

Type: Working Group Document
Title: Mapping Principles of Small Seal Glyphs in Different Shuowen Versions
Source: TCA and China
Reference: [N5306](#)
Date: 2025-06-12

1. General Principles:

- (1) Both the Tenghuaxie version (THX) and the Chenchangzhi version (CCZ) belong to the Daxu Ben (大徐本, the version of Big Xu). They are basically consistent in glyphs, explanations, and ordering. The Xiaoxu Ben (小徐本, the version of Little Xu) and the Duan Zhu (段注) version differ from the Daxu Ben in their views on certain characters, resulting in differences in character forms, explanations, or ordering.
- (2) The coding sequence follows the order of Zhengzhuan (canonical Small Seal Script characters, 正篆), Chongwen (variant characters, 重文), and newly added characters (新附字) as they appear in the Tenghuaxie version of Daxu Ben. For character forms that cannot be matched with the Tenghuaxie version of Daxu Ben, they are appended after the last character in THX, following the sequential order of Chenchangzhi version (CCZ), Xiaoxu Ben (QJZ), and then the Duan Zhu (DYC).
- (3) In terms of written forms, Small Seal script includes Zhengzhuan(正篆), Guwen(古文), and Zhouwen(籀文). The unification principle only discusses the relationship between character forms within the same style system; characters from different style systems are always processed as separate characters. Examples:

Type	Code chart	Disunified/ Unified
Zhengzhuan(正篆)	38022 𠂔 3 神 795E    	K-00035 : Disunified Principle 3.(1)

Type	Code chart	Disunified/ Unified
Zhengzhuan(正篆)	38299 𦰩 12 蒞 44DC TH-00645 C-00663 K-00644 D-00641	K-00644 : Unified Principle 2.(4)
Guwen(古文)	38692 𦰩 56 謀 8B00 TH-01617 C-01682 K-01622 D-01605	K-01622 : Disunified Principle 3.(2)
Guwen(古文)	3863A 𦰩 44 冊 518A TH-01529 C-01594 K-01536 D-01518	K-01536 : Unified Principle 2.(4)
Zhouwen(籀文)	38083 𦰩 6 璫 74BF 𦰩 53E1 TH-00127 C-00131 K-00131 D-00125	K、D : Disunified Principle 3.(1)
Zhouwen(籀文)	388AB 𦰩 82 畫 665D TH-02143 C-02222 K-02148 D-02127	D-02127 : Unified Principle 2.(4)

(4) Characters placed at the same code position must be consistent in both form and explanation. However, if contradictions exist between form and explanation, then the character is based on the form. Examples:

A. D-04654 should be moved from U+392CB to U+39302.

Code chart	Explanation
392CB 𦰩 231 暉 6688 TH-04666 C-04807 K-04676 D-04654	光也。从日，軍聲。
39302 𦰩 231 暉 6688 TH-X149 C-04862	日月气也。从日，軍聲。

B. D-09753 and TH-06565 share the same code point.

Code chart	39AA7 𦰩 357 𦰩 78FA TH-06565 C-06812 K-06577 D-09753	
Explanation	古文「𦰩」。《周禮》有𦰩人。	古文「卵」。

2. Unification Principle :

(1) Taboo Glyphs

Such as C-00023, K-00022, D-00023 of U+38016, are all unified.

Code chart	Disunified/ Unified
38016 禱 3 禱 799B TH-00023    	C、K、D : Unified Principle 2.(1)

(2) Those with incorrect seal script forms:

There are two criteria for determining errors:

a. The glyphs are inconsistent with the explanation.

Example: K-03185, in the explanation from 'K' source is “从竹，『前』聲”，but the glyph shows as “从竹『寿』聲”.

Code chart	Disunified/ Unified
38CD3 𦰩 143 箭 7BAD TH-03177    	K : Unified Principle 2.(2a)

b. Components do not maintain systematic consistency.

Example 1: In the Duan Zhu, the component of the character "歲" is consistently shown as "歲", but in D-00947 it appears as "𪛗", missing a horizontal stroke. so D-00947 can reasonably be considered incorrect and does not need to be encoded separately.

