

Title: Comments on several criteria to distinguish Shuowen Seal Glyphs (Feedback to WG2 N5118)

Source: suzuki toshiya











Date: 2019/12/01

WG2 N5118 gives some criteria to distinguish Seal glyphs are proposed, although its outline is almost same with the Oracle Bone PnP, the efforts by the authors are much appreciated. But, it is unclear whether the proposed idea is “such glyph difference is too significant to be unified (and they should be coded separately), even if the glyphs share same pronunciations and same semantics”, or, they are proposed to clarify the granularity how the experts recognize the glyphic difference. If the proposed idea is former (the criteria is non-unifiable glyphic difference), some of them are concerned to be too microscopic and too intolerant for the implementers.

I suggest the proposed criteria should be used to decide whether there is some recognizable glyphic difference. But it should not be used to decide whether 2 glyphs should be coded separately. As Ken Whistler gave an important comment in the last WG2#67 at Redmond; the identification of the character is expected to be determined with the consideration of glyph shape, phonetics and semantics.

1. Too microscopic disunification prevents future glyph correction

Just before the official proposal of Shuowen Seal as WG2 N5105, some glyph corrections were applied to the proposed character set. According to WG2 N5089 (the meeting report of the last Beijing meeting, 2019 May) “Q5 How the original broken or inconsistent designs were handled”, such kind of the glyphic differences should be handled by IVS, if the original broken glyphs are needed. The first correction is a replacement of 呂 by 呂.

	No.	Vol.	Serial No.	Serial No. in Tenghuaxie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Content
Need to restore	1	卷1	321	00303			莒	艸	12	Zheng-zhuan	依字理恢復原形 《說文解字·呂部》：「呂，聲骨也。象形。」
Component of “呂”	A.	卷5上	60	03235			宮			Zheng-zhuan	
	B.	卷6上	247	04071			稻			Zheng-zhuan	
	C.	卷7上	517	05124			呂			Zheng-zhuan	
	D.	卷12上	45	08391			閭			Zheng-zhuan	

A correction of 莒, by the replacement of 呂 by 呂 (WG2 N5089).

If the rules 1.2 (the number of components or lines are different...) and 1.5 (whether the same set of components are connected each other or not...) in WG2 N5118 are strictly applied, 吕 and 呂 should be coded separately. So, the original wrong glyph in THX for 莒 should be coded separately. In the case of 莒, fortunately the experts could correct the glyph before the publishing it in the standard character set. But if another correction is found to be needed after the standardization, too microscopic disunification would be a barrier against the correction, or it would introduce the incompatibility with the previous versions. For example:

- ① glyph A from THX is standardized.
- ② in later, glyph B from others, with same semantics, is coded separately, because of its glyphic difference.
- ③ in later, glyph A is found to be mistakenly designed and glyph B is found to be correct.

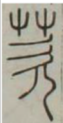

Some users would want to use glyph B at the codepoint of glyph A, to minimize the migration cost. How to resolve such request? To retain the possibility of the glyph correction in future, it is recommended that the separated encoding should be decided by more significant difference, which the users cannot remind one from another.

2. Semantic discussion is more useful to justify future glyph correction

I suggest a consideration of the semantic difference in the decision of the separate encoding. In this section, I give 3 examples: one of “should be disunified”, two of “unification would be acceptable”.

2.1. 茨 should not be designed as 𦵿, they have different semantics.

On the other hand, I have no strong objection to take following glyphic difference between original THX glyph (looking like 𦵿) and corrected 茨, as “non-unifiable”. It is because there is another SW Seal character, 𦵿, looking like as the original THX glyph of 茨. 𦵿 is annotated as “以茅葦蓋屋。从艸次聲。疾茲也”. On the other hand, 茨 is annotated as “雞頭也。从艸欠聲。巨險切”. They would not be exchangeable.

