

Source:	CheonHyeong Sim (沈天珩)
Title:	Proposal to Disunify U+FACB and U+2F9FF
Status:	Individual Contribution on IRG #64 and WG2 #72
Action:	To be considered by China NB, TCA, Japan NB, DPRK and IRG

Background

U+FACB, U+2F9FF and their base character (U+980B) look like the following pictures in the CodeCharts of Unicode 16.0 perspective:

在Unicode16.0的CodeCharts中，U+FACB、U+2F9FF及其基字（U+980B）分別如下所示：

Unicode16.0のCodeChartsでU+FACB、U+2F9FFとそれらの基本文字（U+980B）は別々次のように見える：



Fig.1 U+980B, U+FACB, U+2F9FF in Unicode 16.0

The glyphs of the two compatibility characters look similar to the glyphs from multiple sources in URO, and they are considered as variants to each other. However, actually they are phonetically, graphically and semantically different, should be considered non-cognate, and need to be disunified.

兩個相容表意文字區的字元的字形看上去和基本區中各提交源的字形相似，被認為互為異體字。然而，實際上它們音形義皆不同，應當被視為非同源，需要被解除認同。

2つの日中韓互換漢字ブロックの文字のグリフは、日中韓統合表意文字ブロックでの複数のソースのグリフと似ており、異体字として考えられているが、実は音声的にも、図形的にも、意味的にも異なり、統合しないようにする必要がある。

Also, in the Ideographic Variation Database, VS18 (U+E0101) of U+980B submitted by Hanyō Denshi and Moji Jōhō (aka JMJ) looks similar to U+2F9FF:

此外，在表意文字變體資料庫中，汎用電子和文字情報提交的U+980B的VS18 (U+E0101) 看起來與U+2F9FF相似：

その他、漢字異体字データベースの中で、汎用電子と文字情報の提出したU+980BのVS18 (U+E0101) はU+2F9FFと似ているように見える：

頤 頤 頤 頤
00Ha 00Mo 01Ha 01Mo

Fig.2 IVD of U+980B

Properties of these Characters

Let us see the properties of these characters from each sources one by one. Before doing that, let us see what *Kangxi Dictionary* says first:

我們逐個提交源地看一下這些字元的屬性。在此之前，我們先看一下《康熙字典》怎麼說：

これらの文字的属性をソースずつ見てみよう。その前に、まず『康熙字典』の言い方を見てみよう：

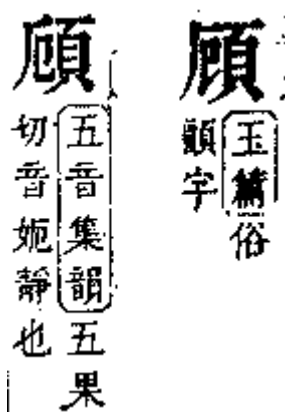


Fig.3-4 康熙字典 同文書局版 [P1400](#) & [P1402](#)

Obviously, 頤 (the one with a component 巳, no matter whether the 1st stroke is 一 or 丿) is a phonogram whose semantic component is 頁 and phonetic component is 厄. Meanwhile, 頤 (the one with a component 巳) is a folk form of 頤.

顯然，「頤」（包含部件「巳」的那一個，無論第一畫是橫是撇）是一個「從頁厄聲」的形聲字，而「頤」（包含部件「巳」的那一個）是「頤」的俗字。

明らかに、「頤」（部品「巳」を含むもの、第1画が横でも左払いでもいい）は「從頁厄^が声」の形声字であり、「頤」（部品「巳」を含むもの）は「頤」の俗字である。

Since 頤 and 頤 are hard to be distinguished quickly in the document with the human eyes, in the following text, I will use 巳頤 and 巳頤 instead.

