

Universal Multiple-Octet Coded Character Set UCS

ISO/IEC JTC 1/SC 2/WG 2 IRG N1610

Date: 2009-10-22

Source:	SUZUKI Toshiya, Faculty of Integrated Arts and Science, Hiroshima University
Title:	Comments on Proposed CJK Unified Ideographs from the source “Index to the Collection of Inscriptions in Yin-Zhou Period”
Actions required:	For consideration at IRG #33 at Da Nang, Vietnam
Distribution:	IRG Members and Ideographic Experts
Medium :	Electronic

Abstract

Checking IRG N1519 part 6 which collects the evidences of the characters proposed to CJK Unified Ideographs Extension C & E (current name), many insufficient evidences are found. The glyphs in IRG N1519 part 6 are collected from “Index to the Collection of Inscriptions in Yin-Zhou period”, so the proposed glyphs are modernized Old Hanzi in Bronze Inscriptions. Some glyphs are not used in the elucidated texts of the book, thus there is a possibility that they were to be removed but left by editorial errors. There are more than 100 such unused glyphs. Also there are the glyphs that the stabilities of their shapes are questionable. The author wants to propose to reschedule the standardization of G_ZJW glyphs until the work of Old Hanzi group on Bronze Inscriptions completes, to standardize the required glyphs with stable shapes.

Comments on proposed CJK Unified Ideographs from the source “Index to the Collection of Inscriptions in Yin-Zhou period”

suzuki toshiya, Faculty of Integrated Arts and Science, Hiroshima University

1. Background

As Japan SC2 committee (JSC2) distributed in the IRG32, I checked IRG N1519 part 6 that collects the evidences of the characters proposed to CJK Unified Ideographs Extension C & E (current name), that are tagged as G_ZJW. All the characters are referring the “Index to the Collection of Inscriptions in Yin-Zhou period” (殷周金文集成引得, abbreviated as I2CIYZ in following), that is a collection of the elucidated texts in the rubbings that were published in the “Collection of Inscriptions in Yin-Zhou period” (殷周金文集成, abbreviated as CIYZ in following). It seems that the proposed characters are invented to typeset the elucidated text digitally, with the glyphs that can remind the shape of Bronze inscriptions. After the part that collects the elucidated text in the rubbings in CIYZ, there is a dictionary-like part that a reader can search the elucidated text from a specified character(s). Although IRG N1519 part 6 does provide the pronunciation and the meaning of the proposed glyphs as conventional Hanzi dictionaries, original I2CIYZ does not provide such information. In I2CIYZ, only a phrase of the elucidated text including specified glyph is shown. By checking the elucidated texts in the dictionary-like part, some such glyphs are found to be questionable on the accuracy of their shape or the requirement of such glyphs.

2. Problems of the I2CIYZ glyphs

There are 4 problems in the glyphs in I2CIYZ.

- A) Some items have too many exemplified glyphs than the elucidated text**
- B) Some exemplified glyphs are incorrect modernization**
- C) Some exemplified glyphs may be reinventions of modernized ShuoWen glyphs**
- D) Some exemplified glyphs are not used in the elucidated text**

About A, B and C problems, the stabilities of the glyph shapes are questionable.

About A, C and D problems, the requirements of the glyphs are questionable.

In the following sections, each of the above problems will be described in detail.

An item of the dictionary-like part of I2CIYZ shows multiple SongTi graphic as the exemplified glyphs. In most case, the exemplified glyphs in a specified item have mutually incompatible shapes and it is difficult to recognize them as forming a glyphic group of ideographic variants, from the viewpoint of modern SongTi typography. This is remarkable contrast from modern Hanzi dictionary whose item is given for a glyphic group formed by the glyphs with ignorable differences.

The reason to list multiple exemplified glyphs can be understood as following. When the Bronze inscription glyphs are classified into the groups sharing same “role” in the text, the glyphic differences in a specified group are too distinct to exemplify by a single SongTi glyph. To avoid misconception assuming too narrow coverage of glyph variants, I2CIYZ lists multiple SongTi glyphs to show appropriate coverage of the possible glyph variants.

If we deal the dictionary-like part of I2CIYZ as the conventional Hanzi dictionary, it is expected that each items have unique idea on shape or unique idea on the per-glyph meaning, or unique role in the elucidated text. But some items in I2CIYZ have inclusive relationship; some glyphs are difficult to determine which item is the best item to be assigned.