No.	Vol.	Serial No.	Serial No. in Tenghua xie version	Original glyph	TTF Font	Corresp Modern Char.	SW Radical	SW Radical Number	Type	Content
1	卷 1	650	00632			茨	艸	12	Zheng-zhuan	從艸次聲

Correction Record for “茨” (WG2 N5089).



Seal glyphs for “茨” .



Seal glyphs for “茨”

商務印書館石印本 and its 中國書店影印本 (4th cell) have already corrected “茨” .

2.2. Designing 莧 as 莧 would not guide the users to different character.

Back to the discussion for 莧, if there is any character looking like 𠂔+呂 with different semantics or phonetics, it would be considerable option to encode them separately. But there is no such. Is the Seal glyph of 𠂔+呂 (THX’s wrong glyph) is hard for the users to remind the Seal glyph of 𠂔+呂 (corrected glyph)? Maybe it is not so hard for the users of CJK Unified Ideographs. Among the cases listed in the section 1.2 and 1.4, there are no similar-but-non-cognate characters like the pair of 莧 and 茨.

I want to remind that the pair of 呂 and 呂 are listed in Annex S as the example of source separation. In the other words, their glyphic difference was regarded to be unifiable. In fact, G0 glyph shape of CJK Unified Ideograph is 𠂔+呂, instead of 𠂔+呂. This proves that the most users would not have serious problem in the character identification under the 呂:呂 unification. In comparison with CJK Unified Ideograph, Shuowen Seal character set would be more compact and stable,

i) Addition or omission of a minor stroke

步•步, 者•者, 臭•臭, 呂•呂, 單•單

Examples of Unifiable Glyphic Difference (ISO/IEC 10646 Annex S)








侶侶	T	呂呂	T	宮宮	T
4FA3 4FB6		5415 5442		5BAB 5BAE	

Examples of Source Separations (ISO/IEC 10646 Annex S)

8392 𦰩 140.7	莧	莧	莧	莧	莧
	G0-5C6C	HB1-B2F7	T1-5C7C	J13-7A77	K1-577A

Unified Glyphs of 莧

2.3. Glyphic difference of 𠂔 in THX and Duan Zhu.

Taking similar shape-phonetic-semantic attitude, the separate encoding of 𠂔  and its Duan Zhu (段注本) glyph  (proposed in WG2 N5118 Section 1.2) would be too intolerant. If the right components  and  are used as 2 different characters (with different pronunciation and semantics), the unification of them would make the users confused about its semantics and pronunciation. Thus, the separate encoding would be reasonable. But as far as I could checked,  is not a self-standing character in Shuowen (in Duan Zhu, it seems that the glyphic design of 𠂔 component is not always same,  in many cases,  in other cases. But I could not find any author's not why he had chosen one from two). Although the glyphic difference is visible, it would not be so hard for the users to remind one from another.

3. Summary

In summary, I suggest that the criteria proposed in WG2 N5118 should be used as “if 2 glyphs have no glyphic differences in these criteria, they should be unified even if they have different phonetics and semantics, because we cannot distinguish 2 characters visibly”, or, “to encode the non-cognate characters separately, there must be some glyphic difference in these criteria”. The expert group replaced several glyphs from THX version to non-THX versions, to enable distinctive design. For such discussion, the criteria in WG2 N5118 are very useful.

In the discussion of the encoding for historic scripts, often it is claimed that “over-unification would loose the secret knowledge hidden in the subtle glyphic difference; even if 2 glyphs are looking like almost same, there would be a possibility that future research would find 2 glyphs have semantic difference. To keep such possibility, the over-unification should not be applied”. But, as far as we are discussing about the authorized texts since Sung dynasty (like of DaXu, of XiaoXu and of Duan Zhu), always we have the annotation texts for all Shuowen Seal characters. We do not have to fear much about the mistakenly unifications of similar-but-non-cognate characters in Shuowen Seal encoding.

I remember similar proposal was discussed and almost accepted in the last Taipei meeting, but the annotated version of WG2 N5118 is not submitted yet.

(end of document)