

Source:	TianHeng Shen(沈天珩, aka CheonHyeong Sim)
Title:	Application for Horizontal Extensions of Multiple Sources in CJK-ExtI
Status:	Individual Contribution on IRG #61
Action:	To be considered by IRG, TCA, Japan, Korea, Vietnam, UTC and UK

In CJK-ExtI (U+2EBF0..U+2EE4A, a new UNC block introduced in Unicode15.1), some characters are duplicated with WS2021. These characters have to be removed from WS2021 later, and some of them with sources other than G-source may possibly be horizontally extended to CJK-ExtI. Note that there is only G-source (GIDC23-XXX) for every character in CJK-ExtI.

**Whether doing the horizontal extension or not should be actually decided by TCA, Japan, Korea, Vietnam, UTC and UK, while this proposal is only for reminding.**

This proposal for horizontal extensions can be mainly divided into three parts.

- >>1. Characters from WS2021
- >>2. Newly introduced half-duplicated characters
- >>3. Other possible horizontal extensions

## 1. Characters from WS2021

Below is the full list of the characters from WS2021 which can be horizontally extended to CJK-ExtI. (The actual shapes different from the actual shapes of G-source are marked red)

### 1.1. T-source: (17 chars)

Codepoint	G-glyph	T-glyph	WS2021 No.	T-source Ref.
2EC21	埤	埤	00843	TD-2351
2EC8B	捌	捌	01429	TC-266C
2ECA8	脍	脍	01738	TC-4A2B
2ECD1	漕	漕	02103	TE-244D
2ED1E	瑑	瑑	02489	TD-4022

2ED2A	琰	琰	02516	TE-6F4A
2ED30	瓊	瓊	02533	TE-362C
2ED31	璿	璿	02534	TE-3C34
2ED4B	稔	稔	02763	TD-2E3B
2ED53	確	確	02801	TE-2622
2ED76	粃	粃	02998	TC-407C
2EDA9	藪	藪	03440	TE-302C
2EDAB	藹	藹	03468	TE-3D3B
2EDC4	寔	寔	00122	TD-4828
2EDE4	鐘	鐘	04130	TE-3924
2EE25	韶	韶	04405	TD-5825
2EE2E	鎔	鎔	04459	TE-3266

#### 1.2. K-source: (2 chars)

Codepoint	G-glyph	K-glyph	WS2021 No.	K-source Ref.
2EC21	埕	埕	00843	KC-00674
2EDE4	鐘	鐘	04130	KC-05820

#### 1.3. V-source: (1 char)

Codepoint	G-glyph	V-glyph	WS2021 No.	V-source Ref.
2ED8B	瞄	瞄	03243	V4-5229

#### 1.4. U-source: (2 chars)

Codepoint	G-glyph	U-glyph	WS2021 No.	U-source Ref.
2ED49	砵	砵	02755	UTC-03172
2EDC4	寔	寔	00122	UTC-03195

#### 1.5. UK-source: (12 chars)

Codepoint	G-glyph	UK-glyph	WS2021 No.	UK-source Ref.
2EC12	埕	埕	00790	UK-20847
2EC14	埕	埕	00802	UK-20845
2EC29	璇	璇	00879	UK-20636
2EC2D	壩	壩	00904	UK-20000

2EC2E	壻	壻	00900	UK-20265
2EC65	崩	崩	01114	UK-20626
2EC78	徕	徕	01295	UK-20601
2ECA5	瞪	瞪	01707	UK-20958
2ED18	瑫	瑫	02481	UK-20029
2ED41	𠂇	𠂇	02736	UK-20642
2EDE4	鐘	鐘	04130	UK-20922
2EE0E	鑄	鑄	04211	UK-20966

## 2. Newly introduced half-duplicated characters

There are two newly introduced half-duplicated characters, they are U+2ED90(艾) and U+2EDD1(逢). They are different in some aspects, so they will be discussed separately.

### 2.1. U+2ED90(艾/艾)

First let us see U+8286 and U+2ED90 in the CodeCharts.







The actual shape of T-source of U+8286 seems to be closer to the actual shape of U+2ED90. It is clear that, the abstract shape of U+8286 (mainly G-source) is 𠂇 while the abstract shape of U+2ED90 is 艾. So what should be explored here is the abstract shape of the actual shape 艾 of T-source.


For T3-2728, let us see the data in CNS11643. (the picture is too high to put it on this page, see the next page)

本機字型			全字庫字型	
				
楷體	黑體	宋體	楷體	宋體

發音/拼音

注音	注音二式	通用	漢語	耶魯	威妥瑪
ㄞˋ  	ai4 ài	ai4 ài	ai4 ài	ai4 ài	ai4 ai4
ㄔˊ  	cha chā	cha cha	cha chā	cha chā	ch'a ch'a
ㄧˋ  	yi4 yì	yi4 yì	yi4 yì	yi4 yì	i4 i4
ㄔˊ  	chai chāi	chai chai	chai chāi	chai chāi	ch'ai ch'ai

From the pronunciation, we can easily see that the abstract shapes mixed together. Since it is a character from T3-source, in other words, it is not from Big5, there is no difficulty to move it to non-BMP.