Figure 1 shows the typical examples are the item 1737 & 3059. The second glyph of the item 1737 is used in the items 3059 again. From the elucidated character “浣”, apparently there is no meaning difference. Thus the second exemplified glyph should be dropped from the item 1737, to solve the inclusive relationship.

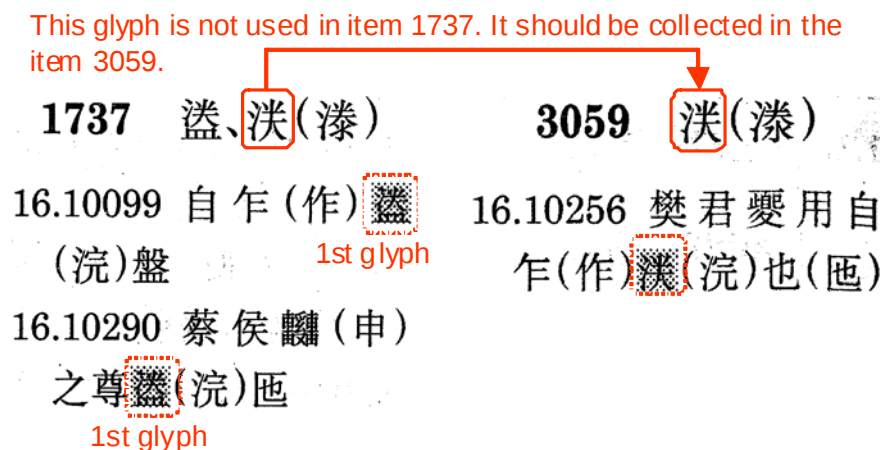


Figure 1: Examples of inclusive relationship between 2 items of I2CIYZ.

- In the case of the conventional Hanzi dictionary whose item shows only one exemplified glyph, the wrong item is easily found by editorial checking. In the case of the dictionary-like part of I2CIYZ, the number of exemplified glyphs in an item is from 1 glyph to 8 glyphs, there is a possibility that some unnecessary exemplified glyphs are overlooked and left in the item. Figure 2 shows the typical example in the item 0831 & 0832. These items are designed to list both of possible mirror reversal variants, but only one side is found in the rubbings in CIYZ. Another side of mirror reversal is imaginary glyph if we restrict the scope to the rubbings in CIYZ. The 2nd glyph of the item 0831 was proposed as G_ZJW00627, the 2nd glyph of the item 0832 was proposed as G_ZJW00430.

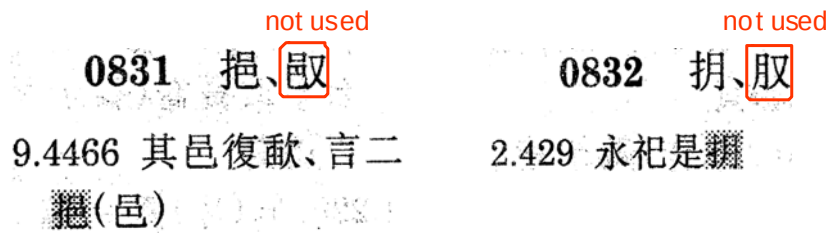


Figure 2: Examples of imaginary glyphs of I2CIYZ (envisioned by “mirror-reverse”)

2.1. Some item have too many exemplified glyphs than the elucidated text

This is problem A, there is a possibility that unnecessary glyphs are proposed, or the glyphs with unstable shapes are proposed.

If an item has multiple exemplified glyphs that are difficult to be unified, the number of exemplified glyphs is expected to be similar to the number of the elucidated text. But there are many items listing too many exemplified glyphs in comparison with the number of the elucidated text. Figure 3 shows the typical examples.

