# Universal Multiple-Octet Coded Character Set International Organization for Standardization

Doc Type: Working Group Document

Title: Proposal to add twenty-one CJK Unified Ideographs to the UCS

Source: Ken Lunde, Adobe Systems Incorporated

Status: Member Body Contribution

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2007-09-28

# **Background**

The twenty-one Adobe-Japan1-6 kanji in this proposal represent those characters that we feel cannot be unified with existing CJK Unified Ideographs, either because their forms make it obvious that unification is not possible, or it is not obvious whether unification is possible. One function of this proposal is for it to eventually make its way to the IRG so that multiple experts in this field can make a final determination as to which of these twenty-one kanji should be separate characters, and which should be unified with an existing CJK Unified Ideograph.

The Adobe-Japan1-6 character collection is defined in Adobe Tech Note #5078, available from Adobe Systems. It enumerates 23,058 glyphs, specifically CIDs 0 through 23057.

# **Proposal Details**

This section lists all twenty-one kanji, along with multiple references, to include a representive glyph image and the closest related CJK Unified Ideograph. The representative glyph is from the Adobe-Japan1-6 character collection, and the typeface design is Kozuka Mincho Regular.

If a proposed kanji is determined to fall within the scope of Annex S, in terms of its ability to be unified with an existing CJK Unified Ideograph, then a corresponding IVS (Ideographic Variation Sequence) shall be registered whereby the existing CJK Unified Ideograph with which it is unified shall serve as the Base Character of the IVS.

### **Glyph Images & Parent Forms**

The following table provides the glyph images for the twenty-one Adobe-Japan1-6 kanji that are being proposed, along with the CJK Unified Ideograph that represents their parents forms:

Adobe-Japan1-6 CID	Glyph Image	Parent Forms	Parent Form Glyph Images	Parent Form CIDs
13763	五	U+5DE5	工	1979
13782	座	U+5EA7	座	2098
13866	勢	U+52E2	勢	2638
14174	簑	U+7C11 U+7C14	簑簑	6003 6004

Adobe-Japan1-6 CID	Glyph Image	Parent Forms	Parent Form Glyph Images	Parent Form CIDs
14278	齣	U+9F63	殿句	7460
20088	按	U+5354	協	1696
20096	嘭	U+56C0	囀	14116
20097	囃	U+56C3	囃	4453
20125	溌	U+64A5	撥	5026
20141	压	U+66DC	曜	3893
20149	樅	U+6A05	樅	5283
20153	殱	U+6BB1 U+6BB2	殱 殲	5363 5362
20156	涼	U+6DBC	涼	3979
20176	痉	U+75D9	痙	5746
20180	盈	U+76C8	对	1264
20194	弘	U+8077	職	2540
20204	葉	U+8449	葉	3903
20240	鈯	U+943A	鐺	7041
20247	羪	U+990A	養	3910
20256	魿	U+9C75	頗	19066
20260	鹸	U+9E78 U+9E7C	鹸 鹼	1896 7677

## **Ideographic Description Sequences**

The following table provides the Ideographic Description Sequences (IDSes) for each of these twenty-one Adobe-Japan1-6 kanji, along with each component of the IDSes represented using the "U+xxxx" notation. The first column also provides radical and stroke information in parentheses, in the form of indexing radical number (1 through 214), the number of strokes in the indexing radical instance, and the number of remaining strokes.

Adobe-Japan1-6 CID	IDS	IDS ("U+xxxx" Notation)
13763 (48.3.0)		U+2FF3 U+4E00 U+200D1 U+4E00
13782 (53.3.8)	□广日□□又土	U+2FF8 U+5E7F U+2FF1 U+2FF0 U+53E3 U+53C8 U+571F

