

Title: Response to WG2n4834

Type: Individual Contribution to WG2 #66 and Small Seal Script Ad hoc Meeting

Source: TCA and China

Action: For consideration by JTC1/SC2/WG2 and UTC

Date: 2017-08-26

For the document WG2 N4834, Mr. Suzuki, who is an information expert from Japan, raised the following questions. TCA's response to the questions is as following :

On the Q1: Is the encoding of Seal as a separate script the best solution?

(1) Why do we propose to encode Shuowen Small Seal scripts separately?

The book Shuowen Jiezi (說文解字, abbreviated as Shuowen) isn't a usual record or dictionary. Instead, the Small Seal script is the final stage of the ancient Chinese characters. Shuowen which is academically recognized as a set of Small Seal scripts, had been standardized from Qin (秦) dynasty (221 B.C. to 206 B.C.), chosen and systematized by Xu Shen (許慎) in East Han (東漢) dynasty (25 to 220 D.C.), then finally proofread and corrected completely by Duan Yucai (段玉裁) in Qing (清) dynasty (1644 to 1911 D.C.). Since it has been compiled as a good set of oracle scripts, we don't think the Small Seal script has to be rearranges again at this moment. Even some odd Small Seal scripts discovered from tombs do not surpass the collection of Shuowen.

The importance of encoding Small Seal scripts is that it's significant base of all ancient Chinese characters. Shuowen's 540 radicals are the base of modern radicals. We used to classify ancient Chinese characters on its conventional practice.

(2) The Small Seal script was originated from the First Emperor of Qin. As the basis of modern ideographs, the Tang-Kai script was mature and popular in the Tang Dynasty, such as Ouyang Xun (歐陽詢), Yan Zhenqing (顏真卿) and Liu Gongquan (柳公權) and other people's posters have become the calligraphic paradigms. There was a time distance of at least 800 years between the Small Seal script and the Tang-Kai script. After a long period of evolution and at least two times of the script revolutions (Li-transcript from the Small Seal script, and Kai-transcript from the Li script), the Small Seal script and the modern ideograph had long been derived into two different kinds of writing scripts, such as different writing methods, different radicals, different strokes and different components. It means that we

cannot to write a Small Seal script by means of the strokes with associated writing sequence of the modern ideographs, and vice versa. In addition, there is no the one-to-one relationship between the Small Seal script(s) and the modern ideograph(s).

There are some examples to illustrate the essential differences among the Small Seal script and the modern ideograph.

- A. Component replacing: while a modern ideograph Li-/Kai-transcript is from a Small Seal script, one or more components were replaced by other(s)

Example 1, the corresponding Small Seal script of the ideograph “帆” is “𠩺”, but the correct ideograph that Li-/Kai-transcript form the Small Seal script “𠩺” is the ideograph “𠩺”, as shown in figure 1.

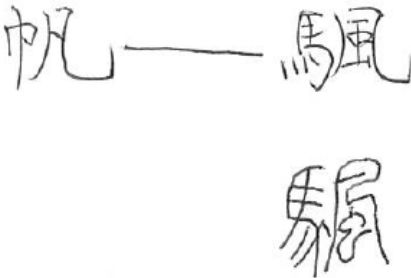


figure 1.

- B. Component Diverted: while a modern ideograph Li-/Kai-transcript is from a Small Seal script, one or more components were diverted by other(s).

Example 2, the corresponding Small Seal script of the ideograph “襍” is “𠩺”, but the correct ideograph that Li-/Kai-transcript form the Small Seal script “𠩺” is the ideograph “襍”, as shown in figure 2.



Figure 2

- C. Component reducing: while a modern ideograph Li-/Kai-transcript is from a Small Seal script, one or more components were reduced.



Example 3, the corresponding Small Seal script of the ideograph “集” is “”, but the correct ideograph that Li-/Kai-transcript form the Small Seal script “” is the ideograph “彙”, i.e., three of “隹” was reduced to be one, as shown in figure 3.



Figure 3

- D. Many-to-one mapping: two or more Small Seal scripts were mapped to one modern ideograph during Li-/Kai-transcript.