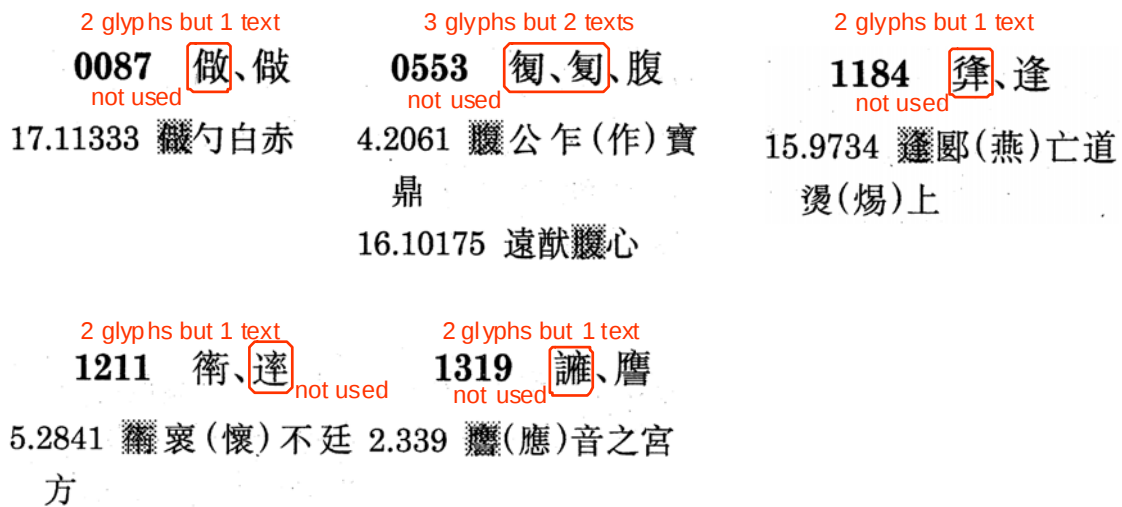


Figure 3: Examples of items listing too many modernized glyphs in I2CIYZ.

2.2. Some exemplified glyphs are incorrect modernization

This is problem B, there is a possibility that the glyphs with unstable or incorrect shapes are proposed.

In some items listing multiple exemplified glyphs, some exemplified glyphs have an extra radical or lack a radical that found in Bronze inscriptions. These glyphs are not similar to the shapes on referred rubbing. If such glyphs are modern Hanzi, they would be listed to notice the meanings of the index, but sometimes invented exemplified glyphs different from the rubbings are listed. Figure 4 shows the typical examples in item 1107.

The item 1107 lists 7 exemplified glyphs for 11 elucidated texts. No pair of 7 exemplified glyphs can be unified by ISO/IEC 10646 Annex S rule. From this item, 5 ideographs had been proposed (1st=G_ZJW01658, 2nd=G_ZJW1668, 3rd=G_ZJW01657, 4th=G_ZJW01667, 6th=G_ZJW01024). Checking the glyph shapes on the rubbing in CIYZ, there are only 3 groups. In fact, only 2 exemplified glyphs (the 1st=G_ZJW01658 and 3rd=G_ZJW01657) are used in the elucidated texts.

The exemplified glyphs including both of “𠂇” (radical person, U+2E85) and “𠂈” (radical walk, U+2ECC) are difficult to recognize as they reflect the shape of referred rubbing. There are 2 routes of modernization of “𠂉” (radical walk, U+2FA1). One is “𠂈” (radical simplified walk, U+2ECC), another is “𠂇” (radical step, U+2F3B), and further simplification to “𠂇” (radical person, U+2E85). The referred rubbings have no glyph include both of them, even if we consider all possible simplification for the modernization, the glyphs including both of “𠂇” and “𠂈” (the third, fourth) are incorrect.

1107 邇、邇、進、
邇、邇、傳、傳 different from the shape of Bronze inscriptions



Figure 4: Examples of incorrect modernizations in I2CIYZ.

2.3. Some exemplified glyphs may be reinventions of modernized ShuoWen glyph

This is problem C, there is a possibility that the duplicated encodings are proposed, the stability of proposed glyph shape and the requirement of the glyph are questionable.

I2CIYZ makes the group of items by the radicals & the number of strokes. The set of radicals follows to ShuoWenJieZie. Thus, most of existing radicals are expected to be coded until Extension B, because IRG N1519 part 6 is not the first proposals to encode the characters modernized from pre-kaishu glyphs. In fact, some radicals of I2CIYZ show multiple glyphs. Figure 5 show the part of the radical #221 from I2CIYZ, the first glyph “𠂔” is already coded as U+29AD6 (proposed by PRC with KangXiZiDian source), the second glyph is now proposed as G_ZJW00686. From the

description, it is supposed that the author was aware of the duplication of U+29AD6 and G_ZJW00686, but he had chosen G_ZJW00686 to modernize other glyphs. There is a possibility that others use U+29AD6 for their modernization, because U+29AD6 is already coded character, and many conventional Bronze inscription dictionaries use U+29AD6 shape.

