ISO/IEC JTC1/SC2/WG2/IRG N2676

Date 2024−03−09

Source:	CheonHyeong Sim (沈天珩, aka TianHeng Shen), Ziheng Gao (高子恒)
Title:	Proposal to Disunify U+5CC0 (峀 & 峀)
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:

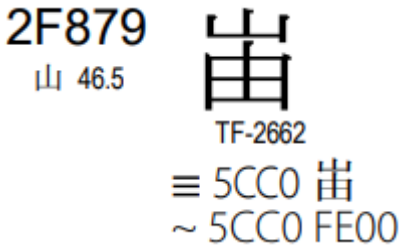


Fig.3 U+2F879 in Unicode 15.1

Obviously, U+5CC0 and U+2F879 mixed the two shapes 峀 and 峀. 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 峀 and 岫

For the former one, in IRGN2210, Eiso Chan gave out some evidences related to G–source and K–source to show that 峀 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 峀 as a character entry, in 形體資料表 column of 岫, 峀 appears as its variant:

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 峀 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 峀 is located between 岫 (pronounced “su”) and 岨 (pronounced “jeo”), so its pronunciation should very likely be “su” (the variant of 岫), and impossible to be “bang” (峀, the variant of 邦), since b comes before s.

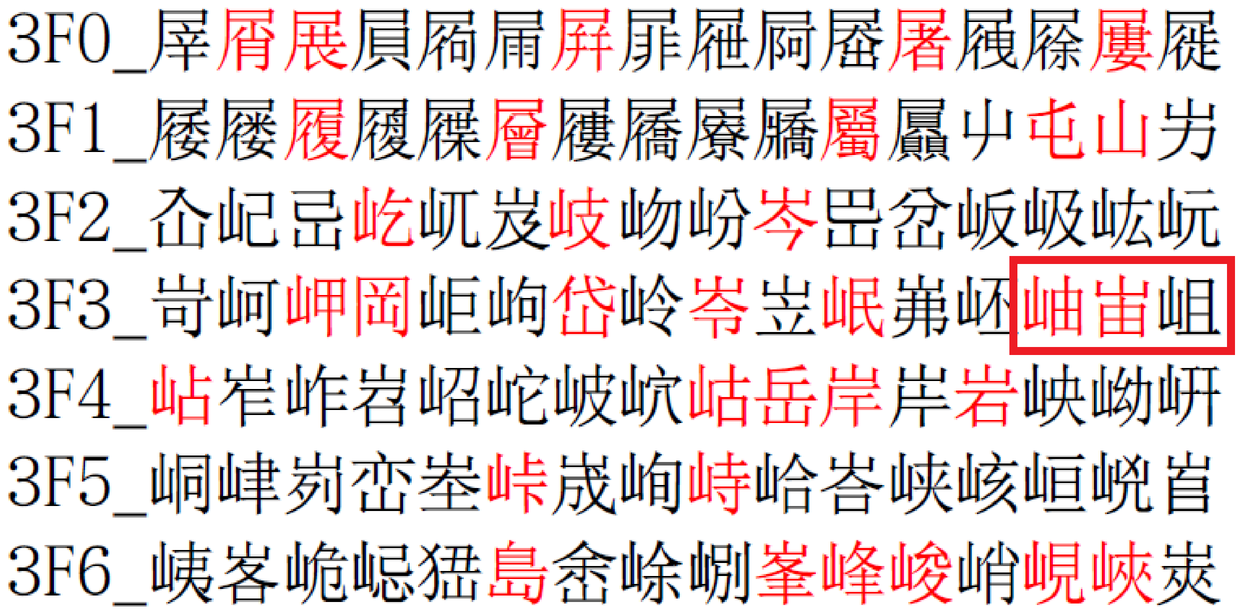
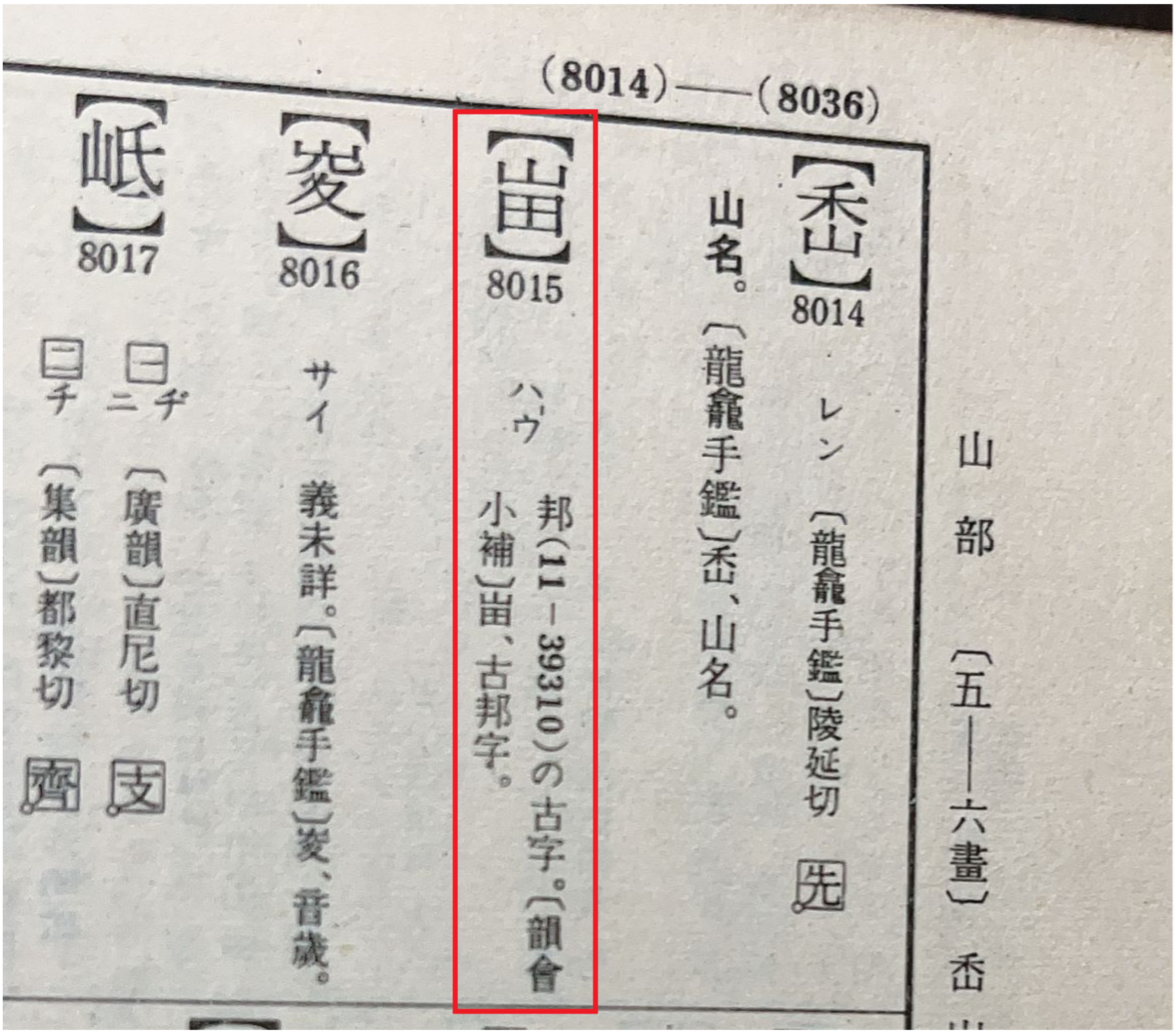