Adobe-Japan1-6 CID	IDS	IDS ("U+xxxx" Notation)
13866 (19.2.11)	□□幸丸力	U+2FF1 U+2FF0 U+5E78 U+4E38 U+529B
14174 (118.6.9)	□竹□並衣	U+2FF1 U+7AF9 U+2FF1 U+2000E U+27607
14278 (211.12.5)	□ 歯 句	U+2FF0 U+6B6F U+53E5
20088 (24.2.6)	□十日カ□冫く	U+2FF0 U+5341 U+2FF1 U+529B U+2FF0 U+51AB U+21FE8
20096 (30.3.11)	□口転	U+2FF0 U+53E3 U+8EE2
20097 (30.3.14)	□□雑	U+2FF0 U+53E3 U+96D1
20125 (64.3.9)	□ 扌 発	U+2FF0 U+624C U+767A
20141 (72.4.5; 96.5.4)	田日玉	U+2FF0 U+65E5 U+7389
20149 (75.4.10)	□木従	U+2FF0 U+6728 U+5F93
20153 (78.4.11)	□歹□戈业	U+2FF0 U+6B79 U+2FF9 U+2298F U+4E1A
20156 (85.3.9)	□氵亰	U+2FF0 U+6C35 U+4EB0
20176 (104.5.5)	口疒圣	U+2FF8 U+7592 U+5723
20180 (108.5.3)	□□乃、Ⅲ	U+2FF1 U+2FFB U+4E43 U+4E36 U+76BF
20194 (128.6.2; 28.2.6)	□耳厶	U+2FF0 U+8033 U+53B6
20204 (140.3.9)	日世日世末	U+2FF1 U+8279 U+2FF1 U+4E17 U+6728
20240 (167.8.6)	□ 釒 当	U+2FF0 U+91D2 U+5F53
20247 (123.6.10; 37.3.13)	□美良	U+2FF0 U+7F8E U+826F
20256 (195.11.10)	□ 魚 針	U+2FF0 U+9B5A U+91DD
20260 (197.7.8)	□卤奂	U+2FF0 U+5364 U+3452

#### References

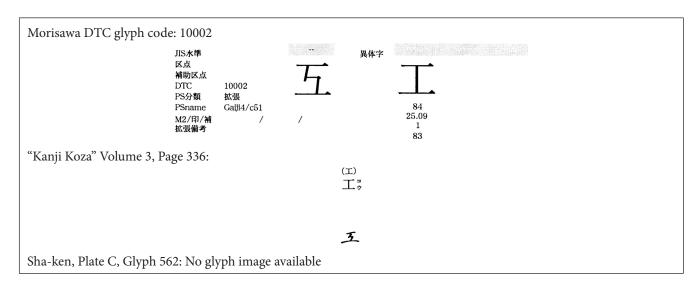
The primary references for these twenty-one characters are the Adobe-Japan1-6 CIDs given in the first column of the first table. The vast majority of Japanese fonts used today are based on the Adobe-Japan1 character collction. The first five characters in this proposal are in Supplement 4 (also known as Adobe-Japan1-4), and the remaining sixteen are in Supplement 5 (Adobe-Japan1-5). Fonts based on these Supplements are considered mainstream and in common use today. In other words, the characters being proposed are considered mainstream, and are expected to be found in most Japanese fonts.

The sections below provide additional references, from the following sources:

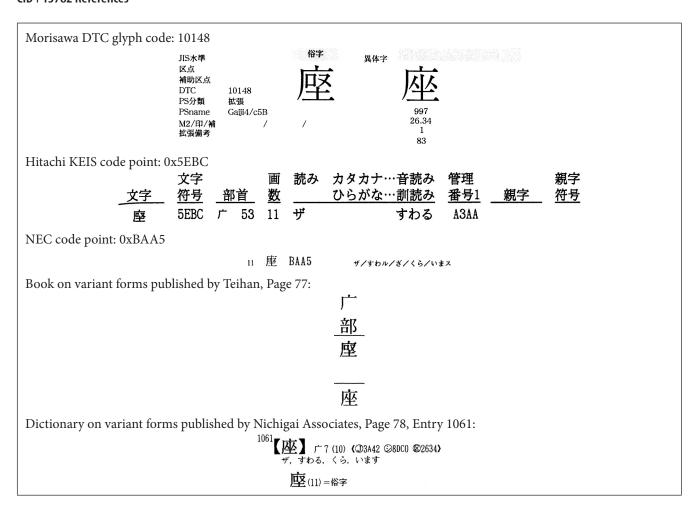
- Morisawa MOR-CODE 2 code points
- Morisawa DTC (Digital Type Center) glyph codes
- Sha-ken glyph codes, from Plates A, B, and C
- NEC code points: 『日本電気標準文字セット辞書 <拡張 >』, ZBB11-1
- Hitachi KEIS code points: 『HITAC 文字コード表 (KEIS83)』, 8080-2-100-10
- Enfour Media code points: Enfour Gaiji DX Series CodeBook
- Asahi Shimbun glyph lists
- "Kanji Koza" Volume 3: 『漢字講座=3 漢字と日本語』, ISBN 4-625-52083-5
- A book on variant forms published by Teihan: 『誤字俗字•正字一覧表』, ISBN 4-924485-29-2
- A dictionary on variant forms published by Nichigai Associates: 『漢字異体字典』, ISBN 4-8169-1249-5
- "Kokuji Dictionary": 『国字の字典』, ISBN 4-490-10279-8

Some of these references include additional information, such as related characters, readings, indexing radicals, and number of total or remaining strokes.

