Date 2024−10−10

Source:	CheonHyeong Sim (沈天珩, aka TianHeng Shen)
Title:	Proposal to Disunify two Kangxi Characters **(Redux)**
Status:	Individual Contribution on IRG #63
Action:	To be considered by IRG and multiple sources

Background

In IRGN2537, I proposed to disunify 8 characters, then TCA had a feedback, and after that there was no further discussion. The feedback from TCA pointed out that they agreed to disunify one of those characters, however even that character is now still not disunified. I asked Selena last year and she said that TCA was waiting for the opinions from China. In this proposal, I would like to propose to disunify 2 of those characters again with additional references, and **hope to get feedback from multiple sources**.

1. U+2335F (𣍟)

2. U+6B25 (欥)

U+2335F (𣍟)

U+2335F, along with its compatibility character U+2F980 looks like the below picture in Unicode 16.0:

First of all, let us see the feedback for IRGN2537 from TCA which indicates that they agreed to disunify this character:

> The TCA agrees with the proposal to disunified 2F980 and U+2335F, because they are two characters with different sounds and meanings.
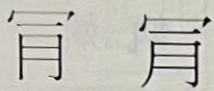> As we can see from the figure below, U+2F980 is the ancient character for "肯", with the sound kěn, the radical is 肉(130.0). In the book 四聲篇海, the radical is 月(74.0), with the sound wěi, which means "吐"(to spit).
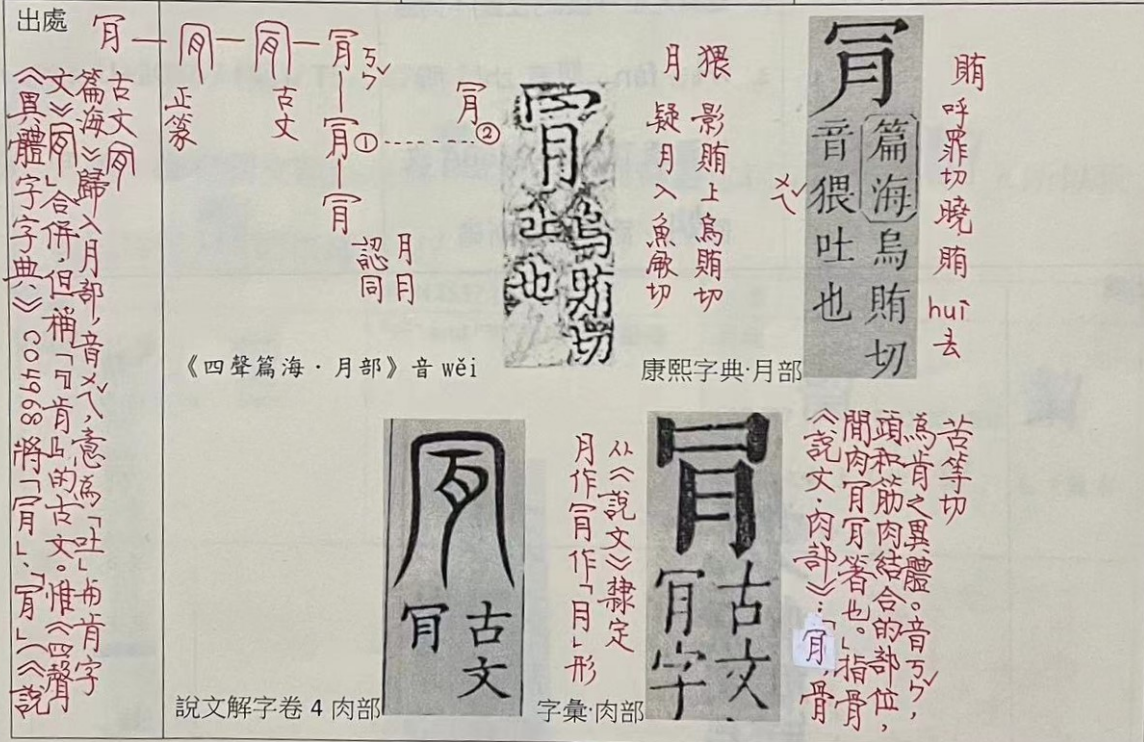> (figure omitted in this document, please see the original feedback)

