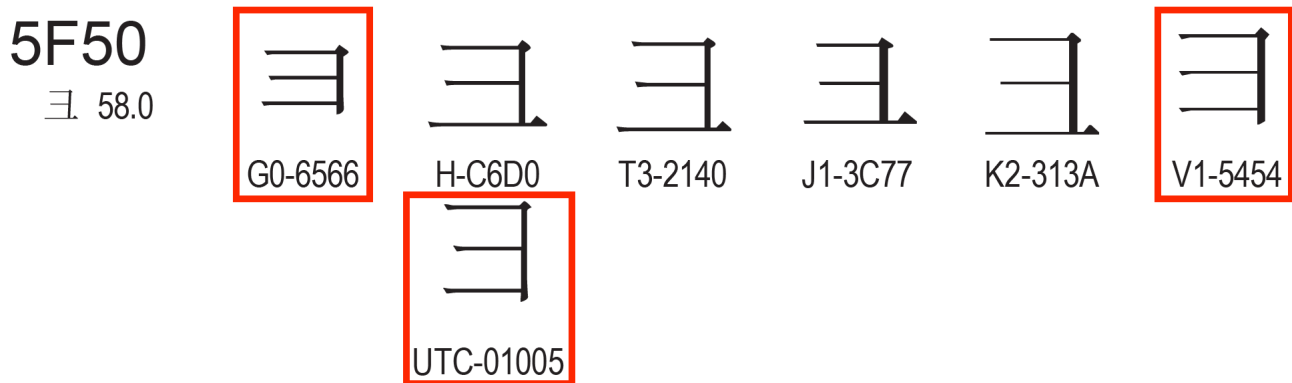
L2/20-112: Proposal to Disunify U+5F50

The following are three related/background documents:

L2/20-098: Proposal to De-Unify One Obsolete Simplified Chinese Character

L2/20-099: Discussion on IRGN2414 among Woo Chi Chung, William Nelson and John Jenkins on Slack

L2/20-100: IRGN2414 Feedback From China

Comments: This document proposes to disunify U+5F50 彐 by requesting that China and Vietnam change its representative glyphs for U+5F50 彐, and to encode at the end of the URO the disunified form, 彐, with G-, V-, and UTC-Source source references.

While there was overall agreement in the ad hoc that U+5F50 彐 deserves to be disunified, it was not clear whether it should be disunified into two (彐 and 彐/彐 as a unifiable pair) or three (彐, 彐, and 彐) ideographs. William Nelson expressed interest in working with the proposal author to prepare a revised proposal.

Recommendations: The ad hoc recommends that the UTC provide the ad hoc comments to the proposal author, and further recommend that he should consider collaborating with William Nelson to prepare a revised proposal that takes into account the ad hoc comments.

5) NamesList.txt Feedback

L2/20-079: Updates for 13 pieces of code charts information in CJK Compatibility Ideographs Supplement block

Comments: This document proposes to add related CJK Unified Ideographs to 13 CJK Compatibility Ideographs. No issues were found in this proposal.

Recommendations: The ad hoc recommends that the UTC accept this proposal, and assign an Action Item to Ken Whistler to update *NamesList.txt* to add the related CJK Unified Ideographs to the 13 CJK Compatibility ideographs per L2/20-079, for Unicode Version 14.0. See *NamesList-additions-20200410.txt* (PDF attachment).

6) IRG Working Set 2020 Submission Deadline

Comments: The IRG Convenor, Dr. Lu, indicated earlier this year that the deadline for member bodies to submit ideographs for IRG Working Set 2020 is 2020-07-17. However, in light of the worldwide COVID-19 situation, some IRG experts have expressed an inability to access the necessary references in their offices or libraries due to regional restrictions. In light of this fact, along with the necessity to cancel IRG #54 altogether, it seems prudent that the IRG Working Set 2020 submission deadline be postponed until at least early 2021.

Experts are in the process of assembling the repertoire of nearly 150 UTC-Source ideographs that are expected to be included in the UTC's submission for IRG Working Set 2020.

Recommendations: The ad hoc recommends that the UTC recommend to the IRG Convenor that the deadline for IRG Working Set 2020 submissions be postponed for at least six months in light of the worldwide COVID-19 situation, and that the new working set be redesignated as IRG Working Set 2021. Full attention should therefore be given to finalizing IRG Working Set 2017 for possible inclusion in Unicode Version 14.0 (September 2021).

7) Other Unihan-related Documents

L2/20-058: Proposal to change the code chart font for 21 blocks in Unicode Version 14.0

Comments: The ad hoc was fully in favor of this particular proposal, and understands its benefits as outlined in the proposal, such as design consistency and streamlining the number of fonts used for code chart production. Alexander Zapryagaev, who provided feedback to the ad hoc prior to the meeting, suggested that four characters in the "Ideographic Symbols and Punctuation" block—U+16FE2 OLD CHINESE HOOK MARK, U+16FE3 OLD CHINESE ITERATION MARK, U+16FF0 VIETNAMESE ALTERNATE READING MARK CA, and U+16FF1 VIETNAMESE ALTERNATE READING MARK NHAY—be included with the scope of this proposal. While it may make sense to eventually cover the characters in this block, especially as it grows, there appears to be no compelling reason to do so at this time.

Recommendations: The ad hoc recommends that the UTC accept this proposal, and assign an Action Item to the code chart editors to integrate into the code chart production process the font provided in L2/20-058, for Unicode Version 14.0.

L2/20-080: Proposal to remove the UCS2003 representative glyphs from the Extension B code charts—Redux

Comments: The ad hoc was fully in favor of this particular proposal, and understands its benefits as outlined in its pages. The presence of the UCS2003 representative glyphs in the CJK Unified Ideographs Extension B code chart has long been a source of confusion for developers and experts alike.

Recommendations: The ad hoc recommends that the UTC accept this proposal, and assign an Action Item to the code chart editors to remove the UCS2003 representative glyphs from the CJK Unified Ideographs Extension B code chart, and to produce an archival CJK Unified Ideographs Extension B code chart that includes only the UCS2003 representative glyphs, for Unicode Version 14.0.

That is all.