

Source:	CheonHyeong Sim (沈天珩, aka TianHeng Shen), Ziheng Gao (高子恒)
Title:	Proposal to Disunify U+5CC0 (𡇗 & 𡇘)
Status:	Individual Contribution on IRG #62
Action:	To be considered by IRG and multiple sources

Background

U+5CC0 looks like the following picture in the CodeCharts of Unicode 15.1:



Fig.1 U+5CC0 in Unicode 15.1

And the following picture shows the CodeCharts of Unicode 16.0 Alpha Review, after the horizontal extension by J-source:



Fig.2 U+5CC0 in Unicode 16.0 (Alpha Review)

Additionally, from Unicode 3.1 (March, 2001), U+5CC0 has a compatibility character U+2F879 with its source TF-2662:

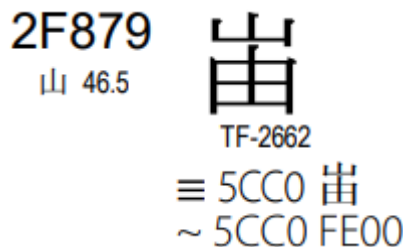


Fig.3 U+2F879 in Unicode 15.1

Obviously, U+5CC0 and U+2F879 mixed the two shapes 𡇗 and 𡇘. These two shapes are now considered as variants, but actually they are phonetically, graphically and semantically different, and need to be disunified. In [IRGN2210](#), Eiso Chan once raised this

issue to TCA but TCA disagreed to disunify this character. However, this is not an issue solely related to TCA. Please see the following description.

Properties of 岫 and 岫

For the former one, in IRGN2210, Eiso Chan gave out some evidences related to G-source and K-source to show that 岫 is a variant of 岫. Due to space constraints, the evidences in IRGN2210 will not be repeated here. In this proposal, I would like to add evidences related to T-source and KP-source.

Although 教育部異體字字典 does not include 岫 as a character entry, in 形體資料表 column of 岫, 岫 appears as its variant:

國家教育研究院 維護

部首查詢 | 筆畫查詢 | 單字查詢 | 複合查詢 | 注音查詢 | 漢語拼音查詢 | 倉頡碼查詢 | 四角號碼查詢 | 附收字查詢

A01110 **岫** 解說 異體字 研訂瀏覽 回查詢結果

異體字

岫 岫

解說

字號	A01110	正字	【岫】 山 - 05 - 08
說文釋形	大徐本：岫，山穴也。从山，由聲。𡵓，籀文，从宀。（似又切） 段注本：岫，山有穴也。从山，由聲。𡵓，籀文，从宀。（似又切）		
字樣說明	右半作「由」：筆順如「田」，中豎筆上須出頭，寫法參「由」字。		
注音	ㄒㄩˋ		
漢語拼音	xiù		
釋義	1. 山洞、巖洞。段注本《說文解字·山部》：「岫，山有穴也。」晉·張協〈雜〉詩一〇首之九：「荒庭寂以閉，幽岫峭且深。」宋·李師中〈詠松〉詩：「半依岫岫倚雲端，獨上亭亭耐歲寒。」		

形體資料表
 重訂直音篇

岫
 音 襍
 山 有
 日 岫
 有
 岫
 上 同

卷六·山部·頁474·右

Fig.4 岫 in 教育部異體字字典

It matches the statement in IRGN2210.

In KP0, the characters are sorted by the Korean pronunciation in the DPRK order, and we could see that 岫 is located between 岫 (pronounced “su”) and 帥 (pronounced “su”), so its pronunciation should undoubtedly be “su”.

Fig.5 KPS 9566 on zi.tools

In KP1, the characters are sorted by the radical first; for the same radical, the characters are sorted by the residual strokes; and when the residual strokes are also the same, they are still sorted by pronunciation (Sim, 2022). We could see that 岫 is located between 岫 (pronounced “su”) and 岫 (pronounced “jeo”), so its pronunciation should very likely be “su” (the variant of 岫), and impossible to be “bang” (岫, the variant of 邦), since b comes before s.

3F0_ 屨
 3F1_ 屨
 3F2_ 屨
 3F3_ 屨
 3F4_ 屨
 3F5_ 屨
 3F6_ 屨

Fig.6 [KPS_10721_reconstructed_by_CheonHyeong_Sim](#)

It also matches the statement in IRGN2210.