While Prof. Hsueh Jen Hsu（許學仁）passed away in May, 2023, Dr. Lu sent an email to all the IRG experts, and attached his manuscript − it chanced to be the discussion about whether to disunify this character or not. This picture is shown below:

Group 7: U+2335F (肎) / T5-2444

➢ 是否要解認同?

ucs	IRGN2537 提議	回應
2F980 肎 月 74.3 T6-287B ≡ 2335F 肎	肎、肎應解除認同。月、月分開編碼。	1. 81 年碼本形 肎；96 年碼 本形 肎 2. 肎 (C04968)釋義之一為: 「肯」的古文，下從肉。見 《說文解字·月部》。
2335F 肎　肎 月 74.3 GKX-0504.23　T5-2444		

出處

《四聲篇海·月部》音 wěi

說文解字卷 4 肉部

字彙·肉部

康熙字典·月部

I still remember that, I received the email when I was having a lesson (Japanese oral interpretation), and after I saw that the attached picture was related to my proposal, I almost cried in public.

It is clear that ▨宀一月 (aka the 月部 one) and ▨宀一月 (aka the 肉部 one) are two different characters. The evidences already listed in IRGN2537（康熙字典，汉语大字典，CNS11643，教育部異體字字典 and 文字情報基盤検索システム）will not be repeated here; below is the evidence from *Daikanwa Dictionary*（大漢和辭典）:
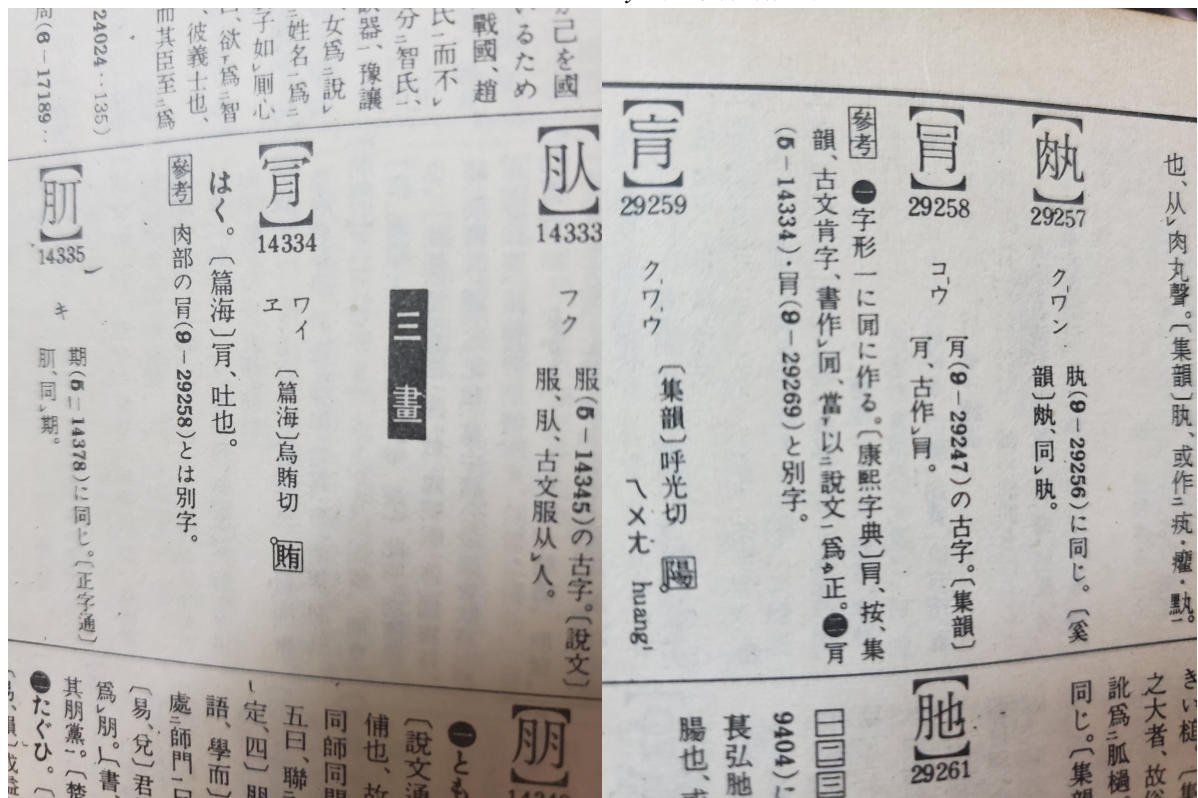

Fig.1−2 大漢和辭典 P5673, P9606

It is also clear that they are two different characters in Japan. After the large horizontal extention by J−source from Unicode 16.0, the former one（aka JMJ−037876）is correctly mapped to U+2335F, and the latter one（aka JMJ−037877）needs to be horizontally extended after the disunification.