由於「顧」和「顧」在文檔中很難用人眼快速區分，在後文中我將以「巳顧」和「巳顧」指代。

「顧」と「顧」は文書の中で人の目で速く区別するのが難しいため、次のテキストでは「巳顧」と「巳顧」で指す。

But, how did 顧 become 巳顧? This is the key to proof the non-cognateness of 巳顧 and 巳顧.

可是，「顧」是怎麼變成「巳顧」的呢？這是證明「巳顧」和「巳顧」非同源的關鍵。

しかし、「顧」はどうやって「巳顧」になったのであろうか。これは「巳顧」と「巳顧」が同源ではないことを証明する鍵である。

楷書



行書



草書



Fig.5 [顧 on 字統网](#)

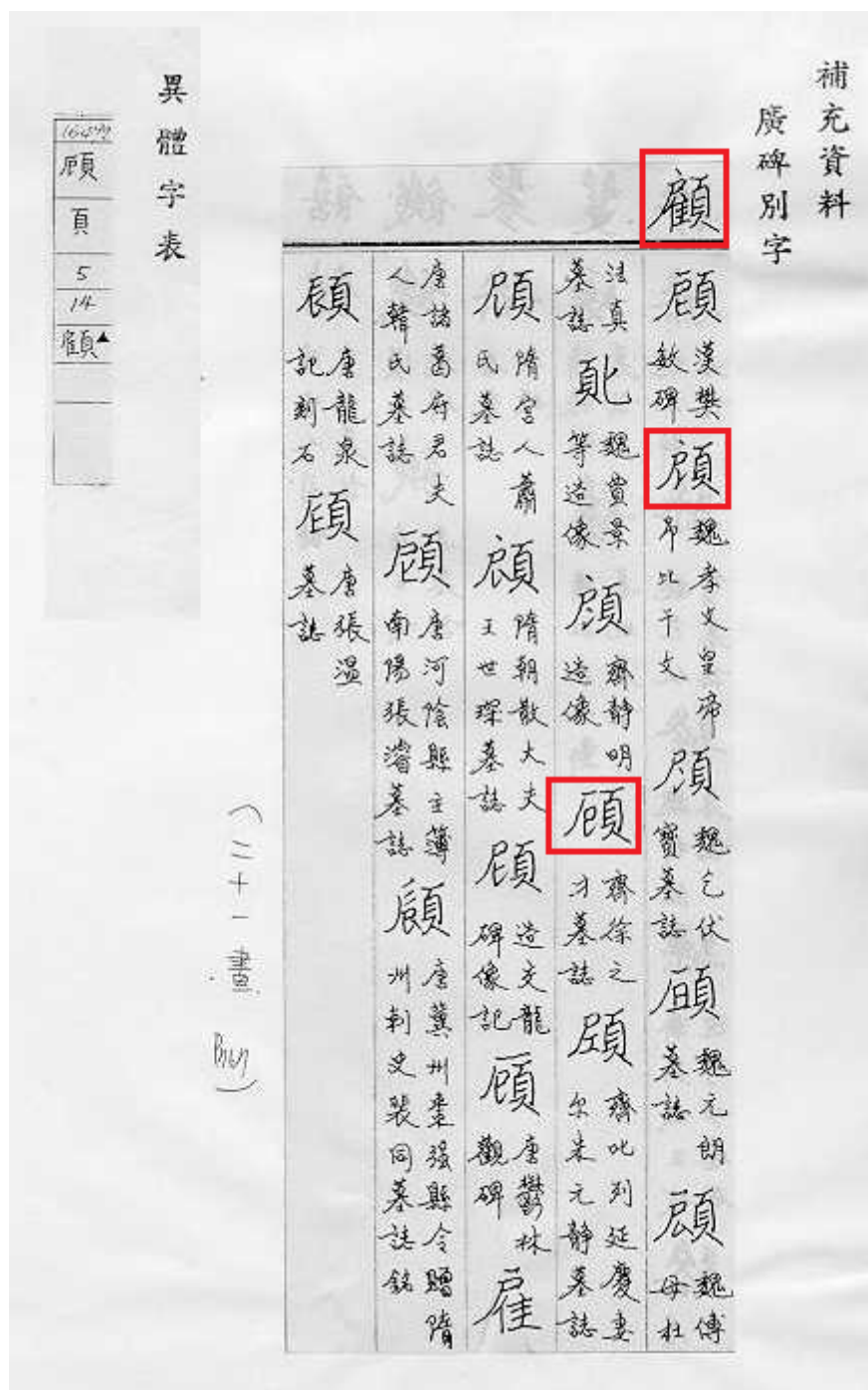


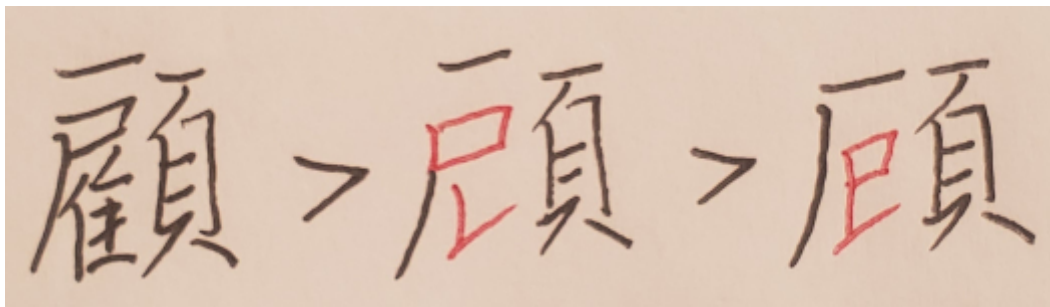
Fig.6 顧 on 異體字字典

We could see that, the component 隹 became 丩, then it combined with the 口 part in 戶, and the remaining part of 戶 became 厂. Thus, the whole glyph became 巳顧. This also explains why the inner part of 巳顧 is 巳 but not 巳:

我們可以看到，部件「隹」變成了「丩」，隨後它與「戶」中的「口」部分結合，「戶」中剩下的部分變成了「厂」。從而，整個字形變成了「巳顧」。這也解釋了為什麼「巳顧」當中的部分是「巳」而不是「巳」：

