

Title: Additional Comments on Shuishu in PDAM2.2 text
Source: suzuki toshiya
Document Type: individual contribution
Date: 2018/04/20

Abstract

In addition to WG2 N4942, some additional points are found that some discussions are expected.

1. Radical-based ordering needs the improvement.

According to the discussion records in N4839 [4], it seems that the characters out of some special categories (anthropomorphic figures, mammals, birds, etc) are ordered by the mixture of the radical and first stroke. Consideration of the first stroke causes some difficulties to spot the radical to be searched, from given glyph shape, like, similar shape but different radical.

Radical-01

1B300	一	SHUIISHU LOGOGRAM QYED7-A
	→ 1B500 一	shuishu radical-01
	→ 4E00 一	
1B301	𠂇	SHUIISHU LOGOGRAM BYEENG3
	→ 4E19 丙	
1B302	𠂈	SHUIISHU LOGOGRAM HNGAAN5 HNGOU1
1B303	𠂉	SHUIISHU LOGOGRAM DA16
	→ 5927 大	
1B304	𠂊	SHUIISHU LOGOGRAM BI2 KU1
1B305	𠂋	SHUIISHU LOGOGRAM QYUM5-A
1B306	𠂌	SHUIISHU LOGOGRAM DI6-A
1B307	𠂍	SHUIISHU LOGOGRAM DYUM1 FONG2 HU3

Radical-08

1B357	𠂎	SHUIISHU LOGOGRAM LYU2 LI2 SUI3
	→ 1B508 𠂎	shuishu radical-08
1B358	𠂏	SHUIISHU LOGOGRAM XONG1 XEEB7
1B359	𠂐	SHUIISHU LOGOGRAM SA1 BONG2
1B35A	𠂑	SHUIISHU LOGOGRAM DI6 HAANG1 FAN1
1B35B	𠂒	SHUIISHU LOGOGRAM XA3
1B35C	𠂓	SHUIISHU LOGOGRAM VYEN3
1B35D	𠂔	SHUIISHU LOGOGRAM BYED8 XONG1
1B35E	𠂕	SHUIISHU LOGOGRAM DEE5

The characters including radical-08 as a component but placed under radical 01

why U+1B305 and U+1B307 are placed under radical-01 (horizontal stroke), U+1B35A is placed under radical-08 (zig-zag stroke)? I guess, the ad-hoc members decided to place there, because the first stroke of U+1B304 is top horizontal bar, but the first stroke of U+1B35A is zig-zag on the left.

Radical-40

1B48C	𠂖	SHUIISHU LOGOGRAM SEE1
	→ 1B52B 𠂖	shuishu radical-40
1B48D	𠂗	SHUIISHU LOGOGRAM FAAM4
1B48E	𠂘	SHUIISHU LOGOGRAM SAO5
1B48F	𠂙	SHUIISHU LOGOGRAM BAI1 QNUG7
1B490	𠂚	SHUIISHU LOGOGRAM DEI2
1B491	𠂛	SHUIISHU LOGOGRAM MEI4 NUG8 SAO5
1B492	𠂜	SHUIISHU LOGOGRAM GGAN5
1B493	𠂝	SHUIISHU LOGOGRAM HUI6
1B494	𠂞	SHUIISHU LOGOGRAM YONG2

Radical-01

1B310	𠂟	SHUIISHU LOGOGRAM BA2
1B311	𠂠	SHUIISHU LOGOGRAM NDEN1
1B312	𠂡	SHUIISHU LOGOGRAM LYEEU2
1B313	𠂢	SHUIISHU LOGOGRAM FAI4 NU4
1B314	𠂣	SHUIISHU LOGOGRAM ZING3
1B315	𠂤	SHUIISHU LOGOGRAM XGAAN2 FU5
Radical-31		
1B447	𠂥	SHUIISHU LOGOGRAM DYUNG3
	→ 1B521 𠂥	shuishu radical-31
1B44E	𠂦	SHUIISHU LOGOGRAM NU2

The characters including radical-40 as a component but placed under radical 01 or 31

why U+1B310 and U+1B315 are placed under radical-01, U+1B44E is placed under radical-31, regardless with the exist of radical-40-like component? I guess, the first strokes of U+1B310, 12, 15 are the top horizontal bar and it is

not the part of radical-40-like structure.