## CID+13763 References

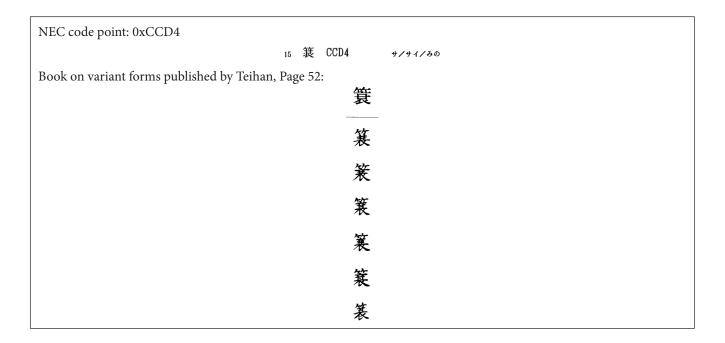


## CID+13782 References

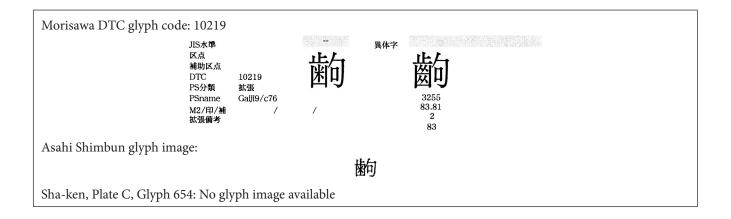


## CID+13866 References

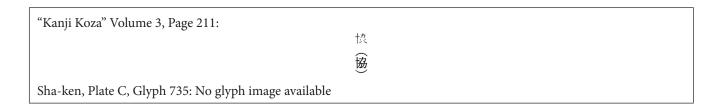
## CID+14174 References



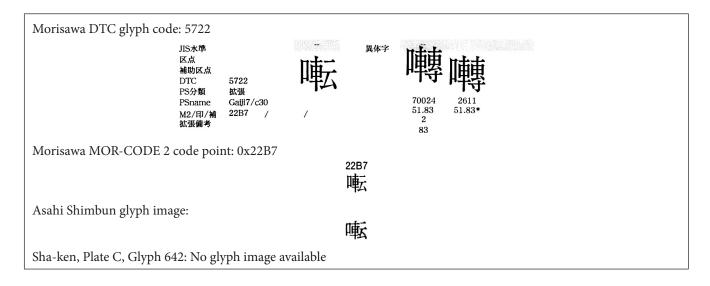
## CID+14278 References



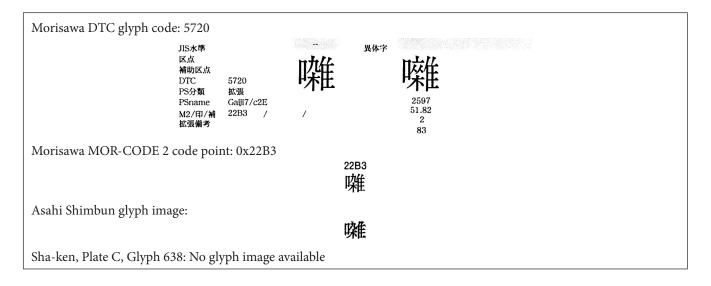
#### CID+20088 References



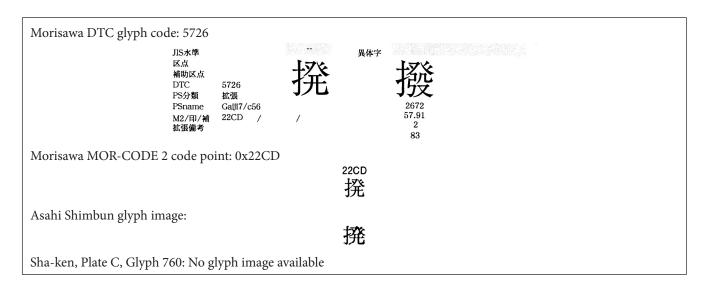
## CID+20096 References



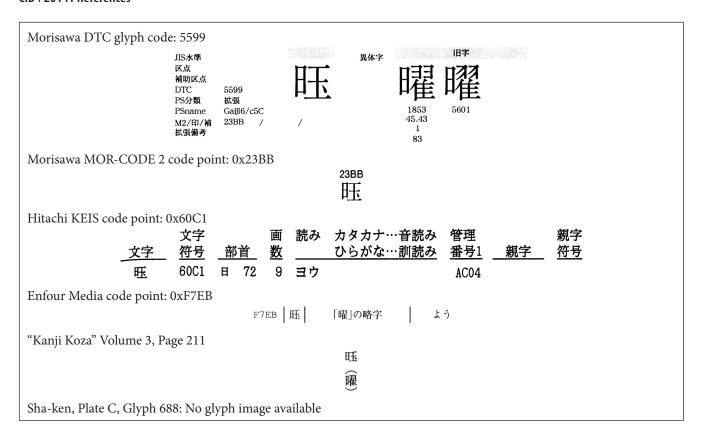
## CID+20097 References



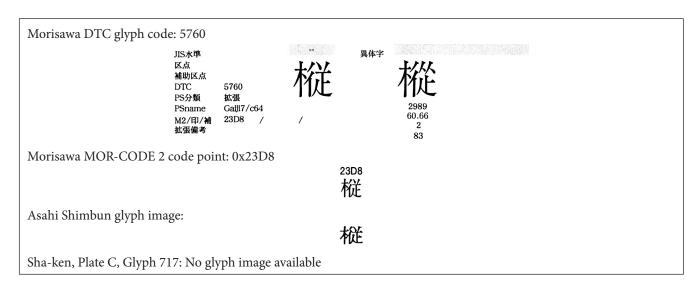
#### CID+20125 References



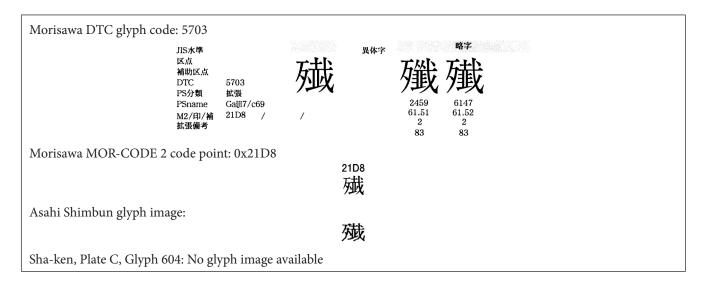