For KP−source, there is no action should be taken. KP1−4A38 is just the 月部 one, and the 肉部 one is not seen in KPS 10721.

To sum up, the table below shows the new source references if the disunification is done:

	G−source	T−source	J−source	KP−source
U+2335F	GKX−0504.23	T5−2444	JMJ−037876	KP1−4A38
U+2????	GKX−0973.20	T6−287B	JMJ−037877	/
U+2F980	/	TU−2F980	/	/

In the table, red indicates a new source reference on that codepoint, while blue indicates a revised source reference on that codepoint. U+2???? indicates the new codepoint, maybe at the end of CJK−ExtC.
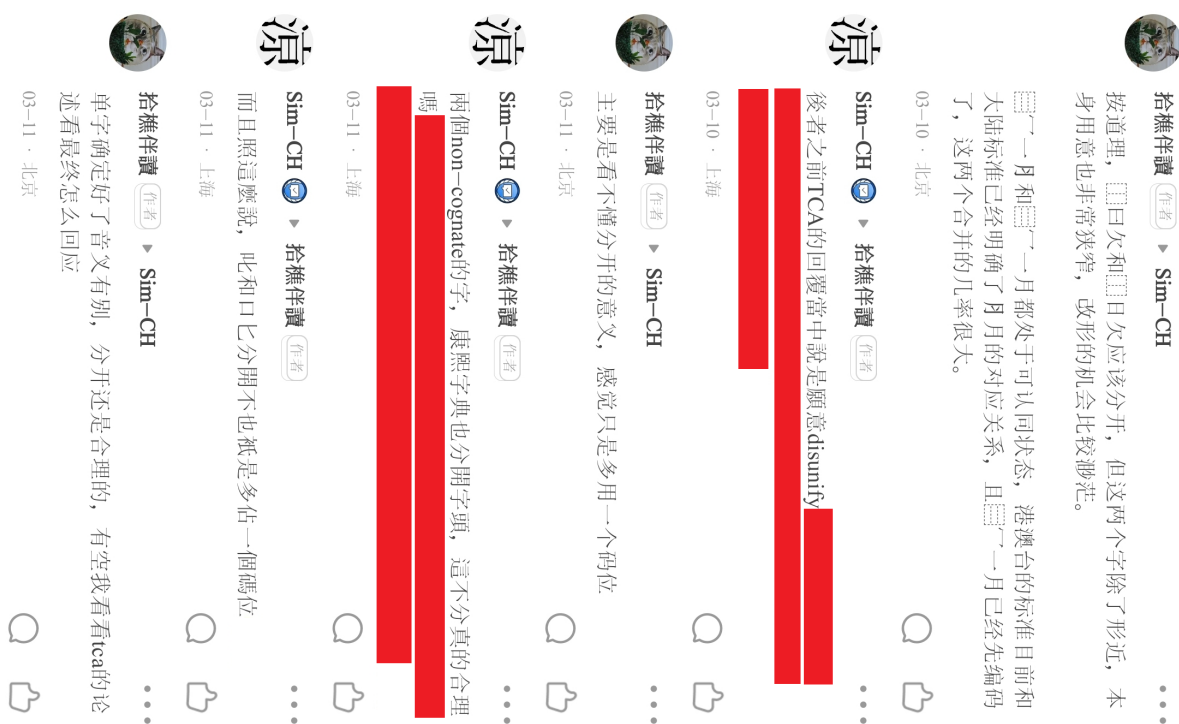
U+6B25（欥）

U+6B25 looks like the below picture in Unicode 16.0:

Let us first see some discussion on Zhihu（知乎, a website like Quora）between Tao Yang（陶扬）and me:

<p align="center">（拾樵伴讀 is Tao Yang, and Sim−CH is me）</p>

He says that,

> 按道理，⯍曰欠和⯍曰欠应该分开，但这两个字除了形近，本身用意也非常狭窄，改形的机会比较渺茫。
>
> In principle, ⯍曰欠 and ⯍曰欠 should be separated, but these two characters are not only similar in shape, but also have a narrow meaning, so the opportunity to change their shapes is relatively slim.
>
> 单字确定好了音义有别，分开还是合理的。

> Once the pronunciation and the meaning of single characters are determined, it is reasonable to separate them.