For the latter one, in IRGN2210, Eiso Chan also gave out some evidences related to G-source and T-source to show that 屨 is a variant of 邦; and for G-source, it is also an independent character used for a place name different from 邦. Due to space constraints, the evidences in IRGN2210 will not be repeated here. In this proposal, I would like to add evidences related to J-source.

In 《大漢和辭典》 (*Daikanwa* Dictionary), 屨 is interpreted as the ancient form of 邦, thus it is a variant of 邦.

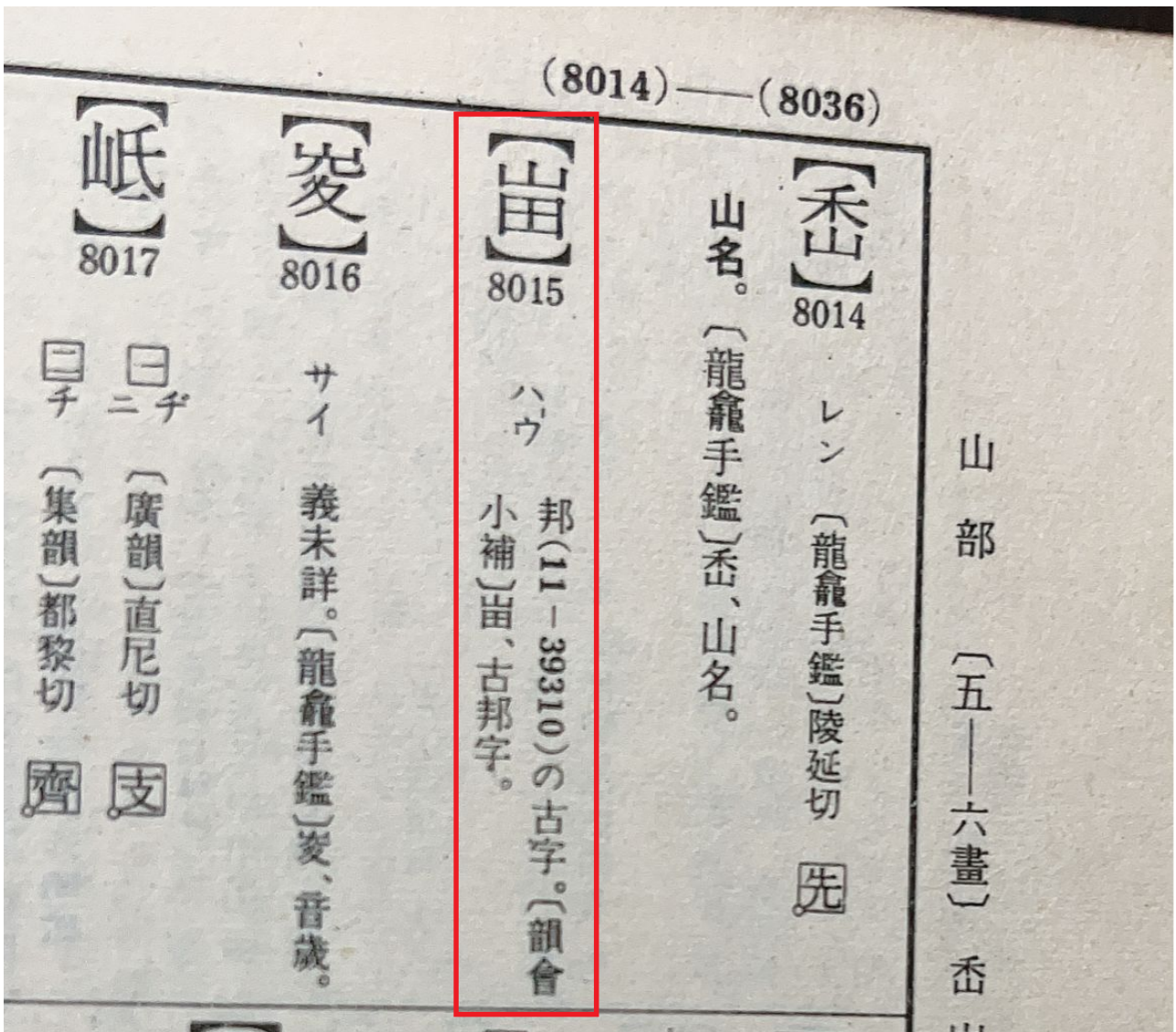


Fig.7 《大漢和辭典》P3646 (i.e. 第4冊 P244)

It also matches the statement in IRGN2210.