Radical-29

1B431	𠂇	SHUIISHU LOGOGRAM ZUI2 → 1B51F 𠂇 shuishu radical-29
1B432	𠂈	SHUIISHU LOGOGRAM GEN3 SUI3 XGEN4 • Water Earthworm of Chariot (constellation)
1B433	𠂉	SHUIISHU LOGOGRAM NDONG5
1B434	𠂊	SHUIISHU LOGOGRAM SEI5
1B435	𠂋	SHUIISHU LOGOGRAM JA6
1B436	𠂌	SHUIISHU LOGOGRAM JI2
1B437	𠂍	SHUIISHU LOGOGRAM ZI3 LYEEU3

Radical-02

1B318	丄	SHUIISHU LOGOGRAM GGAM4 MOG8 → 1B501 丨 shuishu radical-02
1B319	丅	SHUIISHU LOGOGRAM LUNG5
1B31A	𠂎	SHUIISHU LOGOGRAM NNI6
1B31B	𠂏	SHUIISHU LOGOGRAM GGUI5-A
1B31C	𠂐	SHUIISHU LOGOGRAM XAM3

The characters including radical-29 as a component but placed under radical 02

why U+1B31B is placed under radical-02? I guess, the first stroke of U+1B31B is the vertical bar, although the first strokes of U+1B431~U+1B437 is the circle (and followed by connected stroke).

By examples in above, it would be known that the radical selected by the first stroke is not always the most significant glyphic component. There is a concern whether the consideration of the first stroke makes the classification unstable. According to the note in N4638 [1] “according to Shuishu masters, it causes this situation that Shuishu masters deliberately modified the shape of the characters to keep other Shuishu masters from reading them for privacy during the period that each family owned its version. Thus it increases the mystique and leads to many variants for the lack of conformity”, it would be unsafe to assume as the first stroke is stable among the variants, even if the glyphs are similar.

Taking U+1B3C8 (meaning summer, “夏”) as an example, it is placed under the “T”-like-radical (radical-19). Checking the glyph shapes in 水書常用字典 [9], there are many variants and most of them have “丅”-like component, not “T”-like component. Considering that the glyph looking like “丄” (U+1B318) or “丅” (U+1B319) are placed under radical-02 (vertical bar), the “支”-like glyphs in 水書常用字典 would be placed under radical-02 (vertical bar) or radical-21 (try square). In fact, I wish if all of them are placed under radical-10 (又).

1B3C8 𠂑 SHUIISHU LOGOGRAM YA3
→ 590F 夏

𠂑 (𠂑、𠂒、𠂓、𠂔、𠂕、
𠂖、𠂗、𠂘、𠂙、
ja³³
夏
夏季, 《水书·十平
撒象》: ~三月申子辰 ja³³
ha:m¹³ ɲot³² tsi⁵⁵ sən¹³ ci³³
sən³¹ (~季三月忌申日、
子日和辰日)。

“summer” in PDAM2.2

“summer” in 水書常用字典

[9]

