Title: **Issues related to U+29F3B and JMJ-055359**
Source: **suzuki toshiya**
Date: **2014/11/15**
Status: **Individual Contribution**
Action: **Discussion in IRG#43**

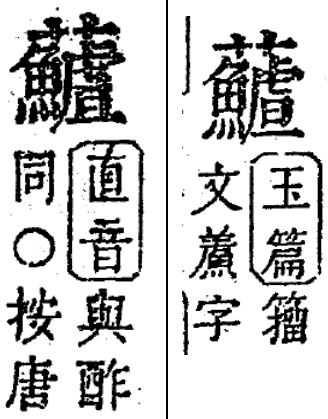
abstract

In IRG N2007, Professor Koichi Yasuoka pointed several similar shaped non-cognate characters submitted by Japan to CJK Unified Ideograph Extension F (see the last page of the document). During the review of F1 set, IRG experts agreed that the shape difference between U+29F3B and JMJ-055359 (F# 06492) is recognizable (旦/且 contrast), so the disunification of them as non-cognate pair is acceptable. However, I found that the source of JMJ-055359 is a Kangxi character, so the disunification might have an impact with existing digital data. I suggest postponing the decision until the evaluation of the impact. Also another aspect by Professor Yasuoka is described in this paper.

1. JMJ-055359 = KX1072.40. KX1072.40 is not coded, or unified with U+29F3B?

Japan national body gave the identity of JMJ-055359 by using Morohashi Daikanwa 32646 (hereafter, I call M32646) under the grass radical, as its evidence. However, we could find the corresponding Kangxi character at KX1072.40 under the grass radical. The source and description are exactly same; M32646 = KX1072.40.

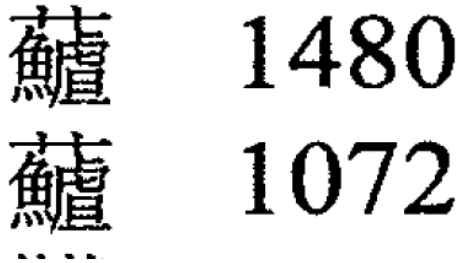
Kangxi		Morohashi Daikanwa (1966-68)	
草部	魚部	草部	魚部
KX1072.40	KX1480.08	【蘧】 32646 ソ 酢(11－39824)に同じ。蘧、同レ酢。	【蘧】 46583 サ 羞(12－46156 篇)蘧、籀文羞。



Morohashi M32646 and its source in Kangxi, KX1072.40

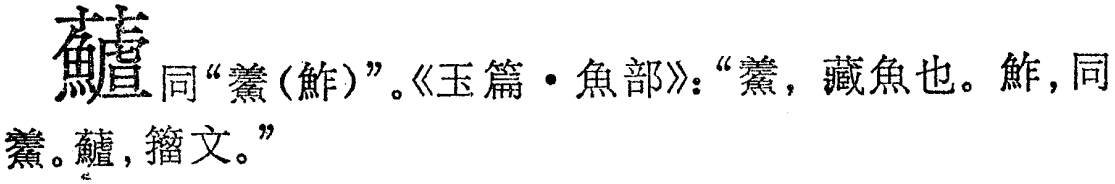
The proposed candidate of the unification of JMJ-055359, U+29F3B, is standardized with the source reference to KX1480.08 under the fish radical (魚部). KX1072.40 versus KX1480.08 shows the 旦/且 contrast in Kanxi, but no character with the explicit source reference to KX1072.40 is coded yet. If China national body worked to encode all Kangxi characters in Ext B, the lack of KX1072.40 could be regarded as it is unified with KX1480.08, at least in China. This hypothesis could be supported by modern dictionaries.

① Kangxi Zidian with Index (康熙字典検索本, 中華書局, ISBN 9787101069747) has a typesetted index. In the typesetted index, the glyph looking like U+29F3B appears twice. One refers KX1072.40, another refers KX1480.08.