By the way, K-source also includes 岷 as KC-07223; but in their [CJK Ideographs Search System](#), the given provenance and context in [the page of KC-07223](#) are not for 岷 but for 岷 (U+31844, KC-07242). Maybe the database has some malpositions, and I hope that the Korean experts could give out the right provenance and context for 岷, to show us whether it is also the variant of 邦 or not.

Summary

To briefly sum up, we can list 岷 and 岷 in multiple sources as the following table: (Note that, red indicates the current sources for U+5CC0, including the newly horizontally-extended J-source in Unicode 16.0 Alpha Review)

	G-source	T-source	J-source	K-source	KP-source
畠	GE-3735	TF-2662	JMJ-068030	K0-617C	KP0-E3E3 (KP1-3F3E)
畠	GKX-1594.81 GHZR-10791.16 (GDM-?????)	T4-2634	JMJ-010468	KC-07223	/

Suggestions on Disunifying

As mentioned above, 畠 and 畠 are phonetically, graphically and semantically different in multiple sources. Thus it is unreasonable to unify them.

Since K0 and KP0 are the basic character sets in South Korea and North Korea, the correspondence between them and Unicode are not able to be changed. We suggest to keep 畠 on U+5CC0 – just as in the initial Unicode version; and move 畠 to a new codepoint (maybe as a UNC, or maybe as a new G-source character to WS2024 as the place name).

In other words, TCA is suggested to move T4-2634 to a new codepoint, TF-2662 to U+5CC0, and change the source reference on U+2F879 to TU-2F879; Japan is suggested to move MJM-010468 to a new codepoint, and add MJM-068030 for U+5CC0; China and (South) Korea are suggested to do a horizontal extension on the new codepoint, or to submit 畠 as a new character.

(End of document)

IRGN2676_KIM_resp1

(= k2680_811_IRGN2676_U+5CC0_KIM_K)

Subject: KIM's response to IRGN2676Disunify5CC0.pdf (2024-03-09)

Date: 2024.03.13.

Status: Individual opinion

Authors: KIM Kyongsok

1. Background

- This is a response to IRGN2676Disunify5CC0.pdf (2024-03-09) by CheonHyeong Sim (沈天珩, aka TianHeng Shen) and Ziheng Gao (高子恒).

- IRGN2676 is much related with the issue #1 in IRGN2210R_U+5CC0_U+2F936__Eiso_CHAN dated on 2017-06-08 at IRG #48.

- At IRG #48, I could find the following comments in the editorial report.

o IRGN2210 Dis-unification of 2 Characters

The editors reviewed all raised issues and will respond before next IRG meeting. For Issue 2, Need consideration

- TCA seems to have said that TCA needed more time to investigate this issue. However, the author could not find a response from TCA after IRG #48 (please correct me if I am wrong).

- It seems that, after IRG #48, the issue of disunifying U+5CC0 mentioned in IRGN2210R has not been discussed.

- The current situation of two characters, 𪗇 U+5CC0 and 𪗈 U+2F879 (Compatibility char.) is well summarized in IRGN2676 and repeated below:

	G-source	T-source	J-source	K-source	KP-source
𪗇	GE-3735	TF-2662	JMJ-068030	K0-617C	KP0-E3E3 (KP1-3F3E)
𪗈	GKX-1594.81 GHZR-10791.16 (GDM-?????)	T4-2634	JMJ-010468	KC-07223	/

5CC0
山 46.5

𪗇
GE-3735

𪗈
T4-2634

𪗈
JMJ-010468

𪗇
K0-617C

𪗇
KP0-E3E3



2. Review results

- The author reviewed the evidences and opinions in the two documents.
- It seems that, based on the evidences in the two documents, the suggestions in the two documents are quite logical and reasonable.

3. Suggestions

- Basically, the suggestions below are based on the suggestions in the two documents.

1) TCA could consider moving T4-2634 glyph and source reference in U+5CC0 to a new character (codepoint) 畠 (NOT to U+2F879, a compatibility char).

2) TCA could consider moving TF-2662 glyph and source reference in U+2F879 to U+5CC0.

3) TCA could consider changing the T source reference of U+2F879 to TU-2F879.

4) Japan could consider moving JMJ-068030 glyph and source reference to a new character (codepoint) mentioned in 1) above.