Example 4, two Small Seal scripts “” and “” were mapped to the same modern ideography “鬱” during their Li-/Kai-transcript, as shown in figure 4.



Figure 4

- E. Functionary separation (職用分工): while a modern ideograph Li-/Kai-transcript is from a Small Seal script, its meaning and usage were something different from the origin.

Example 5, for the Small Seal script “”, its meaning is same as the modern ideograph “竭” (all out), while the meaning of its corresponding ideograph “渴” was somehow changed to be “thirsty”, as shown in figure 5.

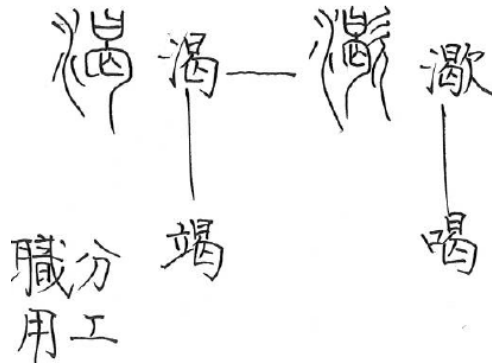


Figure 5

On the Q2: How should duplicate glyphs be handled in a standard encoding?

- (1) The definition of “duplicate glyphs”: In a particular version of “Shuowen (說文)”, such as “Tenghua Xie Ben (藤花榭本)”, the same Small Seal scripts were presented with two different radicals. Such as 「否」, is put in the radical 「口 and 不」.
- (2) For the duplicate Small Seal script, TCA suggests there is only one encoding for them, but they can be two different radicals as their attributes. Such as 「否」, the radical will be 「口 and 不」. The following group from “Tenghua Xie Ben” may be of duplicate words.

司	右	873	議刪	2-1205
司		1894		3-1000
吁	吁	912		2-1248
吁		3053	議刪	4-1272
教	教	2502	議刪	4-559
教		3846		5-1013

吹	吹	828		2-1147
吹		5499	議刪	7-790

8	否	否	933	議刪	2-1271
	否		7665		9-950

Q3: Compatibility with existing reference and research implementations.

(1) Why we do not choose other versions, but Tenghua Xie Ben (藤花榭本)?

The earliest versions of Shuowen are residual volumes of radical wood and radical mouth of Tang dynasty handwritten book (唐寫本木部殘卷、口部殘卷), they only include about 200 characters, which are 2% of Shuowen. Furthermore, three of existing Song (宋) dynasty (960 to 1279) versions, small characters version by Wang Chang of Qingpu (青浦王昶宋小字本), small characters version by Wang Zhong (汪中所藏宋小字本) and version by Huang Zhichun (黃氏志淳本), they may originate from the same sources. And their glyphs have more damages, disappearances or rough scribes. Therefore, they can't be materials of design of Small Seal encoding.

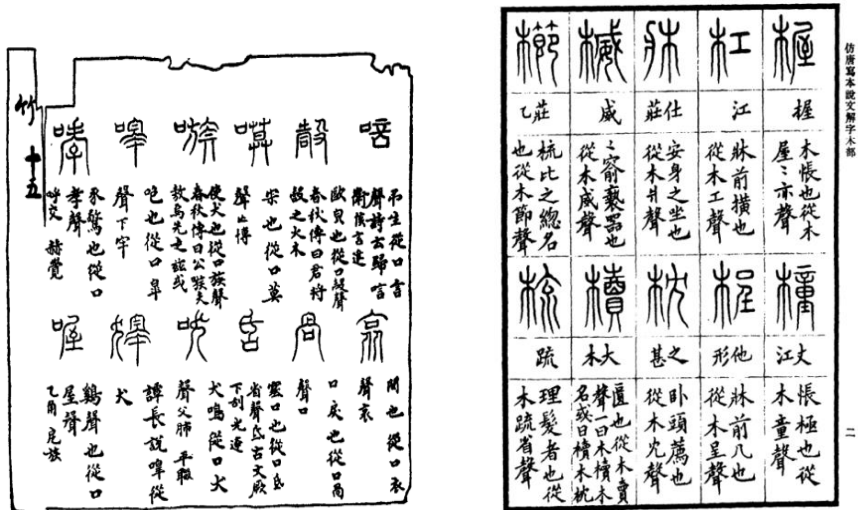


Figure 1 唐寫本木部殘卷 and 口部殘卷

Currently, three of the accepted better versions are Jiguge version(汲古閣本), Tenghuaxie version (THX, 藤花榭本) and Pingjingguan version(PJG,平津館本), which are based on Song dynasty version. So we can compare the others versions as below:



Figure 2 THX(left) and PJG(right).