#### CID+20141 References



#### CID+20149 References



#### CID+20153 References



#### CID+20156 References

"Kanji Koza" Volume 3, Page 340	涼,
	涼
Sha-ken, Plate C, Glyph 700: No glyph image available	

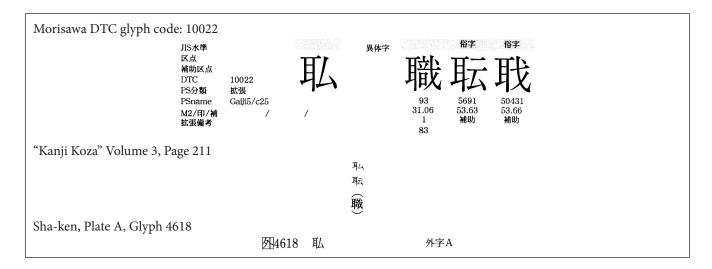
#### CID+20176 References

Morisawa DTC glyph code: 10150 異体字 区点 補助区点 DTC 10150 PS分類 **PSname** Gaiji8/c2A M2/印/補 拡張備考 NEC code point: 0xC5C0 痉 C5C0 ケイ/ひきつル Asahi Shimbun glyph image: 痉 Sha-ken, Plate A, Glyph 5344 图5344 痉 外字A

## CID+20180 References

Hitachi KEIS code point: 0x65D7 文字 画 読み カタカナ…音読み 親字 <u>符号</u> ひらがな…訓読み <u>符号</u> **m** 108 8 エイ みちる A913 盈 B1CE Book on variant forms published by Teihan, Page 51 皿 部 盈 盈 盈 Sha-ken, Plate B, Glyph 7433 图7433 盁 外字B