One thing I need to clarify is that, I do not request (but only suggest) to change the glyph for any sources. There are many different characters (i.e. their abstract shapes are different) with the same actual shape in Unicode CJKUI, even in the recently encoded CJK−ExtH and the upcoming CJK−ExtJ, for examples:

And I totally agree that we should disunify if we could determine that the two characters are phonetically and semantically different, even if they are graphically similar or the same in some of the regional conventions. The evidence already listed in IRGN2537 (康熙字典) will not be repeated here; below is the evidence from *Daikanwa Dictionary* (大漢和辭典):

Fig.3−4 大漢和辭典 P6333, P5434

大漢和辭典 also says that □曰欠 and □日欠 are different characters, which is the same as 康熙字典. However, the two characters are not separated in either the JH (Hanyō Denshi) character set or the JMJ (Moji Jōhō) character set, and the readings are mixed in JMJ's 文字情報基盤検索システム(JMJ−014753). Here I would like to request Japan to separate these two characters − whether to differentiate the glyphs does not matter since there are already precedent cases like JMJ−013840 and JMJ−013841.

The thorny thing here is that, both 康熙字典 and 大漢和辭典 give only the pronunciation of □日欠 and do not provide the meaning. However, □日欠 is also a Nôm character, and it has clear pronunciation and meaning different from □日欠. Even if □日欠 from 康熙字典 or 大漢和辭典 seems to be not cognate with □日欠 from Nôm, they share both the abstract shape and the actual shape, which is a similar case as 机 (jī, means machine in Chinese) and 机 (つくえ, means table in Japanese), so they should be regarded as a same character. When we search for □日欠 in 字典𡨸喃引解 (Tự Điển Chữ Nôm Dẫn Giải), we will see:

Fig.5 online_TĐCNDG_on_nomfoundation.org

We can confirm that the reading sớm corresponds to □日欠 but not □曰欠 because it means "morning" or "early"; also, we may see many other Nôm characters with the reading sớm contain the semantic component 早, 旦 or 日:

Fig.6-7 越汉词典 商务印书馆 P963; Nôm_Lookup_Tool_on_nomfoundation.org

At least we can confirm that 𣅑日欠 appears in 康熙字典, 大漢和辭典 and 字典𤤰喃引解, and it is phonetically and semantically different from 𣅑曰欠. Since 𣅑曰欠 is relatively more commonly used, I suggest to keep 𣅑曰欠 on U+6B25, and encode 𣅑日欠 on a new codepoint.

It looks like that we will do nothing for U+6B25 but only to encode a new character, however, I would like to point out that it is really a disunification request since multiple sources mix the two characters − besides 文字情報基盤検索システム which mixed the readings of two different characters on one single codepoint, in Unihan Database, the kHanyuPinyin value for U+6B25 is "21495.010:yì,huān,yù 32135.090:yù,yì,huān", which also directs to the fact that this codepoint mixed two different characters (𣅑曰欠 corresponds to yì&yù, while 𣅑日欠 corresponds to huān).

To sum up, the table below shows the new source references if the disunification is done:

	G−source	H−source	T−source	J−source	K−source	V−source
U+6B25	G5−4A39	HB2−CCE9	T2−274A	JMJ−014753	K2−3E2D	/
U+2????	GKX−1600.15	/	/	J?????	/	VN−?????

In the table, red indicates a new source reference on that codepoint. U+2???? indicates the new codepoint, maybe at the end of CJK−ExtC. J????? indicates the new J−source if Japan accepts my request to separate the two characters, I am not sure if they would like to use JMJ source or maybe something else (e.g. a new source for 大漢和辭典); VN−????? indicates the new V−source, but I do not know how much (zi.tools says it is VN−F073D but U+F073D in the newest Nôm Na Tống font is not 𣅑日欠).

(End of document)