Obviously, THX’s glyph figures are much clearer than PJG’s. That’s the reason why we choose THX at very first. Second, we will expand the glyph resting on DYC, PJG, XiaoXu version, and so on. Duan Yucai is regarded as the greatest expert of Shuowen in the Qing dynasty. He changed some components of glyph systematically, that is

valuable.

- (2) TCA and China convened a meeting of philological experts to discuss encoding of Small Seal scripts. Our experts include those from China: such as 王寧(Wang Ning), 王立軍(Wang Liun), 王曉明(Wang Xiaoming), 胡佳佳(Hu Jiajia), 師霞(Bu Shixia), 凌麗君(Ling Lijun), 孟琢(Meng Zhuo), 董婧辰(Dong Jingchen) ; TCA: 李璫(Xian Li), 蔡信發(Xin-fa Tsai), 季旭昇(Hsiu-sheng Chi), 許學仁(Xue-jen Hsu), 宋建華(Chien-hua), 魏林梅(Lin-mei Wei), as well as the experts from the Project of Chinese Character Repertoire are all agreed that the Tenghua Xie Ben (藤花樹本) is the best version for the text of the Small Seal as the international code.
- (3) The experts from China and TCA have agreed that THX is sufficient as the baseline for encoding. The differences between THX and other materials are minor enough for determining the first collection of unified characters to encode. Therefore, TCA and China do not think there is a need to provide further justification for use of THX.

Conclusion

The Small Seal is the base of ancient character textual criticism and can't be replaced. It's a certified character set from the academic community. To respect historic text, there are two requests below:

- (1) Encode Shuowenjiezi unified glyphs in Small Seal Block independent of CJK Unified Ideographs.
- (2) We anticipate encoding it in 2 phases. First, encode the Small Seal glyph of THX. Second, expand characters different from THX in Chen Changzhi version (陳昌志本), DYC, PJG, XiaoXu version (小徐本) and so on.

Experts of the working group

China: 王寧(Wang Ning), 王立軍(Wang Liun), 王曉明(Wang Xiaoming), 胡佳佳(Hu Jiajia), 師霞(Bu Shixia), 凌麗君(Ling Lijun), 孟琢(Meng Zhuo), 董婧辰(Dong Jingchen)

TCA: 李璫(Xian Li), 季旭昇(Hsiu-sheng Chi), 許學仁(Xue-jen Hsu), 宋建華(Chien-hua), 魏林梅(Lin-mei Wei),

Organization: Chinese Foundation for Digitization Technology,

The Project of Chinese Character Repertoire, China Electronics
Standardization Institute

~end of document~

針對 Document WG2n4834，日本資訊專家鈴木俊哉先生提供以下議題，TCA 提出以下立場作為回應。

第一個問題：小篆說文納入編碼是最好的解決方案嗎？

1. 為什麼《說文》小篆要單獨編碼？

所提出的問題，是因為對《說文》小篆的基本性質沒有充分瞭解。《說文解字》不僅是一般的傳世文獻，更不同於是一般意義的字書。小篆是中國古文字的過渡到今文字（現代漢字）階段的關鍵最後階段。它經過秦代的規範，又經過許慎的《說文解字》的字形優選、編排整理，和段玉裁的全面考據、校改，而成為學界公認的小篆字集。它相當於已經整理好的甲骨字集，不需要再行整理。目前能夠看到的零星小篆，例如：秦權量、《泰山刻石》、尚有爭議的《嶧山碑》，及歷代墓誌碑的碑額用字等等，都不能超越過《說文解字》之所集。