#### CID+20194 References



## CID+20204 References

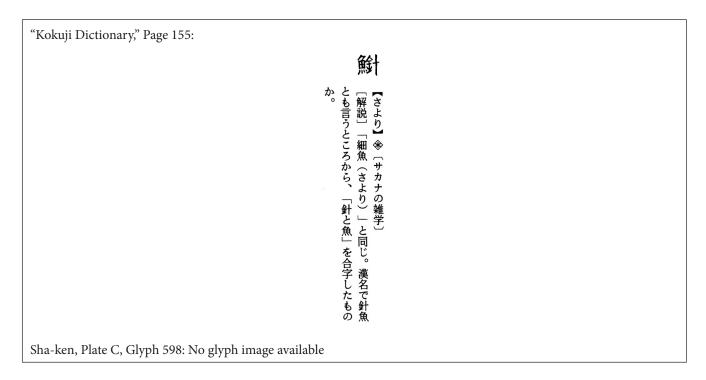
Sha-ken, Plate C, Glyph 672: No glyph image available

#### CID+20240 References

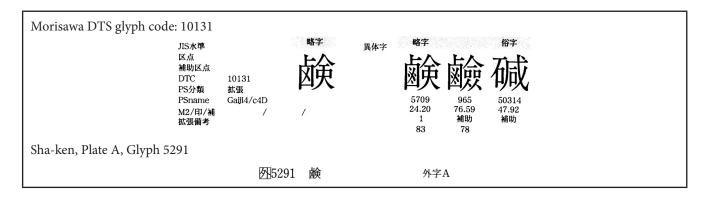
## CID+20247 References

Sha-ken, Plate C, Glyph 578: No glyph image available

#### CID+20256 References



## CID+20260 References



# **Administrative**

- 1. Title: Proposal to add twenty-one CJK Unified Ideographs to the UCS
- 2. Requester's name: Ken Lunde, Adobe Systems Incorporated
- 3. Requester type: Member body
- 4. Submission date: 2007-09-28
- 5. Requester's reference (if applicable): N/A
- 6. Choose one of the following:
- 6a. This is a complete proposal: Yes
- 6b. More information will be provided later: No

# Technical—General

- 1. Choose one of the following:
- 1a. This proposal is for a new script (set of characters): No
- 1b. Proposed name of script: N/A
- 1c. The proposal is for addition of character(s) to an existing block: Yes
- 1d. Name of the existing block: CJK Unified Ideographs
- 2. Number of characters in proposal: 21
- 3. Proposed category: B.1-Specialized (small collection)
- 4a. Is a repertoire including character names provided? Yes
- 4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? N/A
- 4c. Are the character shapes attached in a legible form suitable for review? Yes
- 5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? **Ken Lunde (lunde@adobe.com), Adobe Systems Incorporated**
- 5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: **KozMinPr6N-Regular.otf**, **AFDKO tools**
- 6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? Yes
- 6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? **No**
- 7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? **No**
- 8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at http://www.unicode.org for such information on other scripts. Also see Unicode Character Database http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

## Technical—Justification

- 1. Has this proposal for addition of character(s) been submitted before? If YES, explain. No
- 2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? **Yes**
- 2b. If YES, with whom? John Knightley, Taichi Kawabata & Toshiya Suzuki
- 2c. If YES, available relevant documents: PRI 98 & PRI 108
- 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Anyone who uses Japanese fonts based on the Adobe-Japan1-4 character collection or greater
- 4a. The context of use for the proposed characters (type of use; common or rare): Relatively rare
- 4b. Reference: N/A

- 5a. Are the proposed characters in current use by the user community? Yes
- 5b. If YES, where? Potentially any documents created by Japanese fonts based on the Adobe-Japan1-1 character collection or greater
- 6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? **No**
- 6b. If YES, is a rationale provided? N/A
- 6c. If YES, reference: N/A
- 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? No
- 8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? **No**
- 8b. If YES, is a rationale for its inclusion provided? N/A
- 8c. If YES, reference: N/A
- 9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? **No**
- 9b. If YES, is a rationale for its inclusion provided? N/A
- 9c. If YES, reference: N/A
- 10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? **Yes**
- 10b. If YES, is a rationale for its inclusion provided? Yes
- 10c. If YES, reference: See document
- 11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)? **No**
- 11b. If YES, is a rationale for such use provided? N/A
- 11c. If YES, reference: N/A
- 11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? N/A
- 11e. If YES, reference: N/A
- 12a. Does the proposal contain characters with any special properties such as control function or similar semantics? No
- 12b. If YES, describe in detail (include attachment if necessary): N/A
- 13a. Does the proposal contain any Ideographic compatibility character(s)? No
- 13b. If YES, is the equivalent corresponding unified ideographic character(s) identified? N/A