

The templates of the typical comments to Oracle Bone glyph database

suzuki toshiya (Hiroshima University, Japan)

This document provides the comment templates for non-Chinese speakers to submit the comments to Oracle Bone database. The process to make Oracle Bone glyph database is summarized in the section 1, and the possible comments are described in the section 2 (based on the texts in “Notes” column in Old Hanzi database dated 2011-02-14), and a list to translate the typical comments from English to Chinese is given in the section 3.

1. The assumed process of Glyph DB production for these templates

The comment to Glyph DB should clarify what action should be taken (the update or cleansing of the source picture, the exclusion from the final output, the discussion of unification, the discussion of sorting etc). One action could make following action. For example, the update of source picture may insert new actions; the consideration that the glyph from updated source could be unified or not. The comment submitter should clarify if the following action is needed or not. Thus, the outline of the glyph DB production process should be described. This template assumes the process outlined in Figure 1.

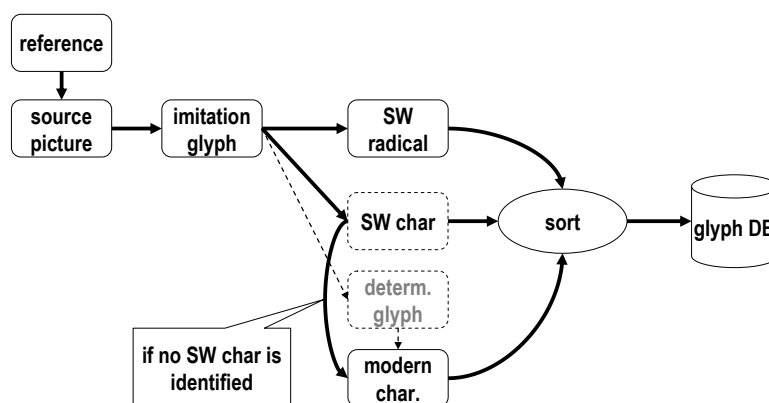


Figure 1: The assumed process of Oracle Bone glyph DB production

In the original Old Hanzi P&R (until version 2), DB had a column to hold “determ. glyph”, the shape translation of Oracle Bone glyph by Kai glyphic components. But it was decided to remove it. Thus the process related with “determ. glyph” is shown by dashed line. There are many comments requesting the correction of determ. glyph in existing notes of the glyph DB, but should not be added from now.

In addition, it is agreed to sort the glyphs in DB by Shuowen ordering, but the DB has

no column to hold corresponding SW glyph, thus it is shown in by dashed line, but the comment about the relationship between the Oracle Bone glyph and SW glyph would be helpful.

From the existing comments in the “Notes” column, the typical comments can be outlined as Figure 2

