Title: CJK & Unihan Working Group Recommendations for UTC #180 Meeting
Author: Ken Lunde
Date: 2024-07-11

This document provides the CJK & Unihan Working Group recommendations for the UTC #180 meeting, based on a meeting that took place on 2024-07-05 from 6 to 8:30PM PDT, which was attended by Yi Bai, Lee Collins, Peter Constable, Kushim Jiang, Ken Lunde, William Nelson, CheonHyeong Sim, Stephan Hyeonjun Stiller, Harald Tveiten, Xieyang Wang, and Yifán Wáng via Zoom. The Chair of the CJK & Unihan Working Group, Ken Lunde, chaired the meeting.

The CJK & Unihan Working Group reviewed public feedback and documents that were received since the UTC #179 meeting took place. Comments are marked in **green**, and Recommendations to the UTC are marked in **red**. A "plain text" Copy&Paste-ready version of the Recommendations is also provided as the PDF attachment named *recommendations-20240705.txt*.

PRI #483 Feedback

The five items of PRI #483 (*Proposed Update UAX #38*) feedback that are in this section were reviewed and discussed by the WG, and for convenience, the following data files, which are PDF attachments, include the recommended Unihan database additions, changes, and removals based on them:

- *Unihan-additions-20240705.txt*
- *Unihan-changes-20240705.txt*
- *Unihan-removals-20240705.txt*
- *kMandarin-additions-20240705.txt*
- *kMandarin-changes-20240705.txt*

With the exception of the proposed *kMandarin* property value additions and changes, the recommended Unihan database additions, changes, and removals are also shown inline, following the **Recommendations**.

01) ID20240419201558

```
Date/Time: Fri Apr 19 20:15:58 CDT 2024
ReportID: ID20240419201558
Name: Eiso Chan
Report Type: Public Review Issue
Opt Subject: 483

In April 19th, 2024 (甲辰龙年谷雨), Tencent Sogou released one video on U+3236D
鸫 (WS2017-04919). Please see https://m.weibo.cn/status/5024879044727275
https://www.xiaohongshu.com/discovery/item/66220b820000000001032dd8 . There
are also two follow-up reports. Please see https://mp.weixin.qq.com/s/
f7jD0OQqR0YhTW2_z5EOsA https://mp.weixin.qq.com/s/hQrZJLkMtQaPxDyqKqEKrg .
```

Those materials and the original evidence provided by UK all show the Putonghua
reading is tún. And the original evidence for the corresponding traditional
form U+3235B 鶉 (WS2017-04893) shows the fanqie is 徒軍切 which matched the
Putonghua reading.

I request to add the following properties values.

U+3235B kFanqie 徒軍

U+3235B kMandarin tún

U+3236D kMandarin tún

Thanks for Mr. Feng Lei's helps, and I have posted the links to ORT.

Comments: The WG reviewed and discussed this feedback, and noted that U+3235B already has 徒軍 as its *kFanqie* property value per the Unicode Version 16.0 Beta review, so no action is necessary. The proposed *kMandarin* property value of U+3235B will need to be checked by the CLDR-TC prior to the UTC accepting it. The proposed *kMandarin* property value of U+3236D conflicts with its value per the Unicode Version 16.0 Beta review. The background is that document L2/22-084 by the same author proposed the same *kMandarin* property value for this ideograph, which the CLDR-TC vetted per UTC Action Item 171-A44. The result of CLDR-TC's vetting is reflected on page 20 of document L2/24-017 in which they disagreed, and recommended that *zhūn* should be used instead, which is present in the Unihan database that is aligned with the Unicode Version 16.0 Beta review. Based on the presented evidence in this feedback, the WG feels that this *kMandarin* property value should be re-checked by the CLDR-TC.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Sharon Chen, CJK: Ask the CLDR-TC to check the proposed kMandarin property value addition and proposed kMandarin property value change, then report back to the UTC. See PRI #483 feedback [ID20240419201558] and Section 01 of document L2/24-165.

02) ID20240528204850

Date/Time: Tue May 28 20:48:50 CDT 2024
ReportID: ID20240528204850
Name: Eiso Chan
Report Type: Public Review Issue
Opt Subject: **483**

IRGN2691 shows U+2E146 㝆 is also used for the names of one series of musical
instruments in Yunnan Province. Please the footnote of p.6 of IRGN2691A.

I request to add the kMandarin property value of U+2E146 㝆 as bǐ as IRGN2691
shows.

Comments: The WG reviewed and discussed this feedback, and agreed that this would be an appropriate *kMandarin* property value addition. As usual, this *kMandarin* property value addition will need to be checked by the CLDR-TC prior to the UTC accepting it.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Sharon Chen, CJK: Ask the CLDR-TC to check the proposed kMandarin property value addition, then report back to the UTC. See PRI #483 feedback [ID20240528204850] and Section 02 of document L2/24-165.