部品「隹」が「丩」に変わり、その後「戸」の「口」の部分と結合し、「戸」の残りの部分が「厂」に変わったことがわかる。これにより、グリフ全体が「巳顧」になった。こ

れは、なぜ「𠂔顧」の内部は「𠂔」ではなく「𠂔」なのかを説明している：



T-source

Let us see what CNS 11643 says about U+980B and U+2F9FF:

我們來看一下CNS11643對於U+980B和U+2F9FF怎麼說：

本機字型			全字庫字型	
顧	顧	顧	顧	顧
楷體	黑體	宋體	楷體	宋體

CNS	Unicode	BIG-5E	EUC
3-4552	980B 顧	9C60	8EA3C5D2

發音/拼音					
國語讀音					
注音	注音二式	通用	漢語	耶魯	威妥瑪
ㄍㄨˋ	e3	e3	e3	e3	ê3
ㄍㄨˋ	ě	ě	ě	ě	ê3

Fig.7 [U+980B on CNS 11643](#)

本機字型			全字庫字型	
顧	顧	顧	顧	顧
楷體	黑體	宋體	楷體	宋體

CNS	Unicode	EUC
6-6430	2F9FF 	8EA6E4B0

發音/拼音					
國語讀音					
注音	注音二式	通用	漢語	耶魯	威妥瑪
  	gu4	gu4	gu4	gu4	ku4
	gù	gù	gù	gù	ku4

Fig.8 [U+2F9FF on CNS 11643](#)

It is clear that 𪛗顧 matches the pronunciation 五果切 from *Kangxi Dictionary*, while 𪛗顧 matches the pronunciation of 顧. It is the same way as how *Kangxi Dictionary* distinguishes the two characters.