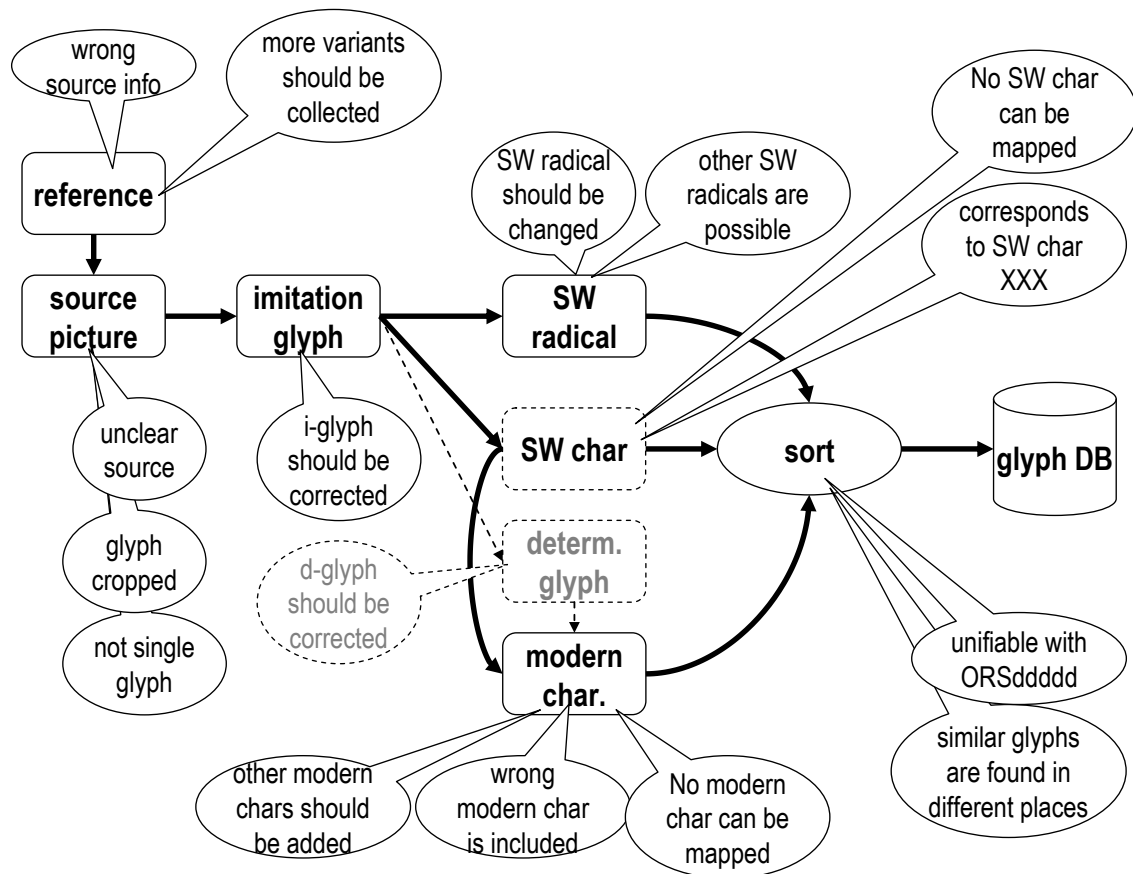


Figure 2: The typical comments to Oracle Bone glyph DB

2. The templates of the comments

In this section, the short English comment (in the subsection title) and the purpose to use it are summarized.

2.1. Comments for the references

This subsection lists the comments to the referential materials or the possible data that are not submitted yet.

2.1.1. “referential info is wrong”

The source picture cannot be identified from the source ID column, and the source picture is suspected to be taken from different sources.

2.1.2. “more variants should be collected”

The Oracle Bone glyph DB collects the glyph from the referential materials regardless of the coverage of existing Oracle Bone dictionaries, thus there is a possibility that some variants in existing Oracle Bone could be overlooked. To request more variants to be discussed, this comment should be given.

2.2. Comments for the source picture

This section lists the comments to the submitted glyphs. Sometimes the re-submission with better source is requested.

2.2.1. “source is unclear”

Some source pictures are difficult to identify the glyph shape correctly, because of the damages in the original objects, poorly inked/rubbed picture, poor digitization, or scaling at invalid aspect ratio. To propose the appropriate action, the comment submitter should clarify the problem is in the original object (原片不清), or in the referential printed matter (圖片不清), or in the digitization process.

2.2.2. “glyph is cropped”

Sometimes the glyph is extracted from the edge of the object and its shape is incomplete. If the cropped part is so small that the compensation by guessing is recognized to be possible, such source is accepted. If the comment submitter suspects the compensation is difficult and better source should be checked (or propose to exclusion from the final output as “insufficient evidence”), this comment should be used. The expected action, rechecking for better source or exclusion from the final output should be clarified.

2.2.3. “not single glyph”

Unlikely with modern CJK ideographs, Oracle Bone glyphs are not regularized to the square bounding. Some compositions of glyphic components are not easy to decide if it is a single glyph or a composition of multiple glyphs. If the comment submitter suspects the source picture includes multiple glyphs, this comment should be used.

2.2.4. “insufficient context”

To identify an Oracle Bone glyph, sometimes the neighboring glyphs are required. If there is no sufficient information to identify the glyph and it should be excluded from the final output, this comment should be used.

2.3. Comment for imitation glyph

The imitation glyph should show the legible glyph shape by black ink on white background. If the imitation glyph is recognized to be different from the source picture, it should be commented.

2.3.1. “imitation glyph should be drawn”

In Old Hanzi P&R, the imitation glyph is specified as “truthful trace” of the glyph in source picture. However, some DB entries give only black-white negated picture and

difficult to recognize the glyph shape without background noise. For such entries, this comment should be used.

2.3.2. “imitation glyph should be corrected”

When the DB entry submitter works with the trimmed glyph picture instead of original whole picture, there is a possibility that some noises (like cracks or bumps of the objects) are misunderstood as a part of glyph. If the comment submitter suspects the glyph includes such noise, this comment should be used.

2.3.3. “imitation glyph from different source”

Sometimes the DB entry is submitted with wrong pair of the imitation glyph and source picture. The imitation glyph is quite difficult to be related with the source picture, even if the differences listed in above are considered. The comment submitter should use this comment in such case. The requested action (replace the correct source picture, or exclusion from the final output) should be clarified.

2.4. Comment for corresponding SW char

As defined in Old Hanzi P&R, there are different sort policies in the glyph DB, for the glyphs mappable to SW, unmappable to SW but mappable to modern char, unmappable to both. Therefore the information if the glyph is unmappable to SW glyph (“未識字”) and/or unmappable to modern character (“未釋字”) is important. The information that the glyph cannot be mapped to modern character should be described in the section for the comment to corresponding modern character.

2.4.1. “unmappable to SW glyph”

If the comment submitter thinks the glyph cannot be mapped to SW glyph, this comment should be used.

2.5. Comment for SW radical

Even for the glyph that cannot be mapped with SW glyph, the SW radical should be determined.

2.5.1. “SW radical should be changed to ...”

If the existing SW radical in of the entry is suspected to be wrong, the correct or possible SW radicals should be given. The comment submitter is expected to give SW radical number.

2.5.2. “other SW radicals are possible”

Some Oracle Bone glyphic components are not distinguished in the original era, but distinguished in SW era. Thus, a group of similar glyphs could be divided to different radicals. If the comment submitter afraid that similar glyph can be found in different SW glyph, this comment should be used.