康熙字典検索本，総画索引 p. 101 (26 画),「一𠂤」

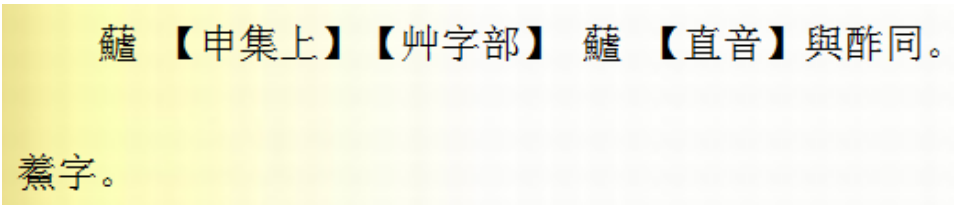
② SuperCJK assigned Hanyu Da Zidian ID to U+29F3B as 53339.120. The description is same with KX1480.08, but the placement of HZ53339.12 is not under the fish (radical of U+29F3B), but under the glass radical (radical of KX1072.40).



HZ53339.12 (漢語大字典, volume 5, p. 3339)

③ In many digitized text of Kangxi Zidian, U+29F3B is used to show the glyph for KX1072.40 too. Furthermore, the descriptions of KX1072.40 and KX1480.08 are concatenated to single entry. If I search the web pages including both of “ ” and “康熙”, many pages make “ ” related with both of “酢” (semantically corresponding to KX1072.40) and “鱈” (semantically corresponding to KX1080.08).

http://ctext.org/dictionary.pl?if=gb&char=%F0%A9%BC%BB



http://www.kyxyz.net/zyk/tsg/kxzd/zw/mydoc060.htm

𩼻字基本信息

𩼻	【基本信息】	拼音： 注音：　部首：魚部
	【简/繁体笔划】	笔划：25　部外笔划：14　繁体部首：魚　笔划：2
	【常用输入法】	五笔86版：AQOG　笔划：1212352512144442153152
	【字库编码】	CJK 统一汉字扩充-B Unicode十六进制：U+29F3B -
	【汉字结构】	
	【造 字 法】	

𩼻字康熙字典查询

康熙字典条目：𩼻

𩼻

康熙字典部首：艸

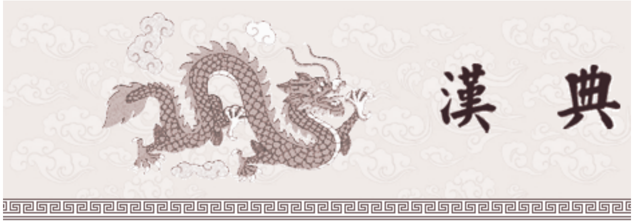
康熙字典笔划：28　部外笔划：22

康熙字典解释：

《直音》與酢同。〇按《唐韻》魚醬曰??，昨誤切，??疑??字之譌，从魚改从缶，於義無取。

《玉篇》籀文耋字。

http://zidian.kxue.com/zi/439167_kangxi.html



基本解释　康熙字典　网友讨论

【申集上】【艸部】 𩼻 ・康熙筆画：28 ・部外筆画：22　【生僻字转文字】

𩼻

《直音》與酢同。〇按《唐韻》魚醬曰䱙，昨誤切，𩼻疑䱙字之譌，从魚改从缶，於義無取。

《玉篇》籀文耋字。

http://www.zdic.net/z/ac/kx/29F3B.htm

Tracking the [illegible] materials of Kangxi (see Appendix) [illegible] U+29F3B [illegible] In summary [illegible] of the impact should be discussed by the member bodies who submitted U+29F3B, at least, China and TCA.

2. Yasuoka's investigation on Morohashi revisions

At present, there is no J-source for U+29F3B. How Japan could discuss whether JMJ-055359 is cognate or non-cognate with the character without J-source? This might be because MJ collection may have a character for U+29F3B but it is not JMJ-055359.

Searching a character for U+29F3B via http://mojikiban.ipa.go.jp/search/SearchBasic.html

MJ055360 could be found. Its corresponding Morohashi is determined as M46583. Therefore, the font released by Moji Joho Kiban project would show MJ055360 as the glyph for U+29F3B. However, MJ055359 and MJ55360 do not show the contrast of KX1072.40 and KX1480.08 (旦/旦).