03) ID20240608195107

```
Date/Time: Sat Jun 08 19:51:07 CDT 2024
ReportID: ID20240608195107
Name: Paul Masson
Report Type: Error Report
Opt Subject: kMandarin for U+4680 舦

The Mandarin reading for this character is not currently in the database.
Presumably it should be jiàn as the simplified form of U+8266 艦.
```

Comments: The WG reviewed and discussed this feedback, and agreed that this would be an appropriate *kMandarin* property value addition. The WG noted that *chèn* might be a better *kMandarin* property value than the one proposed in this feedback, but as usual, this *kMandarin* property value addition will need to be checked by the CLDR-TC prior to the UTC accepting it.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Sharon Chen, CJK: Ask the CLDR-TC to check the proposed kMandarin property value addition, then report back to the UTC. See PRI #483 feedback [ID20240608195107] and Section 03 of document L2/24-165.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Working Group comments in Section 03 of document L2/24-165. See PRI #483 feedback [ID20240608195107].

04) ID20240613044832

```
Date/Time: Thu Jun 13 04:48:32 CDT 2024

ReportID: ID20240613044832

Name: oldherl

Report Type: Error Report

Opt Subject: kDefinition.txt in Unihan database


I found the Unihan kDefinition for character U+9527 锧 is "tungsten (element
74, W)".

But I believe it doesn't have such a meaning in Chinese. The character for
tungsten in Chinese is 钨 U+94A8.

Is it an error? Or is it a meaning in other languages?


Thanks.
```

Comments: The WG reviewed and discussed this feedback, and agreed that the *kDefinition* property value of U+9527 锧 is incorrect, and that the *kDefinition* property value of its traditional form, U+9455 鑕 (*chopping block, executioner's block*), would be a more appropriate property value.

It is useful to explore the history of these ideographs' *kDefinition* property values. The *kDefinition* property value of U+9527 锧 was added in Unicode Version 4.1 (2005) as *tungsten, wolfram*, and then changed to its current property value, *tungsten (element 74, W)*, in Unicode Version 15.0 (2022). The *kDefinition* property value of its traditional form, U+9455 鑕, was added in Unicode Version 2.0 (1996) as *tungsten, wolfram*, and then changed to its current property value, *chopping block, executioner's block*, in Unicode Version 15.0 (2022).

In other words, both ideographs shared the same *kDefinition* property value until Unicode Version 15.0 when they diverged. Unfortunately, changes to the provisional *kDefinition* property were made by John Jenkins without documentation in the UTC document register, so the WG suspects that we may never find the reason for their divergence.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Ken Lunde, CJK: Update the Unihan database to change the provisional kDefinition property value of U+9527, based on PRI #483 feedback [ID20240613044832] and as amended in Section 04 of document L2/24-165, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Working Group comments in Section 04 of document L2/24-165. See PRI #483 feedback [ID20240613044832].

Current:
```
U+9527     kDefinition              tungsten (element 74, W)
```
Adjusted:
```
U+9527     kDefinition              chopping block, executioner's block
```

05) ID20240630213656

```
Date/Time: Sun Jun 30 21:36:56 CDT 2024
ReportID: ID20240630213656
Name: Judith Chen
Report Type: Public Review Issue
Opt Subject: 483
```

I recently spent some time checking the kGB3 values in the Unihan database and have found the following issues:

The kGB3 value of U+53D1 发 should not be `1957`. This is obvious because 发 is a simplified hanzi that had been included in GB0 (GB 2312–80); hence, it is impossible to be included in GB3, a traditional hanzi set. In fact, in GB2 (GB 7589–87, the simplified counterpart of GB3), the hanzi placed at 19-57 (in ku-ten form) is U+72AE 犮. Therefore, I suggest moving the kGB3 value `1957` from U+53D1 发 to U+72AE 犮.

```
The kGB3 value of U+5FF9 忹 should not be `4923`. Before Unicode 6.0, U+5FF9 忹
used to have the kIRG_GSource value `3-5137`; however, if we check GB2, we will
find that the hanzi placed at 49-23 is actually U+225D6 忹. The kIRG_GSource
values of these two characters were corrected in Unicode 6.0 — now it is
U+225D6 忹 that has the kIRG_GSource value `G3-5137`. But unfortunately, only
kIRG_GSource values were corrected, so the kGB3 values are still erroneous.
Therefore, I suggest moving the kGB3 value `4923` from U+5FF9 忹 to U+225D6 忹.

To summarise, this feedback suggests the following changes to kGB3 values in
the Unihan database:

– U+53D1 发: 1957 -> (null)

– U+5FF9 忹: 4923 -> (null)

– U+72AE 犮: (null) -> 1957

– U+225D6 忹: (null) -> 4923

That is all.
```

