

Title: CJK & Unihan Group Recommendations for UTC #178 Meeting

Author: Ken Lunde

Date: 2024-01-11

This document provides the CJK & Unihan Group recommendations for the UTC #178 meeting, based on a meeting that took place on 2024-01-05 from 6 to 8PM PST, which was attended by Yi Bai, Eiso Chan, Kushim Jiang, Ken Lunde, William Nelson, CheonHyeong Sim, Stephan Hyeonjun Stiller, and Yifán Wáng via [Zoom](#). The Chair of the [CJK & Unihan Group](#), Ken Lunde, chaired the meeting.

The CJK & Unihan Group reviewed public feedback and documents that were received since the UTC #177 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-20240105.txt*.

GB 18030-2022 Amendment 1

01) L2/23-271: GB 18030-2022 Amendment, Draft 3 + Disposition of Comments, Draft 2

Comments: The third draft of the GB 18030-2022 amendment includes the [CJK Unified Ideographs Extension I](#) block in lieu of encoding a comparable repertoire of ideographs in Plane 10 per the first and second drafts. In addition, the 26 ideographs at the end of the URO (16) and Extension A (10) blocks are no longer within the scope of Implementation Level 1 per the first and second drafts, and are instead within the scope of Implementation Level 3. Also, the five new Ideographic Description Characters in Unicode Version 15.1 are not within any implementation level, which contradicts the expectation that was recorded in [Note 177-N1](#).

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

Note: Contrary to what Note 177-N1 stated, developers whose products must undergo GB 18030 testing and certification need to be aware that the five new Ideographic Description Characters in Unicode Version 15.1, U+2FFC through U+2FFF and U+31EF, are no longer expected to be associated with any implementation level according to the third draft of GB 18030-2022 Amendment 1, based on document L2/23-271 and Section 01 of document L2/24-012.

PRI #483 Feedback

02) ID20231002194206

Date/Time: Mon Oct 02 19:42:06 CDT 2023

ReportID: ID20231002194206

Name: Eiso Chan

Report Type: Public Review Issue

Opt Subject: **Correction for PRI 483, UAX #38**

The current *kCantonese* property value for U+8D5D 𪛐 is *ngan3*, that is very strange. The *kCantonese* property value for U+8D0B 𪛐 and U+8D17 𪛐 are both *ngaan6*. 《广州音字典》 shows the only one Cantonese reading is also *ngaan6* (*ngan*⁶ in that dictionary). *ngan3* is not suitable for U+8D5D 𪛐 everywhere. The Cantonese word *ngan3* means “to shake, to jiggle”, which I mentioned in p. 3 of IRGN2646.

The *kCantonese* property value for U+8D5D 𪛐 should be changed to *ngaan6* from *ngan3*.

Comments: The group reviewed and discussed this feedback, and agreed that *ngan3* is the appropriate *kCantonese* property value for U+8D5D 𪛐. Furthermore, the change to this provisional UniHan database property has already been applied to the UniHan database, so no UTC action is necessary.

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

No UTC action is required.

03) ID20231214115907

Date/Time: Thu Dec 14 11:59:07 CST 2023

ReportID: ID20231214115907

Name: Mark Davis

Report Type: Public Review Issue

Opt Subject: **483**

The Delimiter for *kDefinition* is incorrect; according to the Description it should be semicolon. "Major definitions are separated by semicolons, and minor definitions by commas."

<https://www.unicode.org/reports/tr38/tr38-33.html#kDefinition>

The two changes would be

Delimiter semicolon

Syntax [^\t";]+

Comments: The group reviewed and discussed this feedback, and agreed that *semicolon*, which would be a new Delimiter, is an appropriate Delimiter for the provisional *kDefinition* property. Its current Delimiter is *N/A*. Ken Lunde applied this new Delimiter to the validation

process of the UniHan database, and confirmed that the *kDefinition* property values validate with the updated Syntax. In addition to changing the Delimiter and Syntax of this property in UAX #38, a description of the new Delimiter should be added to Section 4.1, specifically to insert the following sentence after the first sentence of the fourth paragraph:

Properties which allow multiple property values but allow “space” (U+0020 SPACE) to be included, such as the *kDefinition* property, have a *Delimiter* defined as “semicolon” (U+003B ; SEMICOLON).

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

Action Item for Ken Lunde, CJK: Update Proposed Update UAX #38 to change the Delimiter and Syntax of the provisional *kDefinition* property, and to add a description of the new “semicolon” Delimiter to Section 4.1, based on PRI #483 feedback [ID20231214115907] and Section 03 of document L2/24-012, for Unicode Version 16.0.

04) ID20231217153125

Date/Time: Sun Dec 17 15:31:25 CST 2023

ReportID: ID20231217153125

Name: Ken Lunde

Report Type: Public Review Issue

Opt Subject: **483**

I propose that the provisional *kFrequency* property be dropped from the UniHan database for Unicode Version 16.0. This property was introduced in Version 3.2 (2002), covers barely over 5,000 ideographs (approximately 5% of the repertoire), and hasn't changed at all since its introduction. I am a big fan of frequency data, but I am not a fan of including it in such a database, because the moment it is added, it becomes obsolete.

Comments: The group reviewed and discussed this feedback, and agreed that removing the provisional *kFrequency* property is appropriate.