小篆國際編碼的重要意義，在於它的構形是一切古文字考訂據的重要依據。而且《說文解字》據形系聯的540部部首，是後來辭書歸部分為214部、201部的源頭基礎，按照540部的順序，編輯各代古文字文字編或辭書，已經是整理古文字的常規。就《說文》小篆的進行編碼，自應當先居於其他古文字之編碼的首位。

小篆係源自秦始皇時期，現代漢字引為以楷書為基礎的以楷書現代漢字，則成熟並普及於唐代，例如歐陽詢、顏真卿和柳公權等人的字帖，成為歷代文人的習字範本。小篆與唐代楷書間相距至少800年，其間形體演變至少歷經長時間的演變及至少兩次文字整理革命（小篆轉換為隸之變書、隸書轉換為楷之變書），小篆和現代漢字早已分化成兩種不同的書寫系統。文字：書寫方式不同、部首不同、筆畫不同、組件不同，亦即無法使用現代漢字的筆畫劃及筆順書寫小篆，反之亦然。此外，小篆與現代漢字間並不確實存在一對一關係。

下列數例係用以說明小篆與現代漢字間之主要差異：

A. 部件替換：從小篆經隸楷定並逐漸形成現代漢字的過程中，某些部件已被替換。

例 1：「帆」所對應之小篆為「𣎵」，但「𣎵」的楷定字卻應為「馬風」，如 figure 1 所示。

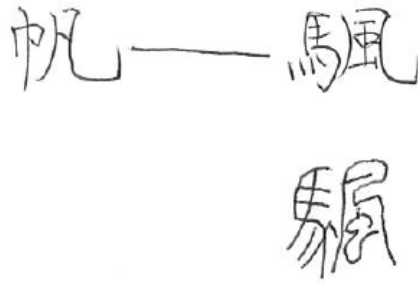


圖 1

B. 部件挪移：從小篆經隸楷定並逐漸形成現代漢字的過程中，某些部件形位已被挪移。

例 2：「雜」所對應之小篆為「𧯛」，但「𧯛」的楷定字卻應為「襪」，如 figure 2 所示。



圖 2

C. 部件減省：從小篆經隸楷定並逐漸形成現代漢字的過程中，某些部件已被省略。

例 3：「集」的小篆為「𧯛」，亦即原先小篆中的三個「隹」被省略成一個「隹」，如 figure 3 所示。



圖 3

D. 多對一隸楷定：從小篆經隸楷定並逐漸形成現代漢字的過程中，兩或多個小篆被對應至同一個現代漢字。

例 4：「𣎵」與「𣎵」兩個小篆的楷定字皆為「鬱」，如 figure 4 所示。



圖 4

E. 職用分工：從小篆經隸楷定並逐漸形成現代漢字的過程中，語意及使用場合產生變異。

例 5：小篆「𣎵」原意為「竭」，但其楷定字「渴」的語意卻已不同於「竭」，如 figure 5 所示。

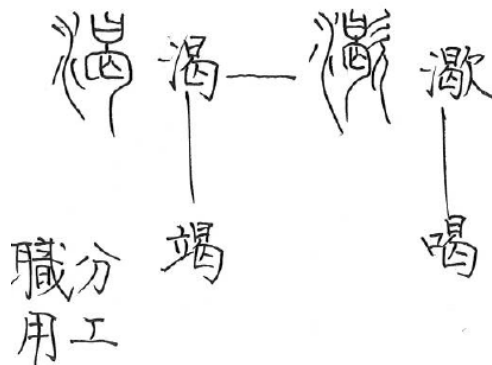


圖 5

第二個問題：如何在標準編碼中處理重複字形？

- (3) 小篆重複字之定義：為在同一說文小篆版本，同一個小篆字形歸在不同部首之下。如「否」字完全相同之小篆字形，分別歸入「口和不」兩個部首中。
- (4) 小篆之重複字可擇一編碼，但該小篆之部首屬性可以收兩個，如「否」字部首可收「口及不」2 個部首。下列為藤花樹本之重複字