Comments: The WG reviewed and discussed this feedback, and agreed that the proposed changes to the provisional *kGB3* property are reasonable.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Ken Lunde, CJK: Update the Unihan database to remove and add provisional kGB3 property values, based on PRI #483 feedback [ID20240630213656] and Section 05 of document L2/24-165, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Working Group comments in Section 05 of document L2/24-165. See PRI #483 feedback [ID20240630213656].

```
Remove:
U+53D1      kGB3                    1957
U+5FF9      kGB3                    4923
New:
U+72AE      kGB3                    1957
U+225D6     kGB3                    4923
```

PRI #484 Feedback

PRI #484 (*Proposed Update UAX #50*) was re-issued shortly after the UTC #179 meeting, but no feedback was received prior to its closing date of 2024-07-02.

PRI #501 Feedback

PRI #501 (*Proposed Update UAX #11*) was issued shortly after the UTC #179 meeting, but no feedback was received prior to its closing date of 2024-07-02.

PRI #502 Feedback

PRI #502 (*Unicode 16.0 Beta*) was issued shortly after the UTC #179 meeting, but no feedback that required the attention of this WG was received prior to its closing date of 2024-07-02.

Public Feedback

The single item of public feedback from document L2/24-161 that is in this section was reviewed and discussed by the WG, and for convenience, the following data file, which is a PDF attachment, includes the recommended Unihan database change based on it:

- *kMandarin-changes-20240705.txt*

06) ID20240416182127

```
Date/Time: Tue Apr 16 18:21:27 CDT 2024
ReportID: ID20240416182127
Name: Paul Masson
Report Type: Error Report
Opt Subject: kMandarin for U+941F 鐟

This character appears to have an alternate pronunciation tì which is not
reflected in the database.
```

Comments: The WG reviewed and discussed this feedback, and noted that the *kMandarin* property value of U+941F 鐟 was *TI4* (aka *tì*) in Unicode Version 6.0 (2010), and that it changed to *zān* in Unicode Version 6.1 (2012), and has remained unchanged since then. Considering that 12 years have passed, it is therefore reasonable to recommend that the CLDR-TC be tasked to vet the suggestion to change the *kMandarin* property value of U+941F 鐟 back to *tì*, particularly because the CLDR-TC was not involved in the change that took effect in Unicode Version 6.1.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Sharon Chen, CJK: Ask the CLDR-TC to check the proposed kMandarin property value change, then report back to the UTC. See feedback [ID20240416182127] in document L2/24-161 and Section 06 of document L2/24-165.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Working Group comments in Section 06 of document L2/24-165. See feedback [ID20240416182127] in document L2/24-161.

UAX #38 / Unihan Database Documents

The document in this section was reviewed and discussed by the WG, and for convenience, the following data files, which are PDF attachments, include all of the recommended Unihan database additions and changes based on them:

- *Unihan_IRGSources-additions-20240705.txt*
- *Unihan_IRGSources-changes-20240705.txt*

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

07) L2/24-138: Recommendations from WG 2 meeting 71 [WG2 N5254]

Comments: The WG reviewed and discussed this document, which includes two recommendations that are relevant to this WG, specifically Recommendation M71.14 and Recommendation M71.16 on page 4.

Recommendation M71.14 is the establishment of the new "CJK Unified Ideographs Extension J" block, which was developed as IRG Working Set 2021 with 4,300 characters. The new block is U+323B0 through U+3347F, and the code point range is U+323B0 through U+3347B. The group noted that a new "T9" source prefix needs to be added to the syntax and description of the normative *kIRG_TSource* property, and that a new "SATM" source prefix needs to be added to the syntax and description of the normative *kIRG_SSource* property.

Recommendation M71.16 is the disunification of U+5CC0, which involves 1) encoding 山田 separately at U+2B73A at the end of the "CJK Unified Ideographs Extension C" block; 2) moving the T-source (T4-2634) and representative glyph of U+5CC0 to U+2B73A; 3) moving the T-source (TF-2662) and representative glyph of U+2F879 to U+5CC0; 4) adding a new G-source (GKX-1594.81) and representative glyph to U+2B73A; and 5) changing the T-source of U+2F879 to TU-2F879 (no change to its representative glyph).

Both recommendations are targeted for ISO/IEC 10646 7th Edition, and therefore should be targeted for Unicode Version 17.0. Michel Suignard already has the new and updated Unihan database property data for both, along with the necessary fonts.

