

Universal Multiple-Octet Coded Character Set  
UCS

ISO/IEC JTC1/SC2/WG2 N4972

Subject:	Request to remove K1-6B6B from U+8C6C
Date:	2018-06-05 (originally submitted on 2018-06-03)
To:	WG2
From:	Jaemin Chung
Pages:	4

U+8C6C 豬—a character in the URO (U+4E00–U+9FA5)—seems to be a unification between 豕 (G, H, T, J, and V) and 豸 (K only), which is never acceptable.

8C6C  
豕 152.9

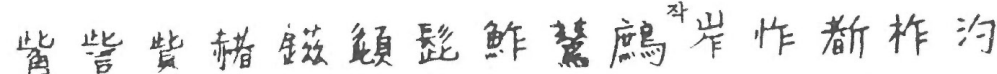
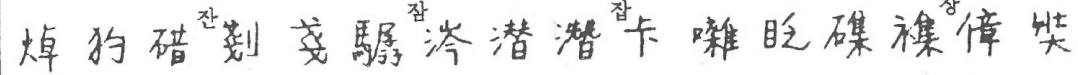

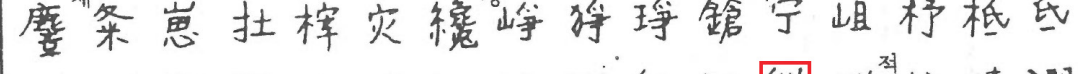
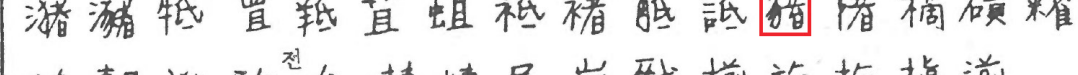
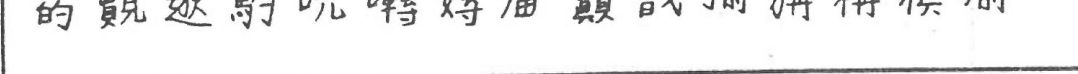
豬   豬   豬   豬   豬   豬  
 GE-3E4A   HB1-BDDE   T1-6F29   J0-6C37   K1-6B6B   V1-6773

I wondered why this kind of huge unification mistake is in the URO, but after checking previous versions of ISO/IEC 10646 and KS C 5657-1991, I now understand what happened.

**Section 1: What actually happened**

0x6B6B (75-75) in the original KS C 5657 (K1; now KS X 1002) does look like 猪者 (handwritten).

Excerpt from KS C 5657-1991

부호표상의 위치	16진 부호	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
제 75 행	6B-20																
	6B-30																
	6B-40																
	6B-50																
	6B-60																
6B-70																	

(Document continued on the next page)

However, when the URO was established in the early 1990s, Korea never submitted 豕者. Instead, it actually submitted a different character 豕者 under the K1-6B6B source reference.

Excerpt from ISO/IEC 10646.1-1993

140/108	<span style="font-size: 2em;">豬 豬 豬</span>						
8C6C	<table border="0"> <tr> <td>1-6F29</td> <td>0-6C37</td> <td>1-6B6B</td> </tr> <tr> <td>1-7909</td> <td>0-7623</td> <td>1-7575</td> </tr> </table>	1-6F29	0-6C37	1-6B6B	1-7909	0-7623	1-7575
1-6F29	0-6C37	1-6B6B					
1-7909	0-7623	1-7575					

The unification was clearly between 豕 and 豕 (acceptable), not between 豕 and 豕 (never acceptable). That is, there was no problem with the unification itself.

But later, Korea silently changed the glyph of K1-6B6B to 豕者 in the 2000 version (and all the subsequent versions) and causes a problem.

Excerpt from ISO/IEC 10646-1:2000

140/108	<span style="font-size: 2em;">豬 豬 豬 豬 豬</span>										
<b>8C6C</b>	<table border="0"> <tr> <td>E-3E4A</td> <td>1-6F29</td> <td>0-6C37</td> <td>1-6B6B</td> <td>1-6773</td> </tr> <tr> <td>E-3042</td> <td>1-7909</td> <td>0-7623</td> <td>1-7575</td> <td>1-7183</td> </tr> </table>	E-3E4A	1-6F29	0-6C37	1-6B6B	1-6773	E-3042	1-7909	0-7623	1-7575	1-7183
E-3E4A	1-6F29	0-6C37	1-6B6B	1-6773							
E-3042	1-7909	0-7623	1-7575	1-7183							

Note that the real 豕者 is encoded later at U+27CEF 豕.