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

Consensus: Remove the provisional *kFrequency* property, based on PRI #483 feedback [ID20231217153125] and Section 04 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Remove the provisional *kFrequency* property from UAX #38, based on PRI #483 feedback [ID20231217153125] and Section 04 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Remove the provisional *kFrequency* property from the UniHan database, based on PRI #483 feedback [ID20231217153125] and Section 04 of document L2/24-012, for Unicode Version 16.0.

PRI #484 Feedback

PRI #484 (*Proposed Update UAX #50*) was issued shortly after the UTC #177 meeting, mainly to add [Section 3.2.4, Quotation Marks](#), but no feedback was received prior to its closing date of 2024-01-02. Although the declared closing date has passed, its closing date should be extended to 2024-04-02 to accommodate feedback to be resolved during the UTC #179 meeting in April of 2024.

Public Feedback

The eight items of public feedback from document [L2/24-008](#) that are in this section were reviewed and discussed by the group, and for convenience, the following data files, which are PDF attachments, include the recommended Unihan database additions, changes, and removals based on them, along with the recommended UAX #45 data file changes:

- *Unihan-additions-20240105.txt*
- *Unihan-changes-20240105.txt*
- *Unihan-removals-20240105.txt*
- *Unihan_IRGSources-additions-20240105.txt*
- *USource-changes-20240105.txt*
- *kMandarin-changes-20240105.txt*

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

05) ID20231106091154

Date/Time: Mon Nov 06 09:11:54 CST 2023

ReportID: ID20231106091154

Name: Junliang Huang

Report Type: Public Review Issue

Opt Subject: **468**

Sorry for commenting on a closed PRI.

UTC-03258 𪛗户己金 is unifiable to U+289D2 𪛗 by UCV #121. 𪛗 is a GHZ character used for person name 朱在𪛗, which is the same person in 南明史. I suggest changing its status from FutureWS to Variant.

UTC-03259 𪛗山𪛗乃 is potentially unifiable to U+23CB8 𪛗.

UTC-03262 𪛗士𪛗工 is likely an error of 壘. The 南明史 evidence mentions that 朱朝𪛗士𪛗工 is 崇禎十三年進士. However, 崇禎十三年庚辰科進士三代履歷 22a and 順治河南通志 17:87 gives 壘 [https://iiif.lib.harvard.edu/manifests/view/drs:48757770\\$714i](https://iiif.lib.harvard.edu/manifests/view/drs:48757770$714i). Given that there are only two stroke differences between 𪛗士𪛗工 and 壘 and 𪛗士𪛗工 is not attested in historical evidence. I suggest changing its status from FutureWS to Variant.

UTC-03264 𪛗火𪛗日 is an error of U+24274 𪛗.

The original 南明史 evidence mentions that 朱睿𤇀火𠂔日, 字翰之, is a painter. His artwork 疏林遠岫 is available on <https://digitalarchive.npm.gov.tw/Painting/Content?pid=5186&Dept=P>, which gives his name 𤇀火𠂔. 𤇀火𠂔 is the preferred form of the current encoded 𤇀火𠂔 (𤇀火𠂔) shape as 𤇀火𠂔 is also attested in 四聲篇海 (成化刊本) 12:18a and his friend 周亮工's work 書畫錄 1:11b. The current encoded 𤇀 is a G4K character used in 文淵閣四庫全書·御定佩文齋書畫譜, also for his name. Anyway, given that there are only one stroke and 𠂔/𠂔 differences between 𤇀火𠂔 and 𤇀. I suggest changing its status from FutureWS to Variant.

UTC-03265 𤇀火𠂔 (𤇀火𠂔) is unifiable to U+3D36 淮 by UCV #304. I suggest changing its status from FutureWS to Variant.

UTC-03267 𤇀金𤇀火𠂔 is a variant of U+28C25 𤇀. 𤇀金𤇀火𠂔 is featured in the table of contents, however, if we turn to page 1464, we will see 南明史 p. 1464 gives 𤇀 instead. So 𤇀金𤇀火𠂔 is a one-off misprint in TOC. I suggest changing its status from FutureWS to Variant.

UTC-03276 𤇀火𠂔 [H6-03]水 is a variant of U+3D57 𤇀. The 南明史 evidence mentions 朱常𤇀火𠂔 [H6-03]水 is 始興縣令. 乾隆始興縣志8:4 and 道光廣東通志29:23 both give U+3D57 𤇀.

UTC-03286 𤇀禾水 is also registered in U+25578 禾 + VS18. I suggest changing its status from FutureWS to Variant.

That's all.

Comments: The group reviewed and discussed this feedback, and agreed that changing the UAX #45 status values from **FutureWS** to **Variant** is appropriate for all eight UAX #45 ideographs, to include adding code points to four of their records.

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

Action Item for Ken Lunde, CJK: Update the UAX #45 data file to change the status values of eight records and to add code points to four of them, based on feedback [ID20231106091154] in document L2/24-008 and Section 05 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Group comments in Section 05 of document L2/24-012. See feedback [ID20231106091154] in document L2/24-008.