The source of the "CJK Unified Ideographs Extension J" block is document WG2 N5257 (aka IRG N2678). The sources of the disunification of U+5CC0 are documents WG2 N5259 (aka IRG N2676) and WG2 N5262. See document WG2 N5283 to see code point U+2B73A at the end of the "CJK Unified Ideographs Extension C" block and the new "CJK Unified Ideographs Extension J" block.

The WG notes that document WG2 N5259 raises 1) a concern that IVSes (*Ideographic Variation Sequences*) for U+5CC0 were registered in the 2014-05-16 version of the IVD (*Ideographic Variation Database*) as part of the *Moji_Joho* collection based on a past IRG decision that the ideograph variant for U+5CC0 was unifiable; and 2) a concern that the proposed disunification could be disruptive to existing usage in Japan. See the excerpt from *IVD_Charts_Moji_Joho.pdf* below. The WG agrees with the IRG and WG2 recommendation that the proposed disunification is appropriate, but notes that some users in Japan may continue to represent the newly proposed ideograph at U+2B73A using the IVS <U+5CC0 U+E0100>.

5CC0

E0100
Moji_Joho
MJ010468

E0101
Moji_Joho
MJ068030

Recommendations: The WG recommends to the UTC the following disposition ⇨

Consensus: Accept 4,300 CJK Unified Ideographs in a new CJK Unified Ideographs Extension J block with code point range U+323B0..U+3347B (block: U+323B0..U+3347F), based on document L2/24-138 (aka WG2 N5254) and Section 07 of document L2/24-165, for Unicode Version 17.0.

Consensus: Accept the disunification of U+5CC0, the encoding of its disunified form at U+2B73A at the end of the Extension C block, and all of the associated property value changes and additions, based on document L2/24-138 (aka WG2 N5254) and Section 07 of document L2/24-165, for Unicode Version 17.0.

Consensus: Authorize a proposed update of UAX #38 for Unicode Version 17.0, based on Section 07 of document L2/24-165.

Note: The newly disunified CJK Unified Ideograph U+2B73A was registered in the 2014-05-16 version of the Ideographic Variation Database as the Ideographic Variation Sequence <U+5CC0 U+E0100> in the Moji_Joho collection. Some existing or future documents could represent this ideograph using that IVS.

Action Item for Michel Suignard, CJK: Provide to Ken Lunde the Unihan17.txt (aka ISO/IEC 10646's CJKSrc.txt) data file, based on document L2/24-138 and Section 07 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Lunde, CJK: Use the Unihan17.txt data file to update the Unihan database, based on document L2/24-138 and Section 07 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Lunde, CJK: Prepare a new PRI for a Proposed Update of UAX #38 that adds the CJK Unified Ideographs Extension J block and U+2B73A to Section 4.4 of UAX #38, and documents the new "T9" source prefix for the kIRG_TSource property and the new "SATM" source prefix for the kIRG_SSource property, based on document L2/24-138 and as amended in Section 07 of document L2/24-165, for Unicode Version 17.0.

Action Item for Rick McGowan, UTC: Post the new PRI for the Proposed Update of UAX #38 with a closing date of 2024-10-16.

Action Item for Ken Whistler, UTC: Update the pipeline to add the CJK Unified Ideographs Extension J block and the disunified ideograph with code point U+2B73A, based on document L2/24-138 and Section 07 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Whistler, EDC: Add mention of the CJK Unified Ideographs Extension J block and the range extension of the CJK Unified Ideographs Extension C block to the Alpha

review notes for Unicode Version 17.0, based on document L2/24-138 and Section 07 of document L2/24-165.

Action Item for Ken Whistler, EDC: Document the CJK Unified Ideographs Extension J block and the range extension of the CJK Unified Ideographs Extension C block in the Core Specification, based on document L2/24-138 and Section 07 of document L2/24-165, for Unicode Version 17.0.

```
Current:
U+5CC0     kIRG_TSource     T4-2634
U+2F879    kIRG_TSource     TF-2662
New:
U+2B73A    kIRG_GSource     GKX-1594.81
U+2B73A    kIRG_TSource     T4-2634
Adjusted:
U+5CC0     kIRG_TSource     TF-2662
U+2F879    kIRG_TSource     TU-2F879
```

UAX #45 / U-Source Documents

The two documents in this section were reviewed and discussed by the WG, and for convenience, the following data file, which is a PDF attachment, includes the recommended UAX #45 *USourceData.txt* data file additions based on it:

- *USource-additions-20240705.txt*

The recommended UAX #45 *USourceData.txt* data file additions are also shown inline, immediately following the **Recommendations**.

Completely unrelated to the document in this section, Ken Lunde is planning to incrementally add to a private repository in the Unicode Consortium's GitHub organization the evidence images and their descriptions for any new UAX #45 ideographs. This is to ease the preparation of the UTC's submissions for future IRG working sets. For larger submissions to UAX #45, the proposal submitter will be asked to provide the evidence images and their descriptions after the UTC accepts them and assigns new U-source source references.