What is more, there is also an actual shape  in T-source, which is mapped to TC-7730. Following is the data of TC-7730 in CNS11643.

本機字型			全字庫字型	
				
Unencoded	Unencoded	Unencoded	楷體	宋體

發音/拼音

注音	注音二式	通用	漢語	耶魯	威妥瑪
ㄔˊ  	chai chāi	chai chai	chai chāi	chai chāi	ch'ai ch'ai
ㄞˋ  	ai4 ài	ai4 ài	ai4 ài	ai4 ài	ai4 ai4
ㄔˊ  	cha chā	cha cha	cha chā	cha chā	ch'a ch'a

This proposal requests TCA to move T3-2728 to U+2ED90, and make a horizontal extension for TC-7730 on U+8286.

## 2.2. U+2EDD1(逄/逢)

This character is undoubtedly the variant of 逄/逢. TCA once submitted this character in the early version of CJK-ExtC, but was eventually deleted; Japan also submitted this character in CJK-ExtF, but was also deleted due to the unification with 逢 (U+8FF3). Since it is finally included in CJK-ExtI, TCA and Japan may consider make a horizontal extension on U+2EDD1, respectively TC-425E and JMJ-058848.

## 3. Other possible horizontal extensions

When preparing the data for CJK-ExtI on [yedict.com](http://yedict.com), Eddie Li found some other possible horizontal extensions, and I will list these cases below. Note that the list is highly likely not comprehensive, and just for TCA's consideration. Totally 104 characters.

Codepoint	G-glyph	T-source Ref.
2EBF2	伋	TC-2263
2EBF3	伋	TC-2265
2EBF6	伋	TC-2837
2EBF7	怪	TC-277B
2EBFA	倝	TC-3679
2EC03	𠂇	TC-424D
2EC04	𠂇	TD-362F
2EC05	𠂇	TC-2F33
2EC09	𠂇	TD-5945
2EC16	𠂇	TC-3941
2EC1E	𠂇	TC-5446
2EC20	𠂇	TD-2421
2EC22	𠂇	TD-2376
2EC2F	𠂇	TD-6845
2EC33	𠂇	TC-3942
2EC39	𠂇	TC-235C
2EC3B	𠂇	TC-2A43

Codepoint	G-glyph	T-source Ref.
2EC41	姁	TC-3A24
2EC43	嬰	TC-3A27
2EC4C	嫫	TC-552E
2EC52	嫫	TD-3869
2EC59	嫫	TE-342E
2EC61	岩	TC-3126
2EC64	𠂇	TC-312C
2EC68	𠂇	TC-475F
2EC6D	𠂇	TD-692F
2EC76	𠂇	TC-3166
2EC84	𠂇	TD-272F
2EC85	𠂇	TD-266B
2EC86	𠂇	TD-3A4E
2EC8D	𠂇	TC-4874
2EC90	𠂇	TC-3C2D
2EC92	𠂇	TC-5C3F
2EC96	𠂇	TD-2844

2EC9A	𠄎	TC-3273
2EC9B	𠄎	TC-3C57
2EC9E	𠄎	TC-497A
2ECA0	𠄎	TD-3B69
2ECA1	𠄎	TD-3B7B
2ECA2	𠄎	TD-4C6A
2ECA9	𠄎	TC-333E
2ECAA	𠄎	TC-334C
2ECAB	𠄎	TC-333A
2ECBB	𠄎	TD-5D6F
2ECC5	𠄎	TC-5B3A
2ECC6	𠄎	TC-5A3C
2ECCB	𠄎	TD-4E4D
2ECD3	𠄎	TC-2C65
2ECDC	𠄎	TC-3E5F
2ECDD	𠄎	TC-4C35
2ECE0	𠄎	TC-4C2C
2ECE6	𠄎	TD-2C60
2ECEA	𠄎	TD-3F4B
2ECEB	𠄎	TD-4F3A
2ED00	𠄎	TC-3F45
2ED02	𠄎	TC-3F50
2ED10	𠄎	TC-5B7B
2ED12	𠄎	TC-5B77
2ED15	𠄎	TC-5C29
2ED1A	𠄎	TD-2D65
2ED1B	𠄎	TD-2D4B
2ED1C	𠄎	TD-405C