很明顯，「𪛗顧」的讀音能夠和《康熙字典》中的「五果切」對上，而「𪛗顧」的讀音也能夠和「顧」的讀音對上。這與《康熙字典》中區分這兩個字的方法一樣。

By the way, T-source has another compatibility character of U+980B which is U+2F9FE. However, it seems that there may be some glyph issues.

順便，T源有U+980B的另一個相容表意文字U+2F9FE。然而，看起來似乎存在字形問題。

2F9FE
頁 181.4

顧
T6-5A33

≡ 980B 顧
~ 980B FE01

Fig.9 U+2F9FE in Unicode 16.0

本機字型			全字庫字型	
顧	顧	顧	顧	顧
楷體	黑體	宋體	楷體	宋體

CNS	Unicode	EUC
6-5A33	2F9FE 	8EA6DAB3

發音/拼音					
國語讀音					
注音	注音二式	通用	漢語	耶魯	威妥瑪
ㄍㄨˋ  	e3 ċ	e3 ċ	e3 ċ	e3 ċ	ê3 ê3

Fig.10 [U+2F9FE on CNS 11643](#)

The T-source glyph of U+980B (T3-4552) is exactly the same as the T-source glyph of U+2F9FE (T6-5A33). Meanwhile, the pronunciation is also exactly the same. In other words, T3-4552 and T6-5A33 are exactly the same character. However, if we have a look at [WG2 N2159](#), i.e. the original proposal from TCA to submit the compatibility ideographs:

U+980B的T源字形（T3-4552）和U+2F9FE的T源字形（T6-5A33）全同。同時，它們的讀音也全同。這表明T3-4552和T6-5A33是同一个字。然而，如果我們看一眼[WG2 N2159](#)，即TCA提交相容表意文字的原始提案：

F9E8	顧	*	6-5A33	頁		顧	6-6430
			1401.151	4			0-980B
F9E9	顧	*	6-6430	頁		顧	3-4552
			1402.140	5			0-980B

Fig.11 T3-4552 and T6-5A33 in WG2 N2159

We could see that the 1st stroke of the original glyph of T3-4552 is 丿, while the 1st stroke of T6-5A33 is 一. Yet the glyph of T3-4552 in CodeCharts never matches that original glyph:

我們會發現T3-4552的原始字形的第一畫是撇，而T6-5A33的第一畫是橫。可是CodeCharts中T3-4552的字形從未與該原始字形匹配：



Fig.12 Old Versions of the Representative Glyphs of U+980B

Unicode 5.2 is the 1st version which contains multiple column charts for CJK Unified Ideographs, in other words, the 1st stroke of T3-4552 in CodeCharts has always been 一. Here I suggest T-source consider modifying **one of these two glyphs** to make them look different. Which glyph to modify can be up to T-source to decide.

Unicode5.2是首個對於中日韓統一表意文字有多列圖表的版本，換言之，在CodeCharts中T3-4552的第一畫一直是橫。在這裏我建議T源考慮修改這兩個字形中的一個來讓它們看起來不一樣。修改哪一個字形可由T源決定。

J-source

Let us see what *Daikanwa Dictionary* says about these two characters:

この2つの文字について『大漢和辭典』の言い方を見てみよう：

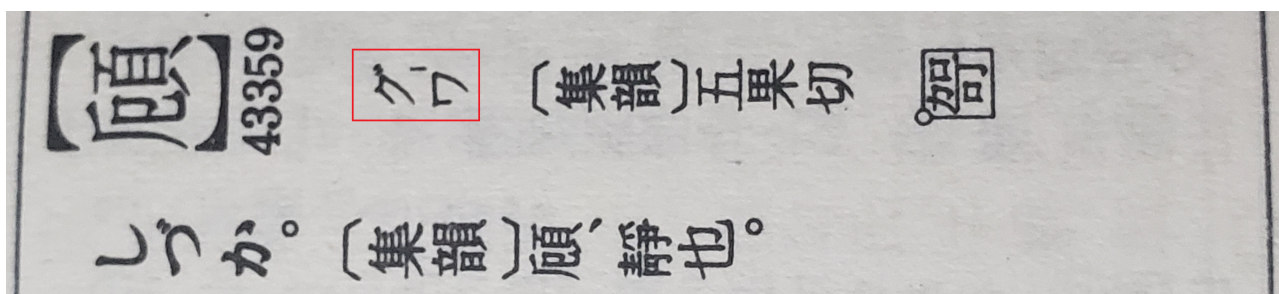


Fig.13 大漢和辭典 P12840, rotated for ease of typesetting

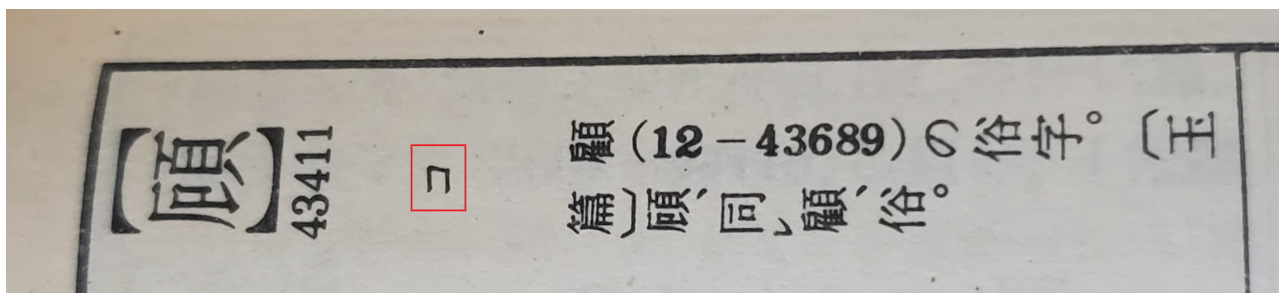


Fig.14 大漢和辭典 P12851, rotated for ease of typesetting

It is also clear that the two characters are distinguished in the same way as how *Kangxi Dictionary* does.

この2つの文字が『康熙字典』と同じように区別されていることも明らかである。

顧	顧	顧	ガ	980B	43359
IP980B	491250	00491250			
顧	顧	顧	コ・かえりみる	980B	43411
KS492000	492000	00492000			

Fig.15 汎用電子情報交換環境整備プログラム成果報告書別冊 P2347

As shown in Fig.15, Hanyō Denshi does also the same. Note that, グワ is a historical form (aka 歴史的仮名遣ひ) of ガ, so that they should be considered the same.

Fig.15によると、汎用電子も同じようにしている。

MJ028161	
 <div> コピーフィールド 玊 顧 </div>	MJ文字図形名
	MJ028161
	MJ文字図形バージョン
	1.0
	総画数
	13
	音読み
	ガ
	訓読み
	部首内画数
	頁部4画

Fig.16 [𠂔顧 \(MJ-028161\)](#) in [文字情報基盤検索システム](#)

KP1-8698	KP1-8699	KP1-869A	KP1-869B	KP1-869C
頤	頤	頤	頤	頤
kam	kyōng	ko	tam	ryōng

Here, *wa* matches the pronunciation 五果切 from *Kangxi Dictionary*, while *ko* matches the pronunciation of 顧. It is also the same way as how *Kangxi Dictionary* distinguishes the two characters. Since *wa* is impossible to be sorted between *kyōng* and *tam*, *ko* is also impossible to be sorted between *ye* and *wan*, we can confirm that KPS 10721 also treats 𠂇頤 and 𠂇顧 as two different characters.