08) L2/24-127: Proposal to add one ideograph for a Korean place name to UAX #45

Comments: The WG reviewed and discussed this proposal, and agreed that adding the ideograph to UAX #45 is reasonable, with an assigned status value of **FutureWS**. The ideograph, whose representative glyph is shown below, can therefore be considered as a candidate for inclusion in the UTC's submission for the next IRG working set (post–IRG Working Set 2024):

膓

Recommendations: The WG recommends to the UTC the following disposition ⇨

Consensus: Accept one new U-source ideograph as UTC-03488 with a UAX #45 status value of FutureWS, based on document L2/24-127 (Lee) and Section 08 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Lunde, CJK: Add one new record to the UAX #45 USourceData.txt data file and its representative glyph to USourceGlyphs.pdf, based on document L2/24-127 and Section 08 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Lunde, CJK: Convey to the author of document L2/24-127 the CJK & Unihan Working Group comments in Section 08 of document L2/24-165.

New:
```
UTC-03488;FutureWS;;137.10;;⿰舟离;UTCDoc L2/24-127;kKorean I LI;16;4
```

09) L2/24-179: Proposal to add 23 Daoist characters used in Talismans to UAX #45

Comments: The WG reviewed and discussed this proposal, and agreed that adding the 23 ideographs to UAX #45 is reasonable, with an assigned status value of **FutureWS**. The ideographs, whose representative glyphs are shown below, can therefore be considered as candidates for inclusion in the UTC's submission for the next IRG working set (post–IRG Working Set 2024):

魕 魅 魆 魆 魅 魓 魓 魑 魓 魓 魑 魖
靈 靈 靈 霏 靈 靈 靈 靈 靈 覬 覬

Recommendations: The WG recommends to the UTC the following disposition ⇨

Consensus: Accept 23 new U-source ideographs as UTC-03489 through UTC-03511 with a UAX #45 status value of FutureWS, based on document L2/24-179 (Wang) and Section 09 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Lunde, CJK: Add 23 new records to the UAX #45 USourceData.txt data file and their representative glyphs to USourceGlyphs.pdf, based on document L2/24-179 and Section 09 of document L2/24-165, for Unicode Version 17.0.

New:
```
UTC-03489;FutureWS;;194.29;;⿱⿱⿳鬼鑫;UTCDoc L2/24-179 1;Related to 金, which is
one of the five elements;37;2
UTC-03490;FutureWS;;194.17;;⿱⿱⿳鬼森;UTCDoc L2/24-179 2;Related to 木, which is
one of the five elements;25;2
UTC-03491;FutureWS;;194.25;;⿱⿱⿳鬼⿰淼林;UTCDoc L2/24-179 3;Related to 水, which
is one of the five elements;33;2
UTC-03492;FutureWS;;194.25;;⿱⿱⿳鬼⿰淼炗;UTCDoc L2/24-179 4;Related to 火, which
is one of the five elements;33;2
UTC-03493;FutureWS;;194.14;;⿱⿱⿳鬼垚;UTCDoc L2/24-179 5;Related to 土, which is
one of the five elements;22;2
```