2ED1D	𠄎	TD-404B
2ED20	𠄎	TD-5036
2ED22	𠄎	TD-5038
2ED2E	𠄎	TE-2E2B
2ED37	𠄎	TC-4D30
2ED42	𠄎	TC-4049
2ED47	𠄎	TC-5D2A
2ED4A	𠄎	TC-5D2C
2ED4E	𠄎	TD-414F
2ED50	𠄎	TD-515D
2ED51	𠄎	TD-515F
2ED60	𠄎	TD-5237
2ED6B	𠄎	TC-5D7E
2ED70	𠄎	TD-526A
2ED73	𠄎	TD-616C
2ED77	𠄎	TD-5334
2ED79	𠄎	TE-2677
2ED81	𠄎	TC-3553
2ED87	𠄎	TC-4F3C
2ED88	𠄎	TC-4C64
2ED8C	𠄎	TC-4F4A
2ED8E	𠄎	TC-2756
2ED8F	𠄎	TC-2D3C
2ED92	𠄎	TC-4170
2ED93	𠄎	TC-5F5D
2ED95	𠄎	TC-5F31
2EDAA	𠄎	TE-373D
2EDAF	𠄎	TD-4541

2EDB4	襪	TE-4535
2EDB6	換	TE-285C
2EDD0	逯	TC-4257
2EDD3	途	TD-333A
2EDD5	遐	TD-464F
2EDD8	適	TD-653B
2EDDA	那	TC-2C32

2EDE1	銖	TD-6578
2EDE5	鑄	TE-3E49
2EE1A	隍	TD-343D
2EE1C	霽	TC-614F
2EE1D	霽	TD-4756
2EE23	馮	TD-4761
2EE24	馮	TD-4762

(End of document)

Universal Multiple-Octet Coded Character Set  
UCS

ISO/IEC JTC1/SC2/IRGN2635\_TCA Feedback

Date: 2023-10-13

Doc. Type:	Member body contribution
Title:	TCA's Feedback to IRGN2635
Source:	TCA
Status:	Input to IRG meeting 61
Action:	For consideration by JTC1/SC2/WG2/IRG
Distribution:	IRG Members and Ideographic Experts
No. of pages:	2
Appendixes	None
Medium:	Electronic

[IRGN2635](#) mentions T-source, so TCA reads it seriously. This document has done a initial analysis on CJK Extension I, and the issues found are divided into 3 topics to explain, TCA would like to put forward some ideas here:

- **Part 1- Characters from WS2021:** The characters mentioned in this section are not coded in Ext-I. As evidenced by [WG2N5238](#) (proposed by China), this issue does not exist.
- **Part 2- Newly introduced half-duplicated characters:** Two similarities with other encoded glyphs were mentioned: U+2ED9D 𪗎<sup>++</sup>义, U+2EDE0 𪗎<sub>一</sub>圣, which should be unified according to the UCV rule. TCA proposed that this part could be further discussed by the IRG. See below for details.
- **Part 3- Other possible horizontal extensions:** It is mentioned in the document that 104 T-source characters can be extended horizontally. Thanks to the proposer's suggestion, TCA has considered applying for horizontal extension.

Regarding Part 2, how to solve this situation that Ext-I has not follow the rule of unification(UCV) completely? TCA has two concerns:

1. The first question is, Ext-I is already coded separately, how to deal with the part that creates the same glyphs but with different code points? Please see the following two examples.
2. The second question is, names and place names have a special legal status and



the requirement for the government usage, and it is not possible to follow the rule of unification completely. If they do not follow the principle of unification, the above mentioned issues will appear again. In this case, what should be done with the unified glyphs?

**The above two issues should be discussed seriously by the experts from WG2, IRG, and UTC, as they will affect the submission of new working sets and the review principles. TCA suggests IVS or other solutions, in the hope that we can solve the problems.**

● 2 examples :

1. U+2EC7B vs U+5F84

Ext-I code/Glyph		Other encoded code /Glyph			Note
2EC7B	2EC7B 𢦏 60.5 GIDC23-140	5F84	5F84 𢦏 60.5 G0-3E36	徑 徑 徑 T3-2875 J0-3742 K2-314A	UCV #87 圣圣

From this set of example, we can find that the glyph of U+2EC7B are exactly the same as the glyphs of T, J and K of U+5F84.

2. U+2ED9D vs U+8286

Ext-I code/Glyph		Other encoded code /Glyph			Note
2ED9D	2ED9D 艸 140.3 艾 GIDC23-430	8286	8286 艸 140.3 艾 G3-6837	艾 T3-2728	UCV#127 义叉

In this group, the glyph for U+2ED9D is the same as the glyph for T-source (T3-2728) in U+8286. If it is T-source moved to U+2ED9D, TCA does not agree with it because it is a commonly used character and "艾" is a basic character(正字) and "艾" is a variant of "艾".

*(End of document)*