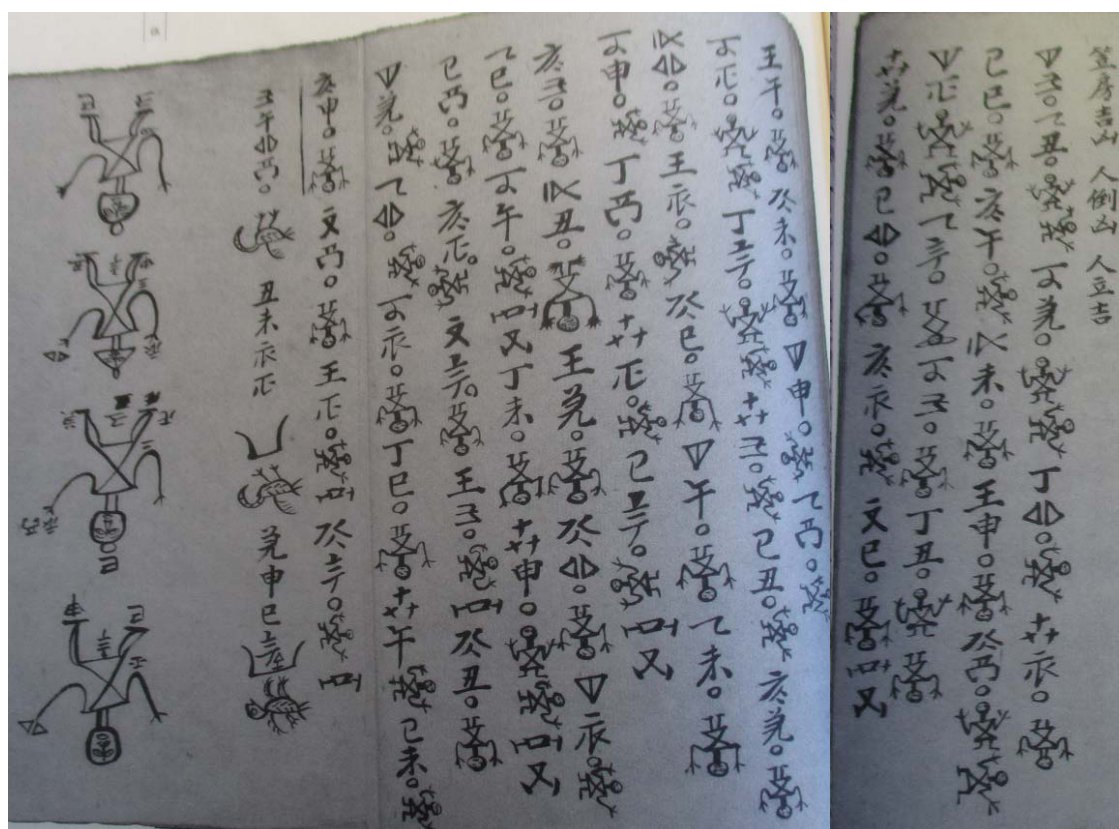
Also, it is expected that the assigned radical is stable against the rotation or mirroring. The numeral 2 and 3 in

PDAM2.2 consist from the vertical strokes (see p. 4 in WG2 N4942), but the numerals 2 and 3 in the past submissions and past researches consist from the horizontal strokes. If such kind of rotation is frequent in Shuishu document, the radical changed by the rotation would be misleading for the users.

I suggest to reconsider the ordering under the sub-collections, as the earliest proposal N4638 [1]. The information related to the calendar, constellations, numerals are quite frequent in Shuishu documents, so it would not be so difficult to identify these context. Although it is not practical in CJK Unified Ideographs to define such sub-collections, because of too frequent usages in non-calendar, non-constellation contexts. But it seems that the dominant usage of calendar, constellation, zodiac characters in Shuishu are still in the context of calendar, constellation and zodiac systems.


1.1. Handling of the rotated glyphs with modified semantics


Also, sometimes the single document gives different semantics for the rotated glyphs.




rotation variant of U+1B4B9 found in 中國水書 [8] (三都卷, 第 96 冊 (S776))

Checking some Shuishu documents, it seems that the rotation of the anthropologic glyphs are popular method to modify the semantics. For example, document S776 in 中國水書, there is a note “人倒凶 人立吉” (upset human indicates bad fortune, upright human indicates good fortune) and following texts use 4 variants: upright, clockwise 90 degree, clockwise 180 degree, clockwise 270 degree rotated human. At present, only upright human is proposed

(U+1B4B9 ). How to encode the texts including rotated human, without loss of the distinction of the contrast of

good or bad? Is it the layout issue, not the character encoding issue? There are 2 characters for good fortune ( 1B3D3

𐍎⁵⁵, U+1B3D3) and bad fortune ( 𐍎¹³, U+1B36E), but it is unclear how we can map 4 variants into 2 characters.

I'm not saying as 4 variants should be coded separately, but it should be clarified how such documents should be coded, because the upset human is quite frequent in Shuishu document.