UTC-03494;FutureWS;;194.16;;□⺮□鬼□宫儿;UTCDoc L2/24-179 6;Related to 宫, which is one of the five ancient Chinese tones (五音);26;2
UTC-03495;FutureWS;;194.19;;□⺮□鬼□商上;UTCDoc L2/24-179 7;Related to 商, which is one of the five ancient Chinese tones (五音);29;2
UTC-03496;FutureWS;;194.12;;□⺮□鬼角;UTCDoc L2/24-179 8;Related to 角, which is one of the five ancient Chinese tones (五音);20;2
UTC-03497;FutureWS;;194.23;;□⺮□鬼□山□犭攵火;UTCDoc L2/24-179 9;Related to 徵, which is one of The five ancient Chinese tones (五音):□山□犭攵 is variant of 徵;31;2
UTC-03498;FutureWS;;194.15;;□⺮□鬼□羽旡;UTCDoc L2/24-179 10;Related to 羽, which is one of the five ancient Chinese tones (五音):炁体 is a term used by the Daoist, so 旡 is a reasonable shape;23;2
UTC-03499;FutureWS;;194.18;;□⺮□鬼□商□○丶;UTCDoc L2/24-179 11;Used to cure 耳聾 (deafness) and 耳鳴 (tinnitus);26;2
UTC-03500;FutureWS;;173.14;;□雨□鬼丝;UTCDoc L2/24-179 12;Used to cure diseases about 小腹 (underbelly);22;3
UTC-03501;FutureWS;;173.26;;□雨□鬼□木林□乂乂乂;UTCDoc L2/24-179 13;Used to cure diseases about 臍 (navel);34;3
UTC-03502;FutureWS;;173.16;;□雨□鬼□王灬;UTCDoc L2/24-179 14;Used to cure diseases about 腸 (intestines);24;2
UTC-03503;FutureWS;;173.14;;□雨□鬼劦;UTCDoc L2/24-179 15;Used to cure diseases about 腰 (waist);22;3
UTC-03504;FutureWS;;173.15;;□雨□鬼□工木;UTCDoc L2/24-179 16;Used to cure diseases about 背 (back);23;3
UTC-03505;FutureWS;;173.20;;□雨□鬼□乂乂乂□乂乂乂;UTCDoc L2/24-179 17;Used to cure diseases about 液 (fluid);28;3
UTC-03506;FutureWS;;173.19;;□雨□鬼□彡目⇆彡;UTCDoc L2/24-179 18;Used to cure 淋症 (stranguria);27;3
UTC-03507;FutureWS;;173.18;;□雨□鬼□巠彡;UTCDoc L2/24-179 19;Used to cure 风症 (a concept used by traditional Chinese medical science);26;3
UTC-03508;FutureWS;;173.16;;□雨□鬼□○○二□○○二;UTCDoc L2/24-179 20;Used to cure 寒症 (a concept used by traditional Chinese medical science);24;3
UTC-03509;FutureWS;;194.58;;□尚□鬼鬼鬼□鬼鬼鬼;UTCDoc L2/24-179 21;Used to cure 寒热病 (疟疾, malaria);68;2
UTC-03510;FutureWS;;173.19;;□雨□鬼□日光;UTCDoc L2/24-179 22;Used to a ritual called 開光點眼 (also called 開光點睛, some people believe that this can help them understand the way that the world works and so be wise):日光 means the light of sun, and the character is used to open (開) your left eye;27;3
UTC-03511;FutureWS;;173.19;;□雨□鬼□月光;UTCDoc L2/24-179 23;Used to a ritual called 開光點眼 (also called 開光點睛, some people believe that this can help them understand the way that the world works and so be wise):月光 means the light of moon, and the character is used to open (開) your right eye;27;3

Other Documents

The 10 documents in this section were reviewed and discussed by the WG.

10) L2/23-284: Proposal to encode two small form CJK characters for Chinese

Comments: The WG revisited this proposal, which resulted in two characters, U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER, being provisionally

assigned during the UTC #178 meeting per Consensus 178-C27, so that a note can be recorded in the minutes of the UTC #180 meeting.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Note: The two characters that were provisionally assigned per Consensus 178-C27, U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER, are used like ideographs, but are not assigned Unihan database properties, because they will not be encoded in a CJK Unified Ideographs nor CJK Compatibility Ideographs block. Cross references will be provided in NamesList.txt from these two characters to their corresponding CJK Unified Ideographs, U+513F 儿 and U+5152 兒, respectively.

11) L2/24-057R: Proposal of the EAST ASIAN AUTO SPACING

Comments: The WG briefly discussed this revised proposal, and was pleased to see that its prior feedback has been reflected in it. At this point, this proposal is better suited for review and recommendation by the Properties & Algorithms Working Group, because it is for a new property that is outside the scope of the Unihan database. The only substantive comment that this WG had about this revised proposal is that while Section 2, *Conformance*, on page 2 states that spacing can be inserted manually if overriding is desired, it does not describe how to remove unwanted spacing that could result from the application of this proposed new property.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Ken Lunde, CJK: Convey to the author of document L2/24-057R the CJK & Unihan Working Group comments in Section 11 of document L2/24-165.

12) L2/24-071R3: Proposal to encode three stable extended Suzhou Numeral–like letters for Cantonese Music

Comments: The WG briefly discussed the third revision of this proposal, which reflects the character name changes as discussed after the UTC #179 meeting. The three characters were provisional assigned during the UTC #179 meeting per Consensus 179-C44. The adjusted character names are as follows:

```
U+16FF4 YANGQIN SIGN SLOW ONE BEAT
U+16FF5 YANGQIN SIGN SLOW THREE HALF BEATS
U+16FF6 YANGQIN SIGN SLOW TWO BEATS
```

The Properties & Algorithms Working Group will be including the above character name changes in its recommendations for this UTC meeting, so no recommendation is necessary from this WG.

Recommendations: The WG recommends to the UTC the following disposition ⇨

No UTC action is required.