Code chart	Disunified/ Unified
383D9 𪛗 22 𪛗 5666 TH-00952    	D : Unified Principle 2.(2b)

Example 2: In the Duan Zhu, the component of the character "今" is consistently shown as "今", but in D-03638 it appears as "𠂔", with a different turning stroke direction, so D-03638 can reasonably be considered incorrect and does not need to be encoded separately.

Code chart	Disunified/ Unified
38EBA △ 181 今 4ECA TH-03651 C-03767 K-03660 D-03638	D : Unified Principle 2.(2b)

(3) Components are systematically modified in form

Such as "門" in K, D sources, and "言" in D source, etc. Examples :

Code chart	Disunified/ Unified
3A20D 門 438* 門 9580 TH-08382 C-08700 K-08395 D-08371	K,D: Unified Principle 2.(3)
3A20F 門 438 關 95C8 TH-08384 C-08702 K-08397 D-08373	K,D: Unified Principle 2.(3)
3866F 言 56* 言 8A00 TH-01582 C-01647 K-01589 D-01571	D: Unified Principle 2.(3)
38672 言 56 語 8A9E TH-01585 C-01650 K-01592 D-01574	D: Unified Principle 2.(3)

(4) Compound characters with intercharacter relationships that have only minor differences (variant forms). Examples:

Type	Code chart	Disunified/ Unified
Zhengzhuān(正篆)	38057 𠂔 3 崇 795F TH-00088 C-00088 K-00088 D-00088	K : Unified Principle 2.(4)
Guwén(古文)	39A56 𠂔 353 廋 5EC4 TH-06491 C-06731 K-06501 D-06479	KD : Unified Principle 2.(4)

(5) For phono-semantic compound characters:

- A. In characters with a left-right structure, the normal pattern is that the left component is the semantic element (indicating meaning) and the right component is the phonetic element (representing pronunciation). If the structure has the phonetic element on the left and the semantic element on the right, it can be unified.
- B. In characters with a top-bottom structure, if they have different structure—with either the semantic element on top and phonetic element below, or the phonetic element on top and semantic element below—but the meaning remains the same, they can be unified.

Type	Code chart	Disunified/ Unified
Zhouwen(籀文)	39B5F 𠂔 370 𠂔 4B94 TH-06740 C-06996 K-06751 D-06727	K-06751 : Unified Principle 2.(5)
Guwen(古文)	3A264 𠂔 439 𠂔 805E TH-08464 C-08787 K-08476 D-08453	K-06758 : Unified Principle 2.(5)
Zhengzhuan(正篆)	38AA2 𠂔 119 𠂔 9DDA TH-02633 C-02723 K-02641 D-02618	K-02641 : Unified Principle 2.(5)
Zhengzhuan(正篆)	39A21 𠂔 350 𠂔 21DE5 TH-06450 C-06678 K-06461 D-06440	K-02641 : Unified Principle 2.(5)

(6) Processing of Guwen or Zhouwen Unification:

When the correctness of character forms in Guwen or Zhouwen is difficult to confirm due to lack of explanation, keep unification without creating new code points. For example:

Code chart	Disunified/ Unified
38640 𠂔 45 𠂔 569A TH-01535 C-01600 K-01542 D-01524	C-01600 : Unified
39EF8 𠂔 410 𠂔 6F22 TH-07626 C-07913 K-07641 D-07614	C-07913 : Unified

3. Disunification Principles :

(1) Same pronunciation and meaning, but with differences in components. Examples:

a. 藍、濫

Code chart	38297 𠂔 12 濫 862B TH-00644 C-00662 K-00643 38298 𠂔 12 濫 862B KD-38298 D-00640	D-00640 : Disunified Principle 3.(1)
Interpretation	瓜蒚也。从艸，監聲。 瓜蒚也。从艸。濫聲。 段玉裁： 各本篆作藍。解誤作監聲。 今依廣韻、集韻訂。魯甘切。八部。	

b. 𡩺、𡩻

Code chart	394D1 𡩺 272 𡩻 7A8C  TH-05152  C-05323  K-05161  D-05139	D-05139 : Disunified Principle 3.(1)
Interpretation	𡩺也。从穴，卯 (mao3)聲。 𡩻也。从穴。𡩻(you3) 聲。 段玉裁：「『𡩻』聲」各本作『卯』聲」。今正。	