Current:

UTC-03258;FutureWS;U+289D2;167.7;;𤇀户己金;UTCDoc L2/22-040 4;;15;4
UTC-03259;FutureWS;U+23CB8;85.5;;𤇀山𤇀乃;UTCDoc L2/22-040 5;;8;4
UTC-03262;FutureWS;;33.6;;𤇀土𤇀工;UTCDoc L2/22-040 8;;9;1
UTC-03264;FutureWS;;86.8;;𤇀火𠂔日;UTCDoc L2/22-040 10;;12;3
UTC-03265;FutureWS;U+3D36;85.11;;𤇀火𠂔佳;UTCDoc L2/22-040 11;;14;4
UTC-03267;FutureWS;;167.17;;𤇀金𤇀火𠂔;UTCDoc L2/22-040 13;;25;4
UTC-03276;FutureWS;;85.9;;𤇀火𠂔 [H6-03]水;UTCDoc L2/22-040 22;;13;4
UTC-03286;FutureWS;U+25778;85.5;;𤇀禾水;UTCDoc L2/22-040 32;;10;3

Adjusted:

UTC-03258;**Variant**;U+289D2;167.7;;𤇀户己金;UTCDoc L2/22-040 4;;15;4
UTC-03259;**Variant**;U+23CB8;85.5;;𤇀山𤇀乃;UTCDoc L2/22-040 5;;8;4
UTC-03262;**Variant**;U+3653;33.6;;𤇀土𤇀工;UTCDoc L2/22-040 8;;9;1
UTC-03264;**Variant**;U+24274;86.8;;𤇀火𠂔日;UTCDoc L2/22-040 10;;12;3
UTC-03265;**Variant**;U+3D36;85.11;;𤇀火𠂔佳;UTCDoc L2/22-040 11;;14;4
UTC-03267;**Variant**;U+28C25;167.17;;𤇀金𤇀火𠂔;UTCDoc L2/22-040 13;;25;4

UTC-03276; **Variant**; U+3D57; 85.9; ; [H6-03] 水; UTCDoc L2/22-040 22; ; 13; 4
UTC-03286; **Variant**; U+25778; 85.5; ; 禾米; UTCDoc L2/22-040 32; ; 10; 3

06) ID20231109191406

Date/Time: Thu Nov 09 19:14:06 CST 2023

ReportID: ID20231109191406

Name: Eiso Chan

Report Type: Public Review Issue

Opt Subject: **kCantonese for U+3150D**

U+3150D 𠵹 is identified as UTC-00420. UAX #45 shows the following information.

UTC-00420; NoAction; ; 30.11; ; [口兜]; kCowles 4114*kMeyerWempe

3029*kCheungBauerIndex 375.08; ; ;

I checked kMeyerWempe 3029 and kCheungBauerIndex 375.08. All of them show the Cantonese reading should be dau3, and it means “cave, den, nest”. We could add the kCantonese property value as below.

U+3150D kCantonese dau3

Current only one source reference is UK-10751, so the UAX #45 could be updated as below if possible.

UTC-00420; ExtH; UTC-00420; 30.11; ; [口兜]; kCowles 4114*kMeyerWempe

3029*kCheungBauerIndex 375.08; ; ;

BTW, the submitted evidence from UK shows the Chinese Hakka-dialect usage.

Comments: The group reviewed and discussed this feedback, and agreed that a circular reference within a UAX #45 record should be avoided, because it doesn't associate the ideograph with the appropriate CJK Unified Ideograph. The group therefore recommends that U+3150D should be horizontally extended to add UTC-00420 as a new U-source source reference and that its UAX #45 record be updated accordingly, which includes adding the missing total strokes (14) and first residual stroke (3) values.

The table below constitutes a proposal to horizontally-extend U+3150D by adding UTC-00420 as a new U-source source reference:

Code Point	Code Chart Excerpt—Version 15.1	U-Source	
		Source Reference	Representative Glyph
U+3150D	3150D 𠵹 30.11 𠵹 UK-10751	UTC-00420	𠵹

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

Consensus: Accept the proposal to add UTC-00420 as a new kIRG_USource property value for U+3150D, based on feedback [ID20231109191406] in document L2/24-008 and as amended in Section 06 of document L2/24-012, for Unicode Version 16.0.

Action Item for Michel Suignard, CJK: Add UTC-00420 as a new kIRG_USource property value for U+3150D, based on feedback [ID20231109191406] in document L2/24-008 and as amended

in Section 06 of document L2/24-012, and provide to Ken Lunde an updated Unihan data file, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the Unihan database to add UTC-00420 as a new kIRG_USource property value for U+3150D and to add a new kCantonese property value to U+3150D, based on feedback [ID20231109191406] in document L2/24-008 and as amended in Section 06 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the UAX #45 data file to reflect the changes to the record for UTC-00420, based on feedback [ID20231109191406] in document L2/24-008 and as amended in Section 06 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the UAX #45 TrueType font to map U+3150D to the glyph for UTC-00420, then provide the updated UAX #45 TrueType font to Michel Suignard and the Charts Group, based on feedback [ID20231109191406] in document L2/24-008 and as amended in Section 06 of document L2/24-012, for Unicode Version 16.0.