13) L2/24-124: Proposal to add a HangulSources.txt data file to the UCD

Comments: The WG reviewed and discussed this proposal, which is to add to the UCD a new *HangulSources.txt* data file, and was strongly in favor of adding it for Unicode Version 17.0. The WG noted that some of the mappings are covered in the following directory on the Unicode website:

`https://www.unicode.org/Public/MAPPINGS/OBSOLETE/EASTASIA/KSC/`

However, the data in that directory is incomplete, in terms of coverage of legacy standards, and includes mappings for characters outside the scope of the proposed new UCD data file. In addition, the **OBSOLETE** label is likely to steer people away from their use.

If accepted, the likelihood of changes being made to the *HangulSources.txt* data file in the future is relatively low, and is expected to become the definitive resource used by implementers for interpreting legacy standards that include some or all of the 11,172 characters in the Hangul Syllables block.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Consensus: Accept HangulSources.txt as a new UCD data file, based on document L2/24-124 and Section 13 of document L2/24-165, for Unicode Version 17.0.

Action Item for Markus Scherer, PAG: Add the HangulSources.txt data file to the UCD, based on document L2/24-124 and Section 13 of document L2/24-165, for Unicode Version 17.0.

Action Item for Ken Whistler, EDC: Update UAX #44 to include the HangulSources.txt data file, based on document L2/24-124 and Section 13 of document L2/24-165, for Unicode Version 17.0.

14) L2/24-125: Proposal to add a new CJK Abbreviations block
L2/24-126: Comments on L2/24-125
L2/24-135: Response to L2/24-126 (Comments on L2/24-125)

Comments: The WG reviewed and discussed this proposal, which proposes a new block for encoding several hybrid ideographs that primarily serve as abbreviations. Formal comments were submitted as document L2/24-126 (Sim), and the proposal author responded by submitting document L2/24-135. The proposal author is therefore expected to submit a revised proposal that addresses the concerns expressed in document L2/24-126, along with the WG comments in this section.

The WG agreed with the feedback in document L2/24-126 in that the characters that are based on hand-written simplification or have only one evidence image should be removed. In addition, any character that uses U+3099 COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK should be further discussed, in particular whether they should be encoded with or without the diacritic. The treatment of U+318D7 杮 combined with U+16FF1 〱 VIETNAMESE ALTERNATE READING MARK NHAY to form a single glyph serves as a precedent for doing so, though the use of U+16FF1 in such contexts is considered to be productive. Perhaps the strongest argument to encode such ideographs atomically is that the absence of the U+3099 changes the

overall semantics of the ideograph. In addition, these ideograph-kana hybrid ideographs are effectively one-off occurrences, and are therefore not considered to be as productive as the use of U+16FF1. Handling such ideograph-kana hybrid ideographs as sequences will put an unreasonable burden on font developers, because it will not be apparent that the only meaningful representation of ⿰言キ is with U+3099 to form ⿰言ギ. The WG agrees that input from the Script Encoding Working Group will be instrumental in bringing this proposal to a stage in which its characters can be provisionally assigned.

The WG agreed that a different block name should be used, and is open to suggestions from the UTC. The WG also agreed that the character names should be generalized, such as to describe the components in terms of CJK radical names and other characters. The current names are Japanese-centric.

About the character properties, the WG concluded that the assigned *Script* value for these characters should be **Hani**, because they will be used as though they were Han ideographs. Their assigned *Script_Extensions* values can then be set to either **Hira**, **Kana**, or **Hang**, depending on which script serves as the non-Han components.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Ken Lunde, CJK: Convey to the author of documents L2/24-125 and L2/24-135 the CJK & Unihan Working Group comments in Section 14 of document L2/24-165.

15) L2/24-150: Proposal for missing three Kana-Ligatures

Comments: The WG previously reviewed and discussed the original version of this proposal, document L2/23-112, and was pleased to see the number of ligatures reduced to a reasonable set with sufficient evidence. Feedback from Japan was solicited for the original proposal, but they were neutral on the matter. The WG agreed that the character names and properties looked good, and mirror those of the following two related characters that were encoded in Unicode Version 3.2 (2002):

```
309F;HIRAGANA DIGRAPH YORI;Lo;0;L;<vertical> 3088 308A;;;;N;;;;;
30FF;KATAKANA DIGRAPH KOTO;Lo;0;L;<vertical> 30B3 30C8;;;;N;;;;;
```

The WG therefore recommends provisional assignment with code points U+1B123 through U+1B125 in the Kana Extended-A block. In addition, to follow the precedent of the two related characters cited above, the second and third characters of this proposal should include the following annotation in the UCD's *NamesList.txt* data file:

historically used in vertical contexts, but now found also in horizontal layout

Recommendations: The WG recommends to the UTC the following disposition ⇨

Consensus: Provisionally assign U+1B123 HIRAGANA DIGRAPH KOTO, U+1B124 KATAKANA DIGRAPH TOKI, and U+1B125 KATAKANA DIGRAPH TOTE in the Kana Extended-A block, based on document L2/24-150 (Kojitani) and as amended in Section 15 of document L2/24-165.