2.6. Comment for determined glyph

The DB entry submitters used various methods to fill the determined glyph column. Until IRG#36 meeting, there were attempts to clean this column, so the glyph DB has many comments to request the correction of the glyph shape in this column. However, this column would not be included in the final output and would be removed in future. Thus no comments should be given.

2.7. Comment for modern corresponding character

The corresponding modern characters are filled by various methods. It might be a character with corresponding meaning, a character with similar shape, a character supposed to be derived from the glyph, a character corresponding to SW glyph. Therefore the multiple modern characters could be included. After IRG#36, P&R version 3 restricts the coverage of the possible character is URO (CJK unified ideographs in BMP, without Extension A). The insertion and exclusion of corresponding modern character should be commented.

2.7.1. “Modern characters should include ...”

If the comment submitter could not find possible modern character in the entry, the submitter should use this comment. The submitter should clarify the reason how proposed character can be related with the glyph; corresponding shape, corresponding meaning, determined by SW, determined by something else.

2.7.2. “Modern characters should not include ...”

If the comment submitter finds wrong modern characters in the list, the submitter should use this comment. The submitter should clarify the reason why it is wrong. The typical reasons would be “the character out of URO”, “no references using this modern character”. In some entry, the glyph structure description by unclear syntax, like, “示(豆/卅)”, is used. The request to move them to the “Notes” column can be submitted by this comment. This comment is not against to leave the modern character column as empty, even if the comment proposes to remove the last character in this column.

2.7.3. “No modern characters can be mapped”

If the comment submitter proposes to leave the corresponding modern character column as empty, this comment should be used. If a glyph cannot be mapped to SW glyph, and cannot be mapped to modern character by its shape, the mapping by meaning is expected. But sometimes the assorted meaning of the glyph is unavailable, or, the glyph is recognized as proper name. In such case, the leaving the modern character column as empty could be better, and this comment should be used.

2.8. Comment for sort

The sorting information is not explicitly stored in the database at present. The comment to gather similar glyphs should be submitted.

2.8.1. “unifiable with ORSdddd”

If the submitter finds an entry in DB that could be unified with other entries, this comment should be used. The submitter is expected to give the ORS number or the modern glyph of the glyph to be unified with.

2.8.2. “similar glyphs might be found in...”

If the submitter has a concern that the similar glyphs can be found in other places of the database, this comment should be used. The location of the similar glyphs is expected to be given by the ORS number or the modern glyph of the glyph to be unified with.

2.9. Comment for clarification

The “Note” column is an accumulation of the discussion in the past. Some discussion caused the action items, but sometimes its result is unclear (the action is taken or not). To request the status clarification, following comments should be used.

2.9.1. “Why deleted?”

In most cases, the deletion of the entry is because of unclear source, cropped glyph or duplicated entry. Some cases, there are deleted entries without the reason. If the comment submitter thinks the glyph in the deleted status should be included, the submitter should request the reason of the deletion.

2.9.2. “Source picture is already replaced?”

This comment can be used when the “Note” column instructs to replace the source picture but it is unclear if the instruction is done.

2.9.3. “Source information is already corrected?”

This comment can be used when the “Note” column instructs to correct the source information but it is unclear if the instruction is done.

2.9.4. “Imitation glyph is already corrected?”

This comment can be used when the “Note” column instructs to correct the imitation glyph but it is unclear if the instruction is done.

2.9.5. “Commented modern characters should be added?”

This comment can be used when the “Note” column gives some modern characters determined by preceding studies but it is not found in the modern character column.

2.9.6. “Commented places are already checked?”

This comment can be used when the “Note” column suggests checking another place or SW radical for similar glyph but it is unclear if the glyph ordering was changed.

2.9.7. “Conclusion of the discussion?”

This comment can be used when the “Note” column notes just as “discussion is needed”, “need discussion” like that, and it is unclear if anything is changed.

2.9.8. “Unified with?”

This comment can be used when the status is “Unified” but without the information of the glyph to be unified.

3. The summarized list of the typical comments to Oracle Bone glyph database

Process	comment	
	English	Chinese
Reference	referential info is wrong	典拠不正?
	more variants should be collected	
Source	source object is unclear	原片不清
	source picture is unclear	圖片不清
	digitization is distorting	
	glyph is cropped	字形不全
	not single glyph	
	insufficient context	
Imitation	imitation glyph should be drawn	
	imitation glyph should be corrected	摹修
	imitation glyph from different source	
SW glyph	unmappable to SW glyph	未識字
SW radical	SW radical should be changed to ...	移入「...」部
	other SW radicals are possible	
Determ.	Determined glyphs should be corrected	Determ. 修改
Modern char	Modern characters should include ...	
	Modern characters should not include ...	
	No modern characters can be mapped	未釋字
Sort	unifiable with ORSdddd	
	similar glyphs might be found in...	互見...
Clarification	Why deleted?	
	Source picture is already replaced?	
	Source information is already corrected?	
	Imitation glyph is already corrected?	
	Commented modern characters should be added?	
	Commented places are already checked?	
	Conclusion of the discussion?	
	Unified with?	