(2)Compound characters with the same components and the same number of components, but with different structure. Examples:

Type	Code chart	Disunified/ Unified
Zhengzhuan (正篆)	386B3 𡩺 56 𡩻 85F9  TH-01650  C-01715  K-01653  D-01636	K-01653、D-01636 : Dsunified Principle 3.(2)

(3)A singlet characters with the same pronunciation and meaning, but with differences in stroke count or stroke relationship (intersecting, connecting, or separate). Example:

Type	Code chart	Disunified/ Unified
Guwen(古文)	39BDC 𡩺 372 𡩻 9E97  TH-06858  C-07119  K-06869  D-06848	K-06869、D-06848 : Disunified Principle 3.(3)

不同《說文》版本中字符的對應原則

1.通則：

- (1) 藤花樹本和陳昌治本都屬於大徐本《說文》，無論是在字形、說解還是排序上，都是基本一致的。小徐本和段注，在少數字的見解上與大徐本不同，從而造成它們在字形、說解或排序上的不同。
- (2) 編碼次序依照藤花樹本大徐《說文》中出現的正篆、重文、新附字的順序排列。如有不能與藤花樹本大徐《說文》對應的形體，則增補於藤本最末字之後，依序按陳昌治本、小徐本、段注本的順序補入。
- (3) 《說文》小篆從書寫形體上來說，包括了篆文(包括正篆和或體)、古文、籀文。認同原則僅討論同一形制下的形體關係，不同形制的一律做不同的字符來處理。例如：

Type	Code chart	Disunified/ Unified
Zhengzhuan(正篆)	38022 𠂔 3 神 795E    	K-00035 : Disunified 原則 3.(1)
Zhengzhuan(正篆)	38299 𦰩 12 蒞 44DC    	K-00644 : Unified 原則 2.(4)
Guwen(古文)	38692 𠂔 56 謀 8B00    	K-01622 : Disunified 原則 3.(2)
Guwen(古文)	3863A 𦰩 44 冊 518A    	K-01536 : Unified 原則 2.(4)
Zhouwen(籀文)	38083 𠂔 6 璽 74BF 𠂔 53E1    	K、D : Disunified 原則 3.(1)
Zhouwen(籀文)	388AB 𠂔 82 𠂔 665D    	D-02127 : Unified 原則 2.(4)

- (4) 放在同一碼位的字，理論上必須形體與說解都一致。但如果存在形體和說解上的矛盾，則依形。如：

A. 392CB 和 39302, D-04654 應與 TH-X149 同碼位。

Code chart	Interpretation
392CB 日 231 暈 6688    	光也 。从日，軍聲。
39302 日 231 暈 6688   	日月气也 。从日，軍聲。

B. D-09753 與 TH-06565 同碼位

Code chart		
39AA7 尸 357 礦 78FA    		
Explanation	古文「礦」 。《周禮》有卅人。	古文「卵」 。

2. 認同原則：

(1) 避諱字，如 U+38016 的 C-00023、K-00022、D-00023，皆為認同

Code chart	Disunified/ Unified
38016 示 3 禩 799B    	C、K、D：Unified 原則 2.(1)

(2) 篆形錯誤者：

判斷錯誤的基準有兩點：

a. 字形與說解不一致。例：

Code chart	Disunified/ Unified
38CD3 竹 143 箭 7BAD    	K：Unified 原則 2.(2)

K-03185，在《小徐本(K)》的說解是「从竹，『前』聲」，而字形卻是「从竹『𠂔』聲」。

b. 構件不符合系統一致性。

例 1：在《段注(D)》的「歲」字部件都是，但 D-00947 為，缺一橫筆，所以 D-00947 有誤，可認同。

Code chart	Disunified/ Unified
383D9     𠂔 22 歲 5666 TH-00952 C-00985 K-00958 D-00947	D : Unified 原則 2.(2)