Fig.6 KPS 10721 reconstructed by CheonHyeong Sim

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 峀 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), 峀 is interpreted as the ancient form of 邦, thus it is a variant of 邦.


山部〔五——六畫〕

(8014)——(8036)

8014 レン 〔龍龕手鑑〕陵延切 先
山名。〔龍龕手鑑〕、山名。

【峀】8015 ハウ
邦(11－39310)の古字。〔韻會小補〕峀、古邦字。

8016 サイ
義未詳。〔龍龕手鑑〕、音歳。

【岻】8017
曰 ニチ 〔廣韻〕直尼切 支
曰 チ 〔集韻〕都黎切 齊

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 峀 as KC-07223; but in their CJK Ideographs Search System, the given provenance and context in the page of KC-07223 are not for 峀 but for 㨨 (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 峀, to show us whether it is also the variant of 邦 or not.

Summary

To briefly sum up, we can list 峀 and 峀 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
峀	GE−3735	TF−2662	JMJ−068030	K0−617C	KP0−E3E3 (KP1−3F3E)
峀	GKX−1594.81 GHZR−10791.16 (GDM−?????)	T4−2634	JMJ−010468	KC−07223	/

Suggestions on Disunifying

As mentioned above, 峀 and 峀 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 峀 on U+5CC0 − just as in the initial Unicode version; and move 峀 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 JMJ−010468 to a new codepoint, and add JMJ−068030 for U+5CC0; China and (South) Korea are suggested to do a horizontal extension on the new codepoint, or to submit 峀 as a new character.

(End of document)