New:

U+3150D	kCantonese	dau3
U+3150D	kIRG_USource	UTC-00420

Current—*USourceData.txt*:

UTC-00420;NoAction;;;30.11;;𠄎; kCowles 4114*kMeyerWempe 3029*kCheungBauerIndex 375.08;;;;

Adjusted—*USourceData.txt*:

UTC-00420;ExtH;U+3150D;30.11;;𠄎; kCowles 4114*kMeyerWempe 3029*kCheungBauerIndex 375.08;;14;3

07) ID20231110201043

Date/Time: Fri Nov 10 20:10:43 CST 2023

ReportID: ID20231110201043

Name: Ken Lunde

Report Type: Error Report

Opt Subject: **UAX #45 USourceData.txt**

The following two UAX #45 ideographs are encoded in Extension H:

UTC-00635;Rejected;;;64.3;;𠄎; kCheungBauerIndex 406.01;;;3

UTC-00911;NoAction;;;32.3;;𠄎; Adobe-CNS1 C+17331;;6;

The first record should be changed to the following, which also includes a total stroke count (6):

UTC-00635;ExtH;U+317EA;64.3;;𠄎; kCheungBauerIndex 406.01;;;6;3

The second one is a duplicate of UTC-02997, which is encoded at U+31587, so its record should be changed to the following, which also includes a first residual stroke (1):

UTC-00911;UTC-02997;;;32.3;;𠄎; Adobe-CNS1 C+17331;;;6;1

Comments: The group reviewed and discussed this feedback, and agreed that the records for UTC-00635 and UTC-00911 should be adjusted per the feedback. The group further recommends that U+317EA should be horizontally extended to add UTC-00635 as a new U-source

source reference. The group also noted that *niu1* is an appropriate *kCantonese* property value for U+317EA, which has already been applied to this provisional property in the Unihan database. The table below constitutes a proposal to horizontally-extend U+317EA by adding UTC-00635 as a new U-source source reference:

Code Point	Code Chart Excerpt—Version 15.1	U-Source	
		Source Reference	Representative Glyph
U+317EA	<div>317EA</div> <div>手 64.3</div> <div>𢦏</div> <div>GZ-4502404</div>	UTC-00635	𢦏

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

Consensus: Accept the proposal to add UTC-00635 as a new kIRG_USource property value for U+317EA, based on feedback [ID20231110201043] in document L2/24-008 and as amended in Section 07 of document L2/24-012, for Unicode Version 16.0.

Action Item for Michel Suignard, CJK: Add UTC-00635 as a new kIRG_USource property value for U+317EA, based on feedback [ID20231110201043] in document L2/24-008 and as amended in Section 07 of document L2/24-012, and provide to Ken Lunde an updated Unihan data file, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the Unihan database to add UTC-00635 as a new kIRG_USource property value for U+317EA, based on feedback [ID20231110201043] in document L2/24-008 and as amended in Section 07 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the UAX #45 data file to reflect the changes to the records for UTC-00635 and UTC-00911, based on feedback [ID20231110201043] in document L2/24-008 and as amended in Section 07 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the UAX #45 TrueType font to map U+317EA to the glyph for UTC-00635, then provide the updated UAX #45 TrueType font to Michel Suignard and the Charts Group, based on feedback [ID20231110201043] in document L2/24-008 and as amended in Section 07 of document L2/24-012, for Unicode Version 16.0.

New:

U+317EA kIRG_USource UTC-00635

Current—*USourceData.txt*:

UTC-00635;Rejected;;64.3;;𢦏𢦏小;kCheungBauerIndex 406.01;;;3
 UTC-00911;NoAction;;32.3;;𢦏𢦏土干;Adobe-CNS1 C+17331;;6;

Adjusted—*USourceData.txt*:

UTC-00635;**ExtH;U+317EA**;64.3;;𢦏𢦏小;kCheungBauerIndex 406.01;;**6**;3
 UTC-00911;**UTC-02997;U+31587**;32.3;;𢦏𢦏土干;Adobe-CNS1 C+17331;;6;**1**

08) ID20231112194053

Date/Time: Sun Nov 12 19:40:53 CST 2023

ReportID: ID20231112194053

Name: Paul Masson

Report Type: Error Report

Opt Subject: **kSemanticVariant for U+9452** 鑒

This character already has one semantic variant listed in the database. Another appears to be U+9373 鑒 with one component missing, but the identical pronunciation in both Cantonese and Mandarin. Please update the database accordingly.

Comments: The group reviewed and discussed this feedback, and agreed that a *kSemanticVariant* property value should be added to U+9373 鑒, and that the *kSemanticVariant* property values of U+9451 鑑 and U+9452 鑒 should be changed accordingly. This is a provisional UniHan database property.

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

Action Item for Ken Lunde, CJK: Update the UniHan database to add and change *kSemanticVariant* property values, based on feedback [ID20231112194053] in document L2/24-008 and Section 08 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & UniHan Group comments in Section 08 of document L2/24-012. See feedback [ID20231112194053] in document L2/24-008.

New:

U+9373	kSemanticVariant	U+9451 U+9452
--------	------------------	---------------

Current:

U+9451	kSemanticVariant	U+9452<kHKGlyph,kMatthews,kMeyerWempe
U+9452	kSemanticVariant	U+9451<kHKGlyph,kMatthews,kMeyerWempe

Adjusted:

U+9451	kSemanticVariant	U+9373 U+9452<kHKGlyph,kMatthews,kMeyerWempe
U+9452	kSemanticVariant	U+9373 U+9451<kHKGlyph,kMatthews,kMeyerWempe

09) ID20231118101513

Date/Time: Sat Nov 18 10:15:13 CST 2023

ReportID: ID20231118101513

Name: Andrew West

Report Type: Error Report

Opt Subject: **CJK Unified Ideographs Extension G code chart**

It has recently been reported to me that the glyphs for U+30D91 (UK-02133) and U+30D94 (UK-02134) are not ideal as the 3rd and 4th strokes of the 谷 component on the left side should be joined (as is shown for the G-source characters U+30D93 and U+30D95). I have already supplied Michel with an updated font for

these two characters, and request that the glyphs are updated for the Unicode 16.0 code chart.

