

**Type:** Working Group Document

**Title:** Preliminary Draft of Mapping Table for Shuowen Seal: from CCZ to N5191

**Source:** suzuki toshiya

**Reference:** [N4688](#), [N5119](#), [N5133 \(data\)](#), [N5187](#), [N5190](#), and

**Date:** [N5191 2023-06-16](#)

## 1. Background

According to WG2 N5119, the report of the Shuowen Seal Ad Hoc group meeting 2019-10-01, the experts formed a consensus to produce the mapping table between the Tonghuaxie version (THX) of Shuowen Jiezi (SWJZ) and Chen Changzhi version (CCZ), which can be a source reference like [CJKSrc.txt](#), [EmojiSrc.txt](#), or [TangutSrc.txt](#). After submitting the full comparison chart between CCZ and the proposed character set N4688 on 2019-10-18, N5133 (2020-04-01) listed the arguable glyph differences. N5187 (2022-06-20) was the response from China and TCA experts; it accepted the glyph distinction by VS instead of the separate encoding for most glyph differences in N5133 if required. Also, there are many modifications of the glyphs of the proposed character set, and they resolved the glyph differences with CCZ. In some cases, the proposed character set keeps the current glyph shape with the differences with CCZ, but no suggestion of their separate encoding. N5190 (2022-02-20) summarizes the glyph changes from the original THX to the latest proposal N5191; some modifications are for the differences listed in N5133, and others are not. Figure 1 summarizes the relationship between N5133 and N5190.



Figure 1: Relationship of the glyph difference and modifications in N5133 and N5190

## 2. Draft Mapping Table from CCZ to N5191, the Shuowen Seal set based on THX

Considering that N5187 accepted to use of VS to handle the subtle glyph differences in N5133 (especially the type A and B in N5133 categorization), it is possible to draft the major part of the mapping table. This proposal is the draft mapping table for the action item agreed upon in N5119. Some TBD entries exist, but 10454 mapped entries in 11108 are already stable. The stable part would be safe to include for the standard when the standard has the first revision of the code chart of Shuowen Seal.

	A	B	C	D	E	F	G
1	seq_CCZ	id_N5191_lexical	id_N5191_glyphic	map_type	corresp_UCS	N5133	N5189
2	00001	00001	-	Unify	一	-	-
3	00002	00002	-	Unify	弌	-	-
4	00003	00003	-	Unify	元	-	-
5	00004	00004	-	Unify	干	-	-

Figure 2: Preliminary Draft Mapping Table from CCZ to WG2 N5191

① seq\_CCZ

The sequential number of the heading Seal glyphs of CCZ, the format is 5 digits with 0 paddings, 00001..11108.

## ② id\_N5191\_lexical

The id assigned to the WG2 N5191 entry corresponds to the id\_CCZ in lexical order. There are 2 groups of glyphs: the Seal glyphs that Xu Xuan believed existed before his restructure and the Seal glyphs that Xu Xuan added explicitly (so-called "newly added character," 新附字). The format for the former group is 5 digits with 0-padding, 00001..10706. The syntax for the latter group is X-prefixed 3 digits with 0 paddings, [Xx]001..[Xx]402. The prefix "X" is uppercase in some entries or lowercase in others. Although N5191 does not describe the reason for distinguishing the uppercase prefix "X001" and the lowercase prefix "x301", the values are the same with N5191.

## ③ id\_N5191\_glyph

The id assigned to the WG2 N5191 entry, whose glyph has better similarity with the entry of id\_N5191\_lexical, with the same syntax. There are 2 cases where id\_N5191\_glyph exists; one is that exactly duplicated glyphs in SWJZ (principle #2 in N5189), and another is that the N5191 glyph is tweaked to avoid the shape conflict with other non-cognate entry (the principle #5 in N5189). For the former case, the id\_N5191\_lexical is empty because the proposed character set in N5191 removed the exactly duplicated glyph in THX. For the latter case, both id\_N5191\_lexical and id\_N5191\_glyphic have the values. The id\_N5191\_lexical is the id to obtain the cognate character in the N5191; the id\_N5191 is the id to find the nearest glyph shape.

## ④ map\_type

The map\_type is a classification of the proposed value. There are the following values.

- Unify              No remarkable glyph difference and should be unified
- Unify (dup)        No remarkable glyph difference and should be unified with the duplicated entry
- VS                  Visible glyph difference, but N5189 accepted VS to distinguish them.
- VS (if required)    Subtle glyph difference, but N5189 accepted VS to distinguish them.
- TBD                No consensus at present
- Bug\_N5191          See the appendix of this document

Figure 3 and Figure 4 illustrate how to determine the map\_type.