司	右	873	議刪	2-1205
司		1894		3-1000
吁	吁	912		2-1248
吁		3053	議刪	4-1272
教	教	2502	議刪	4-559
教		3846		5-1013

吹	吹	828		2-1147
吹		5499	議刪	7-790

8	否	否	933	議刪	2-1271
	否		7665		9-950

第三個問題：與現有的參考和研究實現相容性。

(1)為什麼選取藤花榭本，而不是其他版本？

現存最早的《說文》版本是唐寫本木部殘卷和口部殘卷，二者相加僅存 200 字，占《說文》全書的 2% 左右。而現存的三個宋刻遞修本青浦王昶宋小字本、汪中所藏宋小字本、“黃氏志淳”本，三種版本據考可能同源於一本或相互有傳承關係，它們或字跡殘損脫漏較多，或雕刻粗糙，無法作為國際編碼的底本。

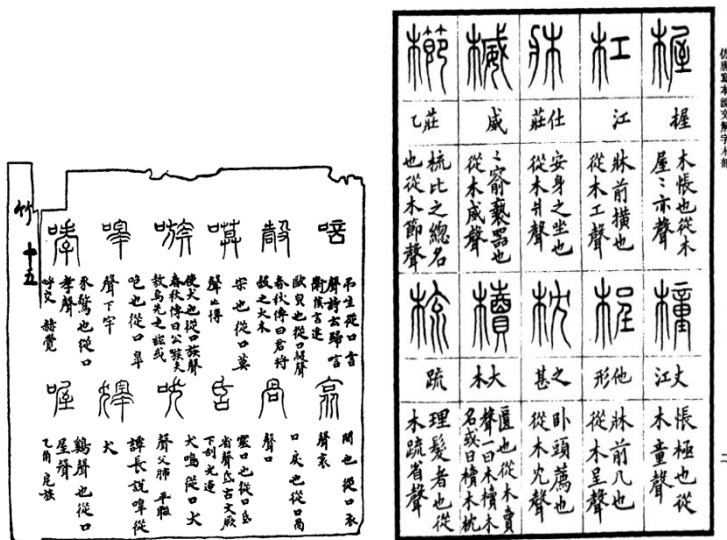


圖 2 唐寫本木部殘卷和口部殘卷

現存公認較好的三本《說文》版本是以宋本為底本校刻的汲古閣本、藤花榭本、平津館本。其中“毛本之病，在子晉之子斧季妄改、剗版，致多誤處，則人未之知也”（段玉裁跋），而其餘兩種版本的情況對比如下：



圖 3 《說文》藤花榭本（左）和平津館本（右）

顯然，在字形的清晰規範程度上，平津館本遠不及藤花榭本。這也是我們第一步選定藤花榭本作為底本的原因。第二步我們再以段注本為主要依據，對藤花榭本字形進行補充。段玉裁作為清代最有成就的《說文》學家，他從系統的角度對小篆字形所做的局部改造，還是具有重要參考價值的。

(2)根據兩岸專家一致共識(包括中國小篆學者代表王寧(Wang Ning),王立軍(Wang Liun),王曉明(Wang Xiaoming),胡佳佳(Hu Jiajia),師霞(Bu Shixia),凌麗君(Ling Lijun),孟琢(Meng Zhuo),董婧辰(Dong Jingchen);臺灣說文小篆學者代表李鏐(Xian Li),蔡信發(Xin-fa Tsai),許學仁(Xue-jen Hsu),季旭昇(Hsiu-sheng Chi),宋建華(Chien-hua),袁國華(Kwok-wa Yuen),魏林梅(Lin-mei Wei),及中國字庫工程皆一致認為藤花榭本為說文小篆納入國際編碼最好的版本。

結論：

綜上所述，我們希望可以以尊重文本為大前提，在不改動古人撰字漢文 獻的情況下將說文小篆字編碼，這對於學術、研究、出版都是有益處的。我們仍建提議維持依照原有的提案，採以下列兩個步驟方式進行，首先以藤花榭本為造字的底本，進行編碼。第二步驟再將其他說文小篆材料與藤花榭本對校比較，將藤花榭本沒有收錄的字型，提交 WG2 可再另行收錄編碼。