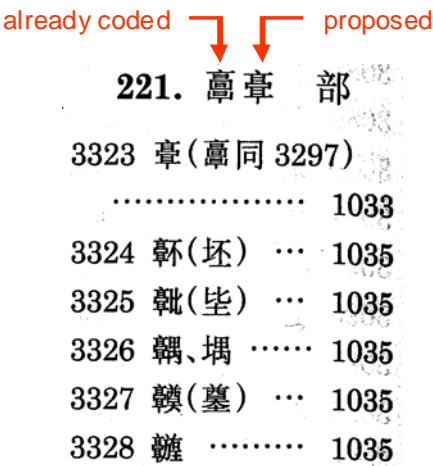


Figure 5: Examples of yet-another modernized glyph taken from SWJZ

2.4. Some exemplified glyphs are not used in the elucidated text

This is problem D, there is a possibility that the unneeded glyphs are proposed.

There are some glyphs that have reasonable shape, but the elucidated texts do not use. Figure 6 shows the typical example in the item 0047. This index lists 2 exemplified glyphs, the first glyph is modern Hanzi commonly used (“使”U+4F7F), and the second glyph is invented glyph and proposed as G_ZJW00333. But if we check the elucidated texts, the invented glyph is never used. As I mentioned the route “𠂔” → “彳” → “亻”, the shape of the invented glyph might be reasonable, but there is a possibility that the it was useless for the author but the author slipped to remove it.

0047 使 使 not used in text

3.513 左隹車尼	5.2809 隹厥友引以告	15.9684 右隹車裔夫郛	16.9925 左隹車工蔡	16.10446 右隹車裔夫	18.12057 左隹車裔夫
3.537 左隹車工尼	于伯懋父	(齊)瘞、工角	16.9926 左隹車工黑	郛(齊)瘞、工疥	孫固、工墳(墳)
3.971 左隹車工墳(墳)	5.2840 隹智(知)社稷	15.9685 左隹車裔夫孫	16.9933 右隹車工疥	16.10447 左隹車裔夫	18.12058 左隹車裔夫
4.2088 左隹車工墳	之質(任)	固、工自(師)質	16.9934 右隹車工疥	孫固、工黑	孫固、工墳(墳)
(墳)	9.4477 左隹車工蔡	15.9686 左隹車裔夫孫	16.10333 右隹車裔夫	16.10450 右隹車工蔡	18.12059 左隹車裔夫
4.2089 左隹車工墳	9.4478 左隹車工墳	固、工墳(墳)	郛(齊)瘞、工	16.10451 右隹車工黑	孫固、工墳(墳)
(墳)	(墳)	15.9692 左隹車裔夫孫	16.10349 左隹車工蔡	16.10472 左隹車造	18.12060 左隹車裔夫
4.2090 左隹車工墳	9.4664 左隹車工尼	固、工上	16.10358 左隹車裔夫	16.10477 右隹車裔夫	孫固、工墳(墳)
(墳)	9.4665 左隹車工黑	15.9693 左隹車裔夫孫	事敦、工質	郛(齊)瘞、工疥	18.12061 左隹車裔夫
4.2091 左隹車工黑	15.9448 右隹車裔夫	固、工墳(墳)	16.10359 右隹車裔夫	18.11472 詔隹	孫固、工墳(墳)
4.2092 左隹車工尼	宋、郛(齊)瘞、工隼	15.9735 隼(舉)擊(賢)	郛(齊)瘞、工虞	18.11814 左隹車工墳	18.12062 左隹車裔夫
4.2093 左隹車工蔡	(觸)	隹能 / 隹得擊(賢)在	16.10397 右隹車工疥	(墳)	孫固、工墳(墳)
4.2094 左隹車工蔡	15.9450 右隹車裔夫郛	(士)良隹(佐)賙(貯)	16.10402 左隹車裔夫	18.11822 左隹車四	18.12063 左隹車裔夫
5.2678 隹于曾	(齊)瘞、工	/ 外之則賙(將)隹能	七欽(欽)、工尼	18.12054 左隹車裔夫	孫固、工墳(墳)
5.2707 右隹車裔夫郛	15.9561 左隹車工尼	(上)勤於天子之席 /	16.10413 左隹車工下	孫固、工墳(墳)	
(齊)瘞、工簡	15.9562 左隹車工黑	隹其老筭(策)賞仲父	士甘	18.12055 左隹車裔夫	
5.2719 叔氏隹質(布)	15.9674 右隹(車)裔夫	16.9924 左隹車工蔡	16.10444 左隹車裔夫	孫固、工墳(墳)	
	吳詒、工賙		孫固、工蔡	18.12056 左隹車裔夫	
	15.9675 左隹車裔夫孫		16.10445 左隹車裔夫	孫固、工墳(墳)	

Figure 6: Example of possibly-correct-but unused glyph in I2CIYZ.