Comments: The group reviewed and discussed this feedback, and agreed that the representative glyph changes for the two UK-source ideographs are appropriate. Michel already has the updated font with the adjusted glyphs, so no UTC action other than a consensus is necessary. The table below shows the current representative glyphs as shown in the Unicode Version 15.1 code charts, along with the updated representative glyphs:

Code Point	Code Chart Excerpt—Version 15.1	Source Reference	Updated Representative Glyph
U+30D91	30D91 谷 150.5 𪛇 UK-02133	UK-02133	𪛇
U+30D94	30D94 谷 150.8 𪛈 UK-02134	UK-02134	𪛈

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

Consensus: Accept the proposal to change the representative glyphs for U+30D91 and U+30D94, based on feedback [ID20231118101513] in document L2/24-008 and Section 09 of document L2/24-012, for Unicode Version 16.0.

10) ID20231124222609

Date/Time: Fri Nov 24 22:26:09 CST 2023

ReportID: ID20231124222609

Name: K T Shek

Report Type: Error Report

Opt Subject: **Unihan_Variants.txt**

Unihan_Variants.txt lists U+26552 (𪛇) as a variant of U+8998 (𪛈) and U+4993 (𪛉):

U+4993 kSemanticVariant U+8998<kMeyerWempe U+26552<kMeyerWempe

U+8998 kSemanticVariant U+4993<kMeyerWempe U+26552<kMeyerWempe

U+26552 kSemanticVariant U+4993<kMeyerWempe U+8998<kMeyerWempe

I believe this is not correct. The correct variant of U+8998 (𪛈) should be U+4021 (𪛉) instead of U+26552 (𪛇). The error is probably caused by the very similar printing of the 目 radical and 耳 radical in MeyerWempe (e.g. in P.384-385 the radical 目 in 𪛉 and 𪛇 are very close to a 耳), resulting in the publisher even misinterpreted 𪛉 as 𪛇 (P.40, RSIndex P.91).

𪛉 shares no similar meaning to 𪛈/𪛉/𪛇. And according to Taiwan's variant dictionary, 𪛈, 𪛉, 𪛇 are variants:

https://dict.variants.moe.edu.tw/variants/rbt/word_attribute.rbt?quote_code=QjA0NTg5

But 貼 has no variant.

https://dict.variants.moe.edu.tw/variants/rbt/word_attribute.rbt?quote_code=QzEwNzUz

I understand that it may not be possible to “fix” the reference in kMeyerWempe to U+4021 because it clearly states that the radical is 耳. But as the information provided by the book is in doubt, I still suggest to consider having the kSemanticVariant of U+26552 removed to avoid confusion or spreading of incorrect information. Thank you!

Comments: The group reviewed and discussed this feedback, and agreed that proposed changes to the provisional *kSemanticVariant* property values are appropriate.

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

Action Item for Ken Lunde, CJK: Update the Unihan database to add, change, and remove kSemanticVariant property values, based on feedback [ID20231124222609] in document L2/24-008 and Section 10 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Group comments in Section 10 of document L2/24-012. See feedback [ID20231124222609] in document L2/24-008.

Remove:

U+26552	kSemanticVariant	U+4993<kMeyerWempe	U+8998<kMeyerWempe
---------	------------------	--------------------	--------------------

New:

U+4021	kSemanticVariant	U+4993	U+8998
--------	------------------	--------	--------

Current:

U+4993	kSemanticVariant	U+8998<kMeyerWempe	U+26552<kMeyerWempe
U+8998	kSemanticVariant	U+4993<kMeyerWempe	U+26552<kMeyerWempe

Adjusted:

U+4993	kSemanticVariant	U+4021	U+8998<kMeyerWempe
U+8998	kSemanticVariant	U+4021	U+4993<kMeyerWempe

11) ID20231218134540

Date/Time: Mon Dec 18 13:45:40 CST 2023

ReportID: ID20231218134540

Name: Lee Collins

Report Type: Error Report

Opt Subject: **Unihan_Readings.txt**

Two entries have wrong transliteration from kana くりや

U+5E96	kJapaneseKun	KURYA
--------	--------------	-------

U+5EDA	kJapaneseKun	KURYA
--------	--------------	-------

KURYA should be KURIYA

Comments: The group reviewed and discussed this feedback, and agreed that the current provisional *kJapaneseKun* property values for U+5E96 庖 and U+5EDA 廚 are incorrect. The group also noted that changes to the provisional *kJapaneseKun* and *kJapaneseOn* properties should be avoided or minimized in the future given that the new provisional *kJapanese* property that was introduced in Unicode version 15.1 provides readings in native Japanese scripts—hiragana and katakana—and has three times the coverage among the CJK Unified ideographs blocks. In other words, the provisional *kJapaneseKun* and *kJapaneseOn* properties may eventually be dropped from the UniHan database.

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

Action Item for Ken Lunde, CJK: Update the UniHan database to change the *kJapaneseKun* property values for U+5E96 and U+5EDA, based on feedback [ID20231218134540] in document L2/24-008 and Section 11 of document L2/24-012.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & UniHan Group comments in Section 11 of document L2/24-012. See feedback [ID20231218134540] in document L2/24-008.