例 2：在《段注(D)》的「今」字部件都是，但 D-03638 為，折筆的方向不同，所以 D-03638 有誤，可認同。

Code chart	Disunified/ Unified
38EBA     △ 181 今 4ECA TH-03651 C-03767 K-03660 D-03638	D : Unified 原則 2.(2)

(3) 其他版本的部件為系統性改形，如《小徐本(K)》、《段注本(D)》的「門」、《段注本(D)》的「言」等。例：

Code chart	Disunified/ Unified
3A20D     門 438* 門 9580 TH-08382 C-08700 K-08395 D-08371	KD : Unified 原則 2.(3)
3A20F     闕 438 闕 95C8 TH-08384 C-08702 K-08397 D-08373	KD : Unified 原則 2.(3)
3866F     言 56* 言 8A00 TH-01582 C-01647 K-01589 D-01571	D : Unified 原則 2.(3)
38672     語 56 語 8A9E TH-01585 C-01650 K-01592 D-01574	D : Unified

	原則 2.(3)
--	----------

(4) 合體字，有字際關係，而僅有細微差異(異寫字)者，可認同：

Type	Code chart	Disunified/ Unified
Zhengzhuan(正篆)	38057  3 崇 795F TH-00088 C-00088 K-00088 D-00088	K : Unified 原則 2.(4)
Guwen(古文)	39A56  353 脊 5EC4 TH-06491 C-06731 K-06501 D-06479	KD : Unified 原則 2.(4)

(5) 形聲字(phono-semantic compound)，字形為左右結構時，通常情形是左為形符(表示字義)右為聲符(代表字音)，如果結構是左聲右形，可以認同。

字形為上下結構時，結構如果有上形下聲、上聲下形的不同，但意義沒有不同時，可給予認同。

Type	Code chart	Disunified/ Unified
Zhouwen(籀文)	39B5F  370 𦘒 4B94 TH-06740 C-06996 K-06751 D-06727	K-06751 : Unified
Guwen(古文)	3A264  439 𦘒 805E TH-08464 C-08787 K-08476 D-08453	K-06758 : Unified
Zhengzhuan(正篆)	38AA2  119 𦘒 9DDA TH-02633 C-02723 K-02641 D-02618	K-02641 : Unified
Zhengzhuan(正篆)	39A21  350 𦘒 21DE5 TH-06450 C-06678 K-06461 D-06440	K-02641 : Unified

(6) 古文、籀文因沒有說解，較難以確認字形正確性時，維持認同，不另編新碼。例如：

Code chart	Disunified/ Unified
38640  45 𦘒 569A TH-01535 C-01600 K-01542 D-01524	C-01600 : Unified
39EF8  410 𦘒 6F22 TH-07626 C-07913 K-07641 D-07614	C-07913 : Unified

3.不認同原則：

(1)音義相同，部件有差異，如：

a.藍、濫

Code chart	38297 艸 12 藍 862B 38298 艸 12 濫 862B TH-00644 C-00662 K-00643 KD-38298 D-00640	D-00640 : Disunified Principle 3.(1)
Interpretation	瓜蒞也。从艸，監聲。 瓜蒞也。从艸。濫聲。 段玉裁： 各本篆作藍。解誤作監聲。 今依廣韻、集韻訂。魯甘切。八部。	

b. 窳、窳

Code chart	394D1 穴 272 窳 7A8C TH-05152 C-05323 K-05161 D-05139	D-05139 : Disunified Principle 3.(1)
Interpretation	窳也。从穴，卯(mao3)聲。 窳也。从穴。卯(you3)聲。 段玉裁：「『卯』聲」各本作「卯」聲。今正。	

(2)合體字，構件和構件數都相同，但構件的位置不同，不認同。如：

Type	Code chart	Disunified/ Unified
Zhengzhuan(正篆)	386B3 言 56 謫 85F9 TH-01650 C-01715 K-01653 D-01636	K-01653、D-01636 : Disunified 原則 3.(2)

(3)獨體字，音義都相同，筆畫數或筆畫關係（相交，相接，相離）有差異。
如：

Type	Code chart	Disunified/ Unified
Guwen(古文)	39BDC 巛 372 麗 9E97 TH-06858 C-07119 K-06869 D-06848	K-06869、D-06848 : Disunified 原則 3.(3)