3. Proposal about the rescheduling of the standardization about G_ZJW glyphs

Considering the problems described in previous section, I want to propose to reschedule the standardization of G_ZJW glyphs after the standardization of Bronze Inscriptions by Old Hanzi group.

As I show a few cases in above, some modernized glyphs in I2CIYZ are only listed in the items but not used in the main text. Checking the pages referred by IRG N1519 part 6, I found about 149 glyphs that are listed in the items but not used in the elucidated text. Figure 7 shows the list of such glyphs.

0255-0047.2 z.jw00333	使	0422-0676.2 z.jw00231	𠂔	0483-0000.x z.jw00473	𠂔	0564-1491.1 z.jw01499	𠂔	0688-2094.1 z.jw00782	𠂔	0765-2531.2 z.jw00744	𠂔	0944-2932.1 z.jw00381	𠂔	1074-3670.3 z.jw01516	𠂔	1128-3987.2 z.jw00556	𠂔	1340-4515.1 z.jw00011	𠂔
0262-0087.1 z.jw00706	𠂔	0429-0702.1 z.jw00391	𠂔	0483-0000.x z.jw01207	𠂔	0564-1494.1 z.jw00803	𠂔	0696-2138.1 z.jw00690	𠂔	0785-2560.1 z.jw01425	𠂔	0971-2956.3 z.jw01566	𠂔	1075-3675.1 z.jw00588	𠂔	1145-4098.1 z.jw01186	𠂔	1389-4566.1 z.jw00279	𠂔
0279-0142.1 z.jw00211	𠂔	0445-0831.2 z.jw00627	𠂔	0483-1139.2 z.jw01539	𠂔	0584-1556.2 z.jw00425	𠂔	0700-2180.1 z.jw01310	𠂔	0800-2570.1 z.jw01301	𠂔	1008-0000.x z.jw01226	𠂔	1076-3692.1 z.jw01591	𠂔	1145-4100.1 z.jw00180	𠂔	1389-4566.2 z.jw00971	𠂔
0284-0198.1 z.jw01850	𠂔	0445-0832.2 z.jw00430	𠂔	0491-1184.1 z.jw01493	𠂔	0586-1576.1 z.jw01081	𠂔	0707-2233.1 z.jw01735	𠂔	0855-2635.2 z.jw01218	𠂔	1010-3131.2 z.jw01237	𠂔	1077-3696.2 z.jw00828	𠂔	1147-4118.2 z.jw01195	𠂔	1389-4566.3 z.jw00170	𠂔
0336-0462.2 z.jw00710	𠂔	0452-0873.2 z.jw01501	𠂔	0497-1223.2 z.jw00206	𠂔	0586-1577.1 z.jw01356	𠂔	0711-2269.1 z.jw01743	𠂔	0855-2637.2 z.jw01217	𠂔	1010-3133.1 z.jw01236	𠂔	1079-3723.2 z.jw00895	𠂔	1148-4133.1 z.jw01393	𠂔	1468-4665.2 z.jw00429	𠂔
0343-0468.2 z.jw00122	𠂔	0452-0874.2 z.jw00995	𠂔	0497-1225.1 z.jw00837	𠂔	0590-1604.1 z.jw00790	𠂔	0724-2336.1 z.jw01595	𠂔	0860-2647.3 z.jw00253	𠂔	1021-3199.2 z.jw00502	𠂔	1088-3752.1 z.jw01884	𠂔	1148-4137.1 z.jw01395	𠂔		
0347-0511.2 z.jw01833	𠂔	0460-0947.1 z.jw00448	𠂔	0499-1233.1 z.jw00006	𠂔	0594-1645.1 z.jw01843	𠂔	0732-2345.2 z.jw01616	𠂔	0863-2657.1 z.jw00372	𠂔	1023-3221.1 z.jw01824	𠂔	1088-3760.1 z.jw01894	𠂔	1173-4161.2 z.jw01403	𠂔		
0347-0511.3 z.jw01827	𠂔	0460-0947.3 z.jw00443	𠂔	0513-1260.2 z.jw00604	𠂔	0599-1674.1 z.jw01867	𠂔	0740-0000.x z.jw01640	𠂔	0864-2666.5 z.jw01055	𠂔	1034-3309.2 z.jw01386	𠂔	1089-3769.1 z.jw01886	𠂔	1179-4198.1 z.jw00663	𠂔		
0347-0511.4 z.jw01547	𠂔	0463-0984.2 z.jw01206	𠂔	0518-1319.1 z.jw01570	𠂔	0612-1705.2 z.jw01321	𠂔	0750-2434.1 z.jw01485	𠂔	0865-2667.1 z.jw01045	𠂔	1038-3343.5 z.jw01789	𠂔	1090-3774.1 z.jw01872	𠂔	1184-4229.3 z.jw00993	𠂔		