Current:

U+5E96	<i>kJapaneseKun</i>	KURYA
U+5EDA	<i>kJapaneseKun</i>	KURYA

Adjusted:

U+5E96	<i>kJapaneseKun</i>	KUR I YA
U+5EDA	<i>kJapaneseKun</i>	KUR I YA

12) ID20240107183744

Date/Time: Sun Jan 07 18:37:44 CST 2024

ReportID: ID20240107183744

Name: Paul Masson

Report Type: Error Report

Opt Subject: **kMandarin for U+9244** 鐵

This character is a semantic variant of U+9435 鐵. As such it should share the same pronunciation for this use, which is tiě, in addition to the existing pronunciation.

Comments: The group reviewed and discussed this feedback, and agreed that the *kMandarin* property value of U+9244 鐵 should be changed from *zhí* to *tiě*. Looking at the history of this property value in the UniHan database is revealing. From Version 2.0 to 6.0, the property value was *TIE3*, which is another way to express *tiě*. The property value was changed to *zhí* in Version 6.1. As usual, this *kMandarin* property value change will need to be checked by the CLDR-TC prior to the UTC accepting them.

Recommendations: The group 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 [ID20240107183744] in document L2/24-008 and Section 12 of document L2/24-012.

Action Item for Ken Lunde, CJK: Convey to the feedback submitter the CJK & Unihan Group comments in Section 12 of document L2/24-012. See feedback [ID20240107183744] in document L2/24-008.

UAX #38 / Unihan Database Documents

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

- *Unihan_IRGSources-additions-20240105.txt*
- *Unihan_IRGSources-removals-20240105.txt*
- *kMandarin-additions-vetted-20240105.txt*
- *kMandarin-changes-vetted-20240105.txt*

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

13) L2/23-211: Proposal to add an RSIndex.txt data file to each release of the Unicode Standard

Comments: This proposal was discussed during the UTC #177 meeting, which resulted in UTC Action Item 177-A26 that tasked Ken Lunde to add a description of the proposed *RSIndex.txt* data file to Section 2.1.2 of UAX #38 via PRI #483 (*Proposed Update UAX #38*). Public feedback as to whether this data file should **not** be added was explicitly solicited, but no negative public feedback was received. The group therefore recommends that this data file be added to each release of the Unicode Standard beginning with Version 16.0.

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

Consensus: Accept the proposal to add the *RSIndex.txt* data file to each release of the Unicode Standard, based on document L2/23-211 (Lunde) and Section 13 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, EDC: Finalize the header text of the *RSIndex.txt* data file, based on document L2/23-211 and Section 13 of document L2/24-012.

Action Item for Ken Whistler, RMG: Update the Unicode Version 16.0 landing page to add link text for the *RSIndex.txt* data file, based on document L2/23-211 and Section 13 of document L2/24-012.

Action Item for Markus Scherer, PAG: Add the *RSIndex.txt* data file to the “charts” directory of Unicode Version 16.0, based on document L2/23-211 and Section 13 of document L2/24-012.

Action Item for Peter Constable, RMG: Add an item to the BRS (Big Red Switch) for Unicode Version 16.0, assigned to Ken Lunde, to update the RSIndex.txt data file, based on document L2/23-211 and Section 13 of document L2/24-012.

Action Item for Peter Constable, RMG: Add an item to the BRS (Big Red Switch) for Unicode Version 16.0, assigned to Markus Scherer, to stage the RSIndex.txt data file, based on document L2/23-211 and Section 13 of document L2/24-012.

14) L2/23-251: Corrections of existing kVietnamese values

Comments: These changes to the provisional *kVietnamese* property have already been applied to the UniHan database, so no UTC action other than to reply to the document author is necessary.

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

Action Item for Ken Lunde, CJK: Convey to the author of document L2/23-251 the CJK & UniHan Group comments in Section 14 of document L2/24-012.

15) L2/23-269: Feedback on year-2022 proposed kMandarin changes/additions for 285 characters (per several UTC actions)

Comments: This document completes the following seven UTC Action Items from 2022 by vetting the recommended informative *kMandarin* property value additions and changes:

170-A76 per L2/22-022 Section 02

170-A92 per L2/22-022 Section 12

171-A31 per L2/22-067 Section 04

172-A44 per L2/22-127 Section 11

173-A21 per L2/22-247 Section 02

173-A32 per L2/22-247 Section 06

173-A52 per L2/22-247 Section 13

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

Consensus: Accept the informative kMandarin property value additions and changes as reviewed and amended by the CLDR-TC, based on document L2/23-269 (Chen) and Section 15 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the UniHan database to add and change the kMandarin property values, based on document L2/23-269 and Section 15 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the author of document L2/23-269 the CJK & UniHan Group comments in Section 15 of document L2/24-012.

Action Item for Ken Lunde, CJK: Convey to the authors of the documents and the submitters of the public feedback, which resulted in the seven UTC Action Items that were completed by document L2/23-269, the CJK & Unihan Group comments in Section 15 of document L2/24-012.

16) L2/24-017: Feedback on year-2022 proposed kMandarin additions for 230 characters (per UTC action 171-A44)

Comments: This document completes the following UTC Action Item from 2022 by vetting the recommended informative *kMandarin* property value additions:

171-A44 per L2/22-067 Section 11

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

Consensus: Accept the informative kMandarin property value additions as reviewed and amended by the CLDR-TC, based on document L2/24-017 (Chen) and Section 16 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the Unihan database to add the kMandarin property values, based on document L2/24-017 and Section 16 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the author of document L2/24-017 the CJK & Unihan Group comments in Section 16 of document L2/24-012.