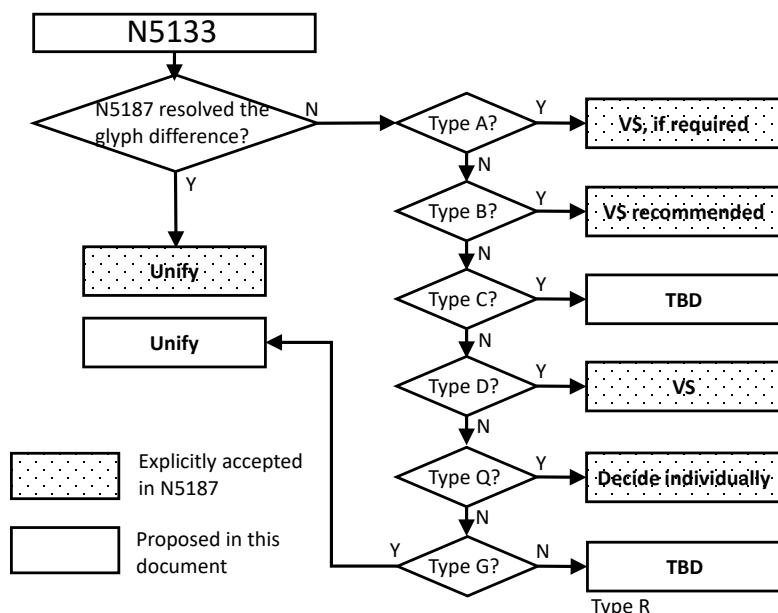


Figure 3: Process to select *map\_type* value for the entries listed in N5133

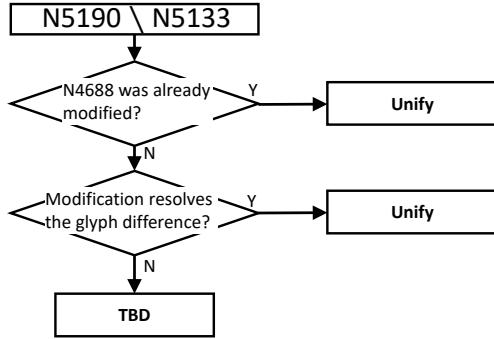


Figure 4: Selection Process of *map\_type* value for the entries listed in N5190 but not in N5133

##### ⑤ corresp\_UCS (informative)

Corresponding CJK Unified Ideographs to help the review and discussion.

##### ⑥ N5133

The value is the type of glyph difference in N5133. The format is N5133:[A-DGQRa-dgqr]. The letter after ":" is the type of the glyph difference defined in N5133. The uppercase, like N5133:A, is the case in which N5191 resolved the glyph difference between CCZ and the proposed character set. The lowercase, like N5133:a, is the case in which N5189 implicitly accepts using VS to distinguish glyphs (if required).

##### ⑦ N5189

The value is the type of reason for the glyph modification or removal, which N5189 described. The format is N5189:[1-6]. The last digit from 1 to 6 is the type of glyph change defined in N5189. The most important case is "N5189:2", which indicates that the lexical position of THX is excluded from the proposal because the entry at the position is exactly duplicated (same glyph and same semantics) with other positions. Another important case is "N5189:5", which indicates the modification of the glyph at the lexical position to avoid the conflict with non-cognate character.

### 3. Draft Mapping Table with Glyph Images

For easier review, the comparison table with the glyph images is available at

<http://gyvern.ipc.hiroshima-u.ac.jp/~mpsuzuki/ShuoWen/review2023/CCZ2THX-202306.pdf>

Figure 5 illustrates the table structure with the glyph images and how it shows the attributes. Also, the reduced table

			Unify (4) N5189:4 N5133:B (7) 噪, 噪 (5)
00949	00949 (1)	00949 (2)	(3)

CCZ	N5191	N5191u	comment
			VS (if required) N5133:a 

N5187 implicitly accepted to handle the glyph difference by VS

			Unify (dup) N5189:2 涙
00914	00914	07729	(4)

The lexical position in N5191 is empty due to exact duplication (N5189:2), so CCZ entry should be mapped to the unified entry of N5191

			TBD (dup) N5189:5 蹠
06138	05927	01491	(5)

The lexical position in N5191 is a tweaked glyph due to possible duplication (N5189:5), so CCZ entry can be mapped to the entry with non-cognate but same-shaped glyph of N5191

Figure 5: Structure of Mapping Table with Glyph Images

## Appendix

The map\_type values for seq\_CCZ 06469 and 06811 are "Bug\_N5191".

The case for 06469 is a wrong duplication of the representative glyph in N5191, as shown in Figure 6. The id\_N5191\_lexical should be 06249.

Serial No.序号	Serial No. in Tenghuaxie version 藤本序號	Unicode	TTF Font	Corresp Modern Char. 今字	SW Radical 說文部首	SW Radical Number
06468	06248	39935	𠂇	丐	丐	327
06469	06249	39936	𠂇	首	首	328
06470	06250	39937	𦵈	稽	首	328

Figure 6: Bug in N5191 for THX06249

The case for 06812 is a wrong duplication of the serial number of THX in N5191, as shown in Figure 7. The id\_N5191\_lexical should be 06566, but N5191 set it as 06565.

06811	06565	39A8C	𢃥	礪	石	357
06812	06565	39A8D	𢃦	礪	石	357
06813	06567	39A8E	𢃧	礪	石	357

Figure 7: Bug in N5191 for THX06566.

(end of document)