	MJ文字図形名▲	住基統一	戸籍統一 登記統一	対応するUCS 対応カテゴリー	X0213 X0212	画数	部首	大漢和
鱈	MJ055360	---	527520 00527520	U+29F3B A	--- ---	26	魚	46583

Search result of U+29F3B in Moji Joho Kiban DB

鱈	鱈	鱈	鱈
MJ055360	MJ055359	戸籍統一	戸籍統一
U+29F3B	CJK F1 #06492	527520	371240

Shape difference of MJ055360 (not proposed) and MJ055359 (F1)

Anyway, Koseki Touitsu Moji defined by the Ministry of Justice, the governmental requirement of MJ collection (for detail, please find IRG N1904) does not show the contrast, so MJ collection could not show the contrast either. Why these collections didn't preserve the 旦/旦 contrast? Recently Professor Koichi Yasuoka made an interesting investigation on this. In earlier revision of Morohashi had the contrast of 旦/旦, but the revisions after 1980 lost the contrast.

http://slashdot.jp/journal/586749/%E5%A4%A7%E6%BC%A2%E5%92%8C%E7%95%AA%E5%8F%B746583%E3%81%AE%E5%8F%B3%E4%B8%8B%E3%81%AF%E3%80%8C%E6%97%A6%E3%80%8D%E3%81%AA%E3%81%AE%E3%81%8B%E3%80%8C%E4%B8%94%E3%80%8D%E3%81%AA%E3%81%AE%E3%81%8B

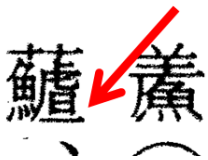
http://slashdot.jp/journal/586898/Re-%E5%A4%A7%E6%BC%A2%E5%92%8C%E7%95%AA%E5%8F%B746583%E3%81%AE%E5%8F%B3%E4%B8%8B%E3%81%AF%E3%80%8C%E6%97%A6%E3%80%8D%E3%81%AA%E3%81%AE%E3%81%8B%E3%80%8C%E4%B8%94%E3%80%8D%E3%81%AA%E3%81%AE%E3%81%8B

Kangxi	Morohashi (1966-1968)	Morohashi 1st revision (1984-1986)	Morohashi 2nd revision (1989-1990)
鱈	【鱈】 32646	【鱈】 32646	【鱈】 32646
鱈	【鱈】 46583	【鱈】 46583	【鱈】 46583

List of representative glyphs for M32646 and M46583 in Morohashi, and Kangxi

Furthermore, although the heading glyph of M46583 in the main content is changed, the character in the description is not changed. The indexing glyph at the indexing page is not changed either.