17) L2/24-016: Disposition of comments on CDAM2.3 to ISO/IEC 10646 6th edition [WG2 N5245]

Comments: Per National Committee comments JP1-007 and JP2-008 on pp 4 and 5 of this disposition-of-comments document, the *kIRG_JSource* property values of U+8FF3 and U+21D12 should be moved to U+2EDE0 and U+21D14, respectively. The group confirmed that these changes are appropriate. These changes correspond to changes to the Moji Jōhō Kiban database, so the appropriate changes to the provisional *kJapanese*, *kMojiJoho*, *kMorohashi* properties have already been applied to the Unihan database.

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

Consensus: Accept the proposal to move the normative *kIRG_JSource* property values of U+8FF3 and U+21D12 to U+2EDE0 and U+21D14, respectively, based on document L2/24-016 (Suignard) and Section 17 of document L2/24-012, for Unicode Version 16.0.

Action Item for Michel Suignard, CJK: Move the two *kIRG_JSource* property values, based on document L2/24-016 and Section 17 of document L2/24-012, and provide to Ken Lunde an updated Unihan data file, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the Unihan database to move the two *kIRG_JSource* property values, based on document L2/24-016 and Section 17 of document L2/24-012, for Unicode Version 16.0.

Remove:

U+8FF3	kIRG_JSource	JMJ-058848
U+21D12	kIRG_JSource	JMJ-034451

New:

U+21D14	kIRG_JSource	JMJ-034451
U+2EDE0	kIRG_JSource	JMJ-058848

UAX #45 / U-Source Documents

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

- *USource-additions-20240105.txt*

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

18) L2/23-285: Proposal to add one character used for one kind of Hakka snack to UAX #45

Comments: The group reviewed and discussed this proposal, and agreed that adding the ideographs 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 IRG Working Set 2024:

糰

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

Consensus: Accept one new U-source ideograph as UTC-03398 with a UAX #45 status value of FutureWS, based on document L2/23-285 (Chan) and Section 18 of document L2/24-012, for Unicode Version 16.0.

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 L2/23-285 and Section 18 of document L2/24-012, for Unicode Version 16.0.

New:

UTC-03398;FutureWS;;119.7;;𪛗;UTCDoc L2/23-285;kMandarin dá;13;1

Other Documents & Topics

19) L2/23-281: Update Suzhou numerals in CJK Symbols font (GitHub issue)

Comments: The group reviewed and discussed this proposal, and agreed that the representative glyphs for U+3024 through U+3029 and U+3038 through U+303A should be adjusted accordingly, which means updating the open source [CJK Symbols](#) font using the glyph data provided as an attachment to [Issue #3](#). The group noted that other glyphs that were mentioned in the issue for which updated glyph data was provide were characterized as cleaning up the subpaths of the glyph, but the glyphs for U+3022 and U+1F260 through U+1F265 also included improvements. In particular, the relative size of the glyphs for the latter range was increased to better fill the em-box. The group recommends that the glyphs for all 19 characters be updated.

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

Consensus: Accept the proposal to change the representative glyphs for U+3021 through U+3029, U+3038 through U+303A, U+303E, and U+1F260 through U+1F265, based on document L2/23-281 (Koo) and as amended in Section 19 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Update the glyphs for U+3021 through U+3029, U+3038 through U+303A, U+303E, and U+1F260 through U+1F265 in the open source CJK Symbols font, then provide the updated CJK Symbols TrueType font to Michel Suignard and the Charts Group, based on document L2/23-281 and as amended in Section 19 of document L2/24-012, for Unicode Version 16.0.

Action Item for Ken Lunde, CJK: Convey to the author of document L2/23-281 the CJK & Unihan Group comments in Section 19 of document L2/24-012.

20) L2/23-283: Proposal to add a property for auto inter-script spacing

Comments: The group reviewed and discussed this proposal, which should have been characterized as preliminary due to the large number of open issues. While the group does not strongly support this proposal, it would not strongly oppose it if the UTC finds merit in it. The group therefore recommends that it be discussed in plenary during the UTC #178 meeting with a larger group of text layout experts.

The group noted that if such a property were to be accepted, it would be sufficiently complex to warrant the drafting of a new Unicode Technical Report (UTR) to adequately document it, along the lines of [UAX #11](#) (*East Asian Width*) for the *East_Asian_Width* property and [UAX #50](#) (*Unicode Vertical Text Layout*) for the *Vertical_Orientation* (vo) property.

The group noted that so-called “house rules” for spacing behavior in text layout can vary, both in terms of the width of the space, and whether a space is used at all. In addition, document-authoring apps tend to provide the most control over the use of spaces in such contexts, but document-displaying apps, such as web browsers, provide the least amount of control over

the use of spaces. Establishing consistent layout behavior is more important for the latter class of apps.

The group also noted that such a property would primarily benefit modern writing systems whose scripts do not use spaces to separate words, such as Chinese and Japanese, whose text may include occasional text strings from writing systems whose scripts *do* use spaces to separate words, such as English. In texts that are predominantly written in a writing system whose scripts use spaces to separate words, explicit spaces are used when including text strings from writing systems whose scripts do not use spaces, such as English text with occasional Chinese or Japanese text strings. It is unclear from the proposal how the proposed property would affect the latter type of texts.