0359-0536.1 z.jw00688	𠂔	0466-0990.2 z.jw00378	𠂔	0523-1355.1 z.jw01030	𠂔	0619-1752.4 z.jw00681	𠂔	0756-2461.1 z.jw00670	𠂔	0865-2676.2 z.jw00383	𠂔	1041-3360.1 z.jw00656	𠂔	1091-3786.2 z.jw01510	𠂔	1246-4254.2 z.jw00674	𠂔		
0418-0624.1 z.jw00138	𠂔	0474-1072.6 z.jw00990	𠂔	0524-1377.1 z.jw00351	𠂔	0620-1755.1 z.jw01338	𠂔	0758-2470.1 z.jw00064	𠂔	0874-2730.2 z.jw00315	𠂔	1043-3374.1 z.jw01772	𠂔	1092-3791.1 z.jw01913	𠂔	1259-4269.2 z.jw00839	𠂔		
0419-0630.1 z.jw00148	𠂔	0474-1072.8 z.jw00900	𠂔	0527-1413.1 z.jw01036	𠂔	0645-1918.1 z.jw00260	𠂔	0758-2473.1 z.jw00677	𠂔	0888-2770.1 z.jw00100	𠂔	1043-3383.1 z.jw01778	𠂔	1093-3820.1 z.jw01535	𠂔	1289-4361.1 z.jw00029	𠂔		
0419-0636.2 z.jw00151	𠂔	0477-1089.2 z.jw00338	𠂔	0533-1431.2 z.jw00731	𠂔	0646-1926.1 z.jw00901	𠂔	0759-2479.1 z.jw00680	𠂔	0910-2905.1 z.jw00174	𠂔	1059-3530.1 z.jw00973	𠂔	1093-3821.1 z.jw01537	𠂔	1296-4432.2 z.jw00051	𠂔		
0419-0637.2 z.jw01860	𠂔	0478-1107.2 z.jw01668	𠂔	0546-1448.1 z.jw00441	𠂔	0679-1979.2 z.jw00945	𠂔	0759-2486.1 z.jw01814	𠂔	0910-2908.2 z.jw00817	𠂔	1067-3605.1 z.jw00525	𠂔	1096-3839.1 z.jw01798	𠂔	1300-4452.3 z.jw00784	𠂔		
0420-0641.1 z.jw00787	𠂔	0478-1107.4 z.jw01667	𠂔	0546-1451.2 z.jw00157	𠂔	0680-2016.2 z.jw01380	𠂔	0760-2496.1 z.jw01907	𠂔	0940-2924.2 z.jw00942	𠂔	1068-3621.2 z.jw01519	𠂔	1114-3951.1 z.jw01584	𠂔	1309-4486.1 z.jw01478	𠂔		
0422-0669.2 z.jw00786	𠂔	0478-1107.6 z.jw01024	𠂔	0555-1472.2 z.jw00167	𠂔	0686-2035.2 z.jw00303	𠂔	0764-2514.1 z.jw01428	𠂔	0941-2926.1 z.jw01806	𠂔	1074-3669.2 z.jw01524	𠂔	1116-3976.2 z.jw01506	𠂔	1309-4493.2 z.jw01391	𠂔		

Figure 7: Exemplified glyphs that are never used in the elucidated text.

The requirement and the stability of such glyphs are questionable; there is a possibility that some unnecessary glyphs are overlooked by editorial error. The discussion about such modernized glyphs requires the wide knowledge about Bronze inscriptions and the conventional dictionaries about Bronze inscriptions, so it is difficult to work as a part of modern ideographs. Considering the difficulties, I want to propose the standardization of G_ZJW glyphs after the standardization of Bronze inscription in Old Hanzi project. In the current working of Old Hanzi experts group for Oracle Bone script, the modern glyphs are added to all Oracle Bone glyphs. If Bronze inscriptions will be managed by same database, their glyph determinations would be good basis to discuss the requirement of new modernized glyph, the stable shape of the modernized glyph, and the unification rule of invented radicals.