Action Item for Ken Whistler, UTC: Update the Pipeline to add the three provisionally assigned characters U+1B123 HIRAGANA DIGRAPH KOTO, U+1B124 KATAKANA DIGRAPH TOKI, and

U+1B125 KATAKANA DIGRAPH TOTE, based on document L2/24-150 and Section 15 of document L2/24-165.

Action Item for Ken Lunde, CJK: Provide a TrueType font for U+1B123 HIRAGANA DIGRAPH KOTO, U+1B124 KATAKANA DIGRAPH TOKI, and U+1B125 KATAKANA DIGRAPH TOTE to Michel Suignard and the Charts Working Group, based on document L2/24-150 and Section 15 of document L2/24-165.

Action Item for Ken Lunde, CJK: Prepare text for the Core Specification that describes U+1B123 HIRAGANA DIGRAPH KOTO, U+1B124 KATAKANA DIGRAPH TOKI, and U+1B125 KATAKANA DIGRAPH TOTE, based on document L2/24-150 and Section 15 of document L2/24-165.

Action Item for Ken Lunde, CJK: Convey to the author of document L2/24-150 the CJK & Unihan Working Group comments in Section 15 of document L2/24-165.

16) L2/24-151: Proposal for two geometric shapes for Japanese traditional calendars

Comments: The WG reviewed and discussed this proposal, and agreed that the Script Encoding Working Group is better suited for reviewing it. The Japan NB should also be contacted to solicit their feedback. No one in the WG was familiar with these symbols. The only substantive feedback from the WG is that their *Vertical_Orientation* (vo) property values should be set to **U** (*Upright*) to match the property values of the other symbols in the set.

Recommendations: The WG recommends to the UTC the following disposition ⇨

Action Item for Ken Lunde, CJK: Convey to the author of document L2/24-151 the CJK & Unihan Working Group comments in Section 16 of document L2/24-165.

Action Item for Ken Lunde, CJK: Contact the Japan NB to solicit feedback on document L2/24-151, based on Section 16 of document L2/24-165.

17) L2/24-155: UTC Submission for IRG Working Set 2024 [IRG N2687]

Comments: This document is the UTC's submission for IRG Working Set 2024 whose submission deadline is 2024-07-26, and is therefore an FYI document for the UTC. The submission consists of 244 UAX #45 (aka U-source) ideographs, and includes a total of 794 evidence images.

Recommendations: The WG recommends to the UTC the following disposition ⇨

No UTC action is required.

Other Topics

18) UTC Action Item 179-A35

Comments: Per UTC Action Item 179-A35, Ken Lunde was tasked to contact the Japan NB to confirm that the J-source representative glyphs for U+2011E, U+2011F, and U+20120 are designed as intended. See Section 14 of document L2/24-067. Ken received the following reply from Shuichi TASHIRO on 2024-06-11:

> *We (JNB) confirmed that we explicitly intended to design the JMJ glyphs for U+2011E, U+2011F, and U+20120 as they are, and there is no need to change them.*

Therefore, no action is necessary.

Recommendations: The WG recommends to the UTC the following disposition ⇨
No UTC action is required.

19) IRG #63 meeting

Comments: Two UNC (*Urgently Needed Character*) proposals with a total of three ideographs are expected to be discussed during the IRG #63 meeting in October, and are therefore expected to be discussed during the UTC #181 meeting in early November. The first UNC proposal, for a single ideograph, was already submitted by China to the IRG as document IRG N2691. The second UNC proposal, for two ideographs, is expected to be submitted by TCA prior to the meeting. If recommended by the IRG, these three ideographs should be targeted for Unicode Version 17.0, and the recommended code points will likely be U+2B73B..U+2B73D at the end of the "CJK Unified Ideographs Extension C" block.

As the newly-appointed IRG Convenor, Ken Lunde started to host a new IRG home page and IRG document register on the Unicode website (the first version went live on 2024-07-08), like Michel Suignard has done for the WG2 home page and document register. For the near future, the new IRG document register will include only new documents, such as those necessary for the IRG #63 meeting and after. Documents from earlier meetings will be added as time and resources permit.

Recommendations: The WG recommends to the UTC the following disposition ⇨
No UTC action is required.

20) Updated Provisional Unihan Database Properties

Comments: The purpose of this section is to provide a synopsis of adjustments that were made to provisional Unihan database properties by Ken Lunde since the UTC #179 meeting, which are shown in the table below:

Property	Additions	Removals	Changes	Notes
kPhonetic	344	20	184	Fine-tuning of property values with stakeholder

Recommendations: The WG recommends to the UTC the following disposition ⇨

No UTC action is required.

That is all.