The group further noted that the proposed property would constitute establishing contexts between character classes that affect layout, and those contexts are largely language- and script-dependent. Although the group was unable to come up with any concrete examples, the group strongly suspects that there would arise a non-trivial number of problematic or ambiguous cases when defining the actual property values for all of the characters in the Unicode Standard, many of which may be context-dependent.

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

No UTC action is required.

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

Comments: The group reviewed and discussed this proposal, which is an updated version of a previous proposal from 2016 by the same authors, document [L2/16-109](#), that proposed two Standardized Variation Sequences (SVSes). The revised document proposes that two new characters be added to the Unicode Standard. The group agreed that the usage and semantics of the two proposed characters are different than their full size counterparts, U+513F 儿 (simplified) and U+5152 兒 (traditional), and that the proposed code points and character names, U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER are reasonable. The group therefore recommends that these two characters be provisionally assigned. The group also noted that these characters should be treated as confusables with their full size counterparts in the [CJK Unified Ideographs](#) block, and that the simplified form is confusable with U+3126 儿 BOPOMOFO LETTER ER in the [Bopomofo](#) block.

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

Consensus: Accept the proposal to provisionally assign U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER in the Ideographic Symbols and Punctuation block, based on document L2/23-284 (West & Chan) and Section 21 of document L2/24-012.

Action Item for Ken Whistler, UTC: Update the Pipeline to add the provisionally assigned characters U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER, based on document L2/23-284 and Section 21 of document L2/24-012.

Action Item for Ken Lunde, CJK: Provide a TrueType font for U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER to Michel Suignard and the Charts Group, based on document L2/23-284 and Section 21 of document L2/24-012.

Action Item for Ken Lunde, CJK: Prepare text for the Core Specification that describes U+16FF2 CHINESE SMALL SIMPLIFIED ER and U+16FF3 CHINESE SMALL TRADITIONAL ER, based on document L2/23-284 and Section 21 of document L2/24-012.

Action Item for Ken Lunde, CJK: Convey to the authors of document L2/23-284 the CJK & Unihan Group comments in Section 21 of document L2/24-012.

22) CJK & Unihan Group Vice-Chair

Comments: The CJK & Unihan Group has been without a Vice-Chair since the passing of [John H. Jenkins](#) on 2023-02-28. The Chair, Ken Lunde, nominated Eiso Chan (陈永聪) to serve as the Vice-Chair, and he agreed. The Unicode Consortium subsequently approved the nomination on 2024-01-10. Please join me in welcoming Eiso as the new Vice-Chair of the CJK & Unihan Group.

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

No UTC action is required.

23) “Unihan Tools & Processes” GitHub Repository—Private

Comments: As stated in Section 02 of document [L2/23-163](#), which was the CJK & Unihan Group recommendations for the UTC #176 meeting, Ken Lunde successfully restored the ability to update the Unihan database. In addition, he also updated the version of the Unihan database that runs on the Unicode server via the [Unihan Database Lookup](#) page, and also updated the radical-stroke indexes and all UAX #45 assets. All of the tools and processes for updating the Unihan database, the radical-stroke indexes, and all UAX #45 assets are now fully documented in the private [Unihan Tools & Processes](#) repository on GitHub.

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

No UTC action is required.

24) Updated Provisional Unihan Database Properties

Comments: For those who were not aware, John Jenkins routinely made adjustments to provisional Unihan database properties, which do not require a UTC consensus. 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 #177 meeting, which are shown in the table below:

Property	Additions	Removals	Changes	Notes
kCantonese	1		1	Per ID20231110201043 in Section 07 of this document Per ID20231002194206 in Section 02 of this document
kDefinition			415	Per UTC Action Item 175-A9

Property	Additions	Removals	Changes	Notes
kJapanese	42	41		Per changes to the Moji Jōhō Kiban database
kMojiJoho	41	41	1	Per changes to the Moji Jōhō Kiban database
kMorohashi	40	36	3	Per changes to the Moji Jōhō Kiban database
kPhonetic	23	3,207	3	See the note below
kStrange	15		3	
kVietnamese			166	Per L2/23-251 in Section 14 of this document

The 3,207 *kPhonetic* property value removals were done in consultation with the primary stakeholder of that provisional Unihan database property. The number of ideographs with this property nearly doubled in Unicode Version 14.0 (2021), from 11,463 to 20,037, with virtually all of the property values being asterisked. An asterisked *kPhonetic* property value means that the ideograph does not actually appear in the source of this provisional property, *Ten Thousand Characters: An Analytic Dictionary*. There is no UTC document for these additions, and John Jenkins didn't leave any notes.

In particular, any phonetic group with more than 50 asterisked property values was deemed suspicious, and their asterisked property values were subsequently removed. This affected the asterisked property values for the following 20 phonetic groups (the number of removed property values are in parentheses): 77 (139), 82 (185), 952 (102), 1050 (56), 1051 (58), 1125 (81), 1218 (51), 1339 (90), 1359 (212), 1365 (73), 1426 (508), 1434 (66), 1473 (62), 1501 (269), 1546 (94), 1575 (407), 1605 (426), 1613 (126), 1636 (140), and 1639 (62). Any phonetic group with 50 or fewer asterisked property values will be manually checked as time permits.

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

No UTC action is required.

That is all.