The Residual Strokes Issue

As we can see in Fig.11, Fig.17 and Fig.18, the residual strokes count of 𠂇顧 should be 5. Actually, not only Moji Jōhō and KPS 10721, but also *Kangxi Dictionary*, [CNS 11643](#) and *Daikanwa Dictionary* say that it should be 5. However, as shown in Fig.1, the residual strokes counts of both U+FACB and U+2F9FF are 4. No matter whether we finally decide to disunify 𠂇顧 and 𠂇顧 or not, the residual strokes counts of U+FACB and U+2F9FF should be revised to 5.

如同我們在Fig.11、Fig.17和Fig.18中看到的那樣，「𠂇顧」的部外筆畫數應為5。實際上，不僅文字情報和KPS10721，《康熙字典》、[CNS11643](#)和《大漢和辭典》都說應為5。然而，如Fig.1所示，U+FACB和U+2F9FF的部外筆畫數均為4。無論我們最終是否決定解除認同「𠂇顧」和「𠂇顧」，U+FACB和U+2F9FF的部外筆畫數都應該修正為5。

Fig.11、Fig.17、Fig.18の示しているように、「𠂇顧」の部首内画数は5であるべきである。実は、文字情報やKPS10721だけでなく、『康熙字典』、[CNS11643](#)、『大漢和辞典』も5であるべきだと言っている。しかし、Fig.1によると、U+FACBとU+2F9FFの部首内画数はいずれも4である。最終的に「𠂇顧」と「𠂇顧」を統合しないようにすることにしたかどうかにかかわらず、U+FACBとU+2F9FFの部首内画数を5に修正すべきである。

Conclusion

To sum up, no glyph needs to be removed from U+980B itself, but only to remove the variation relationship between it and its compatibility characters (U+FACB and U+2F9FF), i.e. to encode T6-6430 and KP1-869A onto a new codepoint. Japan NB may consider moving JMJ-028162 (now encoded as U+980B_U+E0101) to that new codepoint.

綜上所述，U+980B本身不需要移走任何字形，只要解除它和它的相容表意文字（U+FACB和U+2F9FF）的異體字關係即可，即把T6-6430和KP1-869A編碼到一個新的碼位上。日本可考慮把JMJ-028162（目前編碼為U+980B_U+E0101）移動到那個新碼位上。

以上のように、U+980Bそのものから字形を移動する必要はなく、それとその互換漢字（U+FACBとU+2F9FF）との異体字関係を解除すればいい、すなわち、T6-6430とKP1-869Aを新しいコードポイントに符号化すればいい。日本はJMJ-028162（現在はU+980B_U+E0101と符号化）をその新しいコードポイントに移動することを検討することができる。

Meanwhile, China NB may also consider a horizontal extension on that new codepoint, using *Kangxi Dictionary* as the source reference.

同时，中国也可考虑在那个新码位上横向扩展，使用《康熙字典》作为提交源参考。

Additionally, China NB may consider change the source reference on U+980B to GKX-1400.06 or [GHZR-84647.03](#) if necessary, since GE-source is actually a pseudo-G-source for the sake of horizontal extensions in early years. The glyph does not need any changes.

此外，若有需要，中国还可考虑将U+980B的提交源参考换成GKX-1400.06或[GHZR-84647.03](#)，因GE源（第七辅助集）实为为早期的横向扩展而建立的假G源。字形无需改变。

	G-source	M-source	T-source	J-source	KP-source
頤 980B	GE-542C <i>(GKX-1400.06)</i> <i>(GHZR-84647.03)</i>	MD-980B	T3-4552	JMJ-028161	KP1-8696
頤 FACB	/	/	/	/	KP1-869A KPU-0FACB
頤 2F9FF	/	/	T6-6430 TU-2F9FF	/	/
頤 2????	GKX-1402.14	/	T6-6430	JMJ-028162	KP1-869A

(texts in red indicate the new source references on the corresponding codepoint)

(End of document)