27CEF	<span style="font-size: 2em;">豕 豕 豕</span>			
豕 153.9	<table border="0"> <tr> <td>UCS2003</td> <td>GKX-1202.34</td> <td>T7-335F</td> </tr> </table>	UCS2003	GKX-1202.34	T7-335F
UCS2003	GKX-1202.34	T7-335F		

(Document continued on the next page)

## Section 2: Proposed changes

It is very clear that K1-6B6B **cannot** stay at U+8C6C 豬, as Korea made a huge non-unifiable change. The K1-6B6B source reference at U+8C6C 豬 should either be simply dropped or be moved to U+27CEF 豸.

After this change, the UCS code chart should look something like this:

1) K1-6B6B simply dropped

	U+8C6C 豬				
8C6C 豕 152.9	豬	豬	豬	豬	豬
	GE-3E4A	HB1-BDDE	T1-6F29	J0-6C37	V1-6773
	U+27CEF 豸				
	27CEF 豸 153.9	豸	豸	豸	
		UCS2003	GKX-1202.34	T7-335F	

or

2) K1-6B6B moved to U+27CEF 豸

	U+8C6C 豬				
8C6C 豕 152.9	豬	豬	豬	豬	豬
	GE-3E4A	HB1-BDDE	T1-6F29	J0-6C37	V1-6773
	U+27CEF 豸				
	27CEF 豸 153.9	豸	豸	豸	
		UCS2003	GKX-1202.34	T7-335F	
		豸			
		K1-6B6B			

(Document continued on the next page)

### Section 3: Comments, possible issues, and arguments

I understand Korea's position that 𪛇者 is correct for K1-6B6B. Yes, K1-6B6B is 𪛇者. But Korea never submitted 𪛇者 – what it submitted was 𪛇者, **not** 𪛇者.

If the change in the 2000 version (and all the subsequent versions) was an attempt to correct a mistake, then what Korea should have done is to remove K1-6B6B from U+8C6C 豬 and to propose to encode K1-6B6B at a new code point. Keeping the source reference at the same code point while making a huge non-unifiable change is **not** the way to correct a mistake. This is simply wrong.

The current Korea NB members seem to be believing that what Korea originally submitted was 𪛇者 (Korea does not say anywhere that it originally submitted 𪛇者 and silently changed it to 𪛇者). I guess this is simply because the current members are unaware that this silent change was made before they joined the Korea NB. It is unfortunate that the current members need to deal with this even though this is not what they did.

Note that compatibility is not really an issue for this character.

- KS X 1002 (formerly KS C 5657) is just a paper standard in reality. There is no encoding which implemented it. (Actually, the main purpose of KS X 1002 was to put more hangul in the early versions of UCS.)
- Korean fonts which support hanja usually only cover KS X 1001; they usually don't cover KS X 1002. (Besides, lots of Korean fonts don't even cover hanja, since hanja is rarely used in Korean.)

Korea might complain

1) about the removal or remapping of K1-6B6B.

2) that a KS X 1002 character is outside the BMP after K1-6B6B is moved to U+27CEF 豬.

If so,

1) Korea cannot really complain about K1-6B6B being dropped or moved from U+8C6C 豬. This is clearly Korea's fault not submitting the correct character at the time (if Korea submitted the correct character at the time, this would not have happened). Submitting a different character and silently making a huge non-unifiable change only cause a problem and only pollute the UCS code charts. Some might even say that Korea is deceptive after reading this document.

2) what Korea should have done is to propose to encode 𪛇者 in the BMP **before** it is encoded in Extension B. Since the real 𪛇者 is already given a UCS code point, complaining about having a non-BMP character is meaningless.

Or, Korea might simply say “just no.” Please do **not** say “just no” – keeping K1-6B6B 𪛇者 at U+8C6C 豬 even after the details in this document are known only proves that Korea cannot be trusted.

Also, Korea **cannot** use “we already discussed this in 2010” as its reason (Korea said that it already discussed this in 2010 and the decision was to keep this as-is), since the details in this document were not known at the time.

(End of document)