5) Japan could consider adding JMJ-010468 glyph and source reference to U+5CC0.

6) Rep. of Korea (src. ref. KC-07223) and China (src. ref. GKX-1594.81) could consider

6a) doing a horizontal extension on the new character (codepoint) mentioned in 1) above; or

6b) submitting 畠 as a new character.

7) IRG could consider removing UCV #94.

94	畠 畠			092/192 5CC0	畠 畠 E-2635 4-2630 E-0621 4-0620	畠 0-017C E-0532	畠 2F879
----	--------	--	--	-----------------	---------------------------------------	-----------------------	------------

4. RE: the wrong information about KC-07223 in <https://www.koreanhistory.or.kr/newchar/>

- The evidence information about KC-07223 in koreanhistory.or.kr is NOT about KC-07223, but about a different character 擷 (KC-07242, U+31844).

- I will try to contact the manager of this web site. However, it may not be easy to get the correct evidence information since the evidence information is not updated these days.

자형	 이체자 보기		
부수	山	통합코드	KC07223
획수	5	총획수	8
KS코드		JIS코드	
GB코드		BIG5코드	
한국뜻			
한글음가			
영어뜻			
한어병음			
일본어 훈독			
사각호마	파자 (山/田)		

The following information is about KC-07242 (=U+31844), NOT about KC-07223.

문자이미지파일		기관명	한국고전번역원
사업차수	2008	관련DB	고전번역총서
출전서명	용재집 (容齋集)	출전목록	3책
출전위치	[부]김석근(金@根)	출전파일명	exm_UM_6JOc_JS_001318.xml
전후문맥	<전력 type="전력" 식별자="" id="">무오 혼충(戊午魂忠)</전력> <관직 type="관직" 식별자="" id=""></관직> <봉작 id="" 식별자=""></봉작></관력> <타과 type="타과"></타과> <본관 id="296" 식별자="">안동(安東)</본관> <기타></기타> </본인> <친족 가계="친족" 대수="1" 유형="부"> <인명 id="" type="성명" 식별자="">김석근(金{才+稟}根)</인명> <인명 type="자"></인명> <인명 type="호"></인명> <관력> <관직 type="관직" 식별자=""		원전이미지파일명

* * *

WG2 N5262

Title: Comment on WG2 N5259

Source: Shuichi TASHIRO, Japan

Reference: [WG2 N 5259](#), [IVD 2017-12-12](#), [UCVs](#)

Date: 2024-5-8

According to WG2 N 5255 and WG2 N 5259, at IRG Meeting #62, IRG N2676's "Proposal to Disunify U+5CC0" and IRGN2676_KIM_resp1 reached a unanimous consensus as stated in IRG Recommendation M62.07.

IRG N2676 proposes disunifying characters of 𪗇山田, including JMJ-10468 proposed in CDAM2.4, to other code points, from U+5CC0, and suggests assigning new code point for them.

IRGN2676_KIM_resp1 suggests removing UCV#94 additionally.

However, the unification rules applied to U+5CC0 is designated as UCV#94 and it has been shared within the IRG, and past CJK encodings have been developed based on this rule.

It must be noted that altering the unification rules once applied for the encoding is a disruptive act against the standard. Such actions should not be taken in fulfilling our responsibility of maintaining interoperability.

On the other hand, it is acceptable that there is a perspective to considering 𪗇山田 and 𪗇山由 as different characters according to discussions and evidences presented in IRG N2210R, IRG N2676, and IRGN2676_KIM_resp1.

Thus, the only two theoretically possible ways are:

1. Let the over-unification of U+5CC0 as is for backward compatibility.

In fact, both JMJ-010468 (𪗇山田) and JMJ-068030 (𪗇山由) were assigned to U+5CC0 in accordance with the UCV#94 , and registered in the Moji_Joho collection of the IVD 2017-12-12:

5CC0	𪗇	𪗇
	E0100 Moji_Joho MJ010468	E0101 Moji_Joho MJ068030

Based on this, all administrative systems built in accordance with ISO standards in Japan currently use this code for processing resident names and other information.

- 2a. Assign new code points, different from U+5CC0, to both 𪗇山田 and 𪗇山由 in order to accommodate the perspective that they are different characters.
- 2b. Instead of assigning two new code points to both of them, it may be possible to keep U+2F879 for 𪗇山由 and change its property from the current CJK Compatibility Ideographs to Unified Ideographs.