Indexing page for M46583, in the 2nd revision of Morohashi

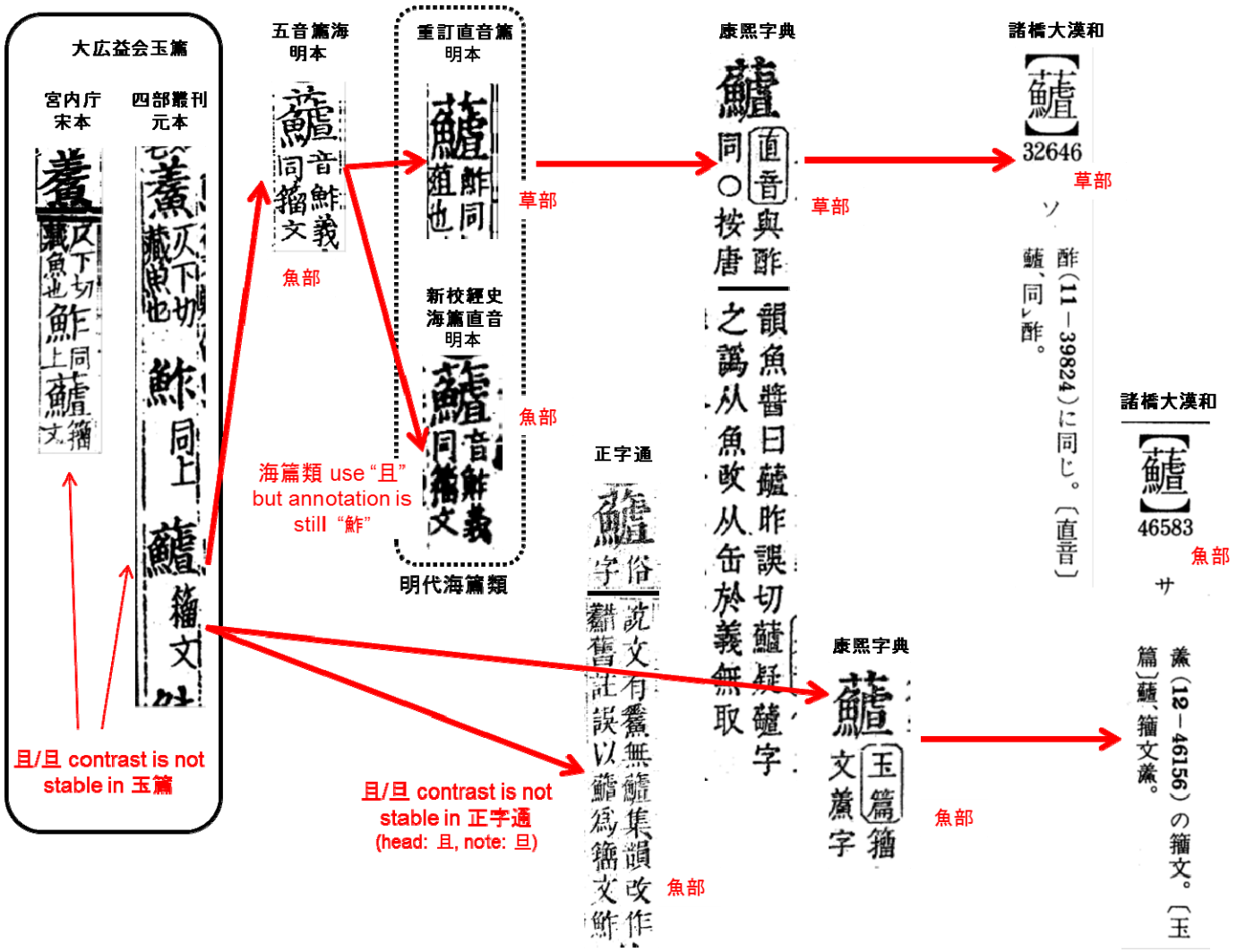


Shape Difference of M46853 in the 2nd revision of Morohashi

In summary, even if Kangxi and the initial version of Morohashi show a shape difference, the distinction of the 2 characters by Japanese governmental sections is not by their shapes. Therefore, even if a consensus to disunify JMJ-055359 from U+29F3B is formed in IRG, the registration of MJ055360 as (one of the) IVSs for for U+29F3B should be carefully discussed. According to WG2 N4625 (Japan submission to add MOJI-JOHO-KIBAN IDEOGRAPHS collection to ISO/IEC 10646 Annex A), the fonts showing MJ055360 for U+29F3B would be quite important for Japanese governmental IT systems.

Appendix KX1072.40 and KX1480.08 are non-cognate?

Although Morohashi and Kangxi descriptions for KX1072.40 (=M32646) and KX1480.08 (=M46583) are different, they are exactly non-cognate? When I check “重訂直音篇” and “新校經史海篇直音” for the character corresponding to KX1072.40, both of them refer “鮓” instead of “酢”. Of course, “酢” and “鮓” are semantically different today (“酢”, “醋” are for vinegar, “鮓” is for salted fish), Kangxi description concerned about the relationship with fish source (魚醬), so such distinction was not critical for Kangxi editors in the discussion of KX1072.40. Considering that “鮓” was related with “ ” (U+29F3B) in 玉篇, the origin of the distinction is how the description in 玉篇 was reduced.



(end of document)