2. Why composite glyph without component glyph?

Some Shuishu characters in PDAM2.2 [6] are supposed to be composite characters to mean rather complicated idioms, like, U+1B4C4, U+1B4C8 and U+1B4D4. They are described as “hungry like a horse, like a pig¹, like a chicken”

1B4C4		SHUISHU LOGOGRAM JI1 NGA6 MA4
		• hungry like a horse
1B4C8		SHUISHU LOGOGRAM JI1 NGA6 HMU5
		• hungry like a pig
1B4D4		SHUISHU LOGOGRAM JI1 NGA6 GGA15
		• hungry like a chicken

Composite characters of horse, pig and chicken in PDAM2.2 [6]

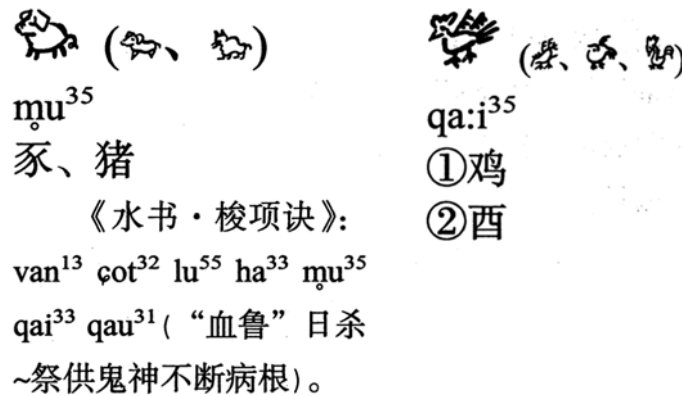
It seems that these glyphs are composite glyphs of an animal and something. It would be natural for the users to expect the characters for the component animals (in this case, horse, pig and chicken). For a horse, there are general-purpose horse and zodiac symbol horse. But for chicken and pig, there are only zodiac symbols.

1B4C5		SHUISHU LOGOGRAM SYAANG1 NNED8 MA4
		• Horse (constellation)
1B4CA		SHUISHU LOGOGRAM MA4
		• horse
1B4CF		SHUISHU LOGOGRAM BYED7 FA3 ZU1
		• Pig (constellation)
1B4D8		SHUISHU LOGOGRAM NGA2 NNID8 JI1
		• Rooster (constellation)

Atomic characters for horse, pig and chicken in PDAM2.2 [6]

According to the structure of proposed phonetic values, clearly there are words for chicken (GGA15) and pig (HMU5) in Shui language. Should we understand as “Shui language have the words for chicken and pig, but Shuishu does not have?”. No, because 水書常用字典 has general-purpose pig and chicken.

¹ In Han language, pig (豚) and boar (猪) are often recognized as non-cognate, but it seems that they are not seriously distinguished in Shui language.



Atomic characters for pig and chicken in 水書常用字典 [9]

Thus, I think, I’m not asking something like “there is an emoji for airplane, why there is no emoji for propeller? there is an emoji for car, why there is no emoji for gastunk?”. According to N4638, 61359 characters from 5 books were already checked on 2014; there was no pig, no chicken? If it is true, how many times “hungry like a pig”, “hungry like a chicken” were used? This situation reminds a concern from USNB in PDAM2.2 ballot; *This large pictographic set should have an established record of publication and stronger indication that the repertoire is complete and well-analyzed before it is standardized.* The term “pictographic” would be rather important, because there would be a concern whether these composite characters should be coded at single codepoint, or coded by some sequence like complex emoji.

3. All representative glyphs are sufficiently distinctive?

According to the section “**Reasons for deletions**” of N4758, it seems that the submitters avoided to encode 2 different glyphs with same semantics, even if their glyph shapes are significantly different. For example, “具”-like character meaning “death” is kept in PDAM2.2 Shuishu chart, but “乚”-like character meaning “death” is withdrawn between N4696 and N4758. This is consistent with the process described in N4758.

62	132			DE11	tai ¹	死	S8	Reordered and corrected meaning
----	-----	--	--	------	------------------	---	----	---------------------------------

N4758, the character meaning “death” which is kept in PDAM2.2 Shuishu

2	59			DE11	tai ¹	死	S1	variants
---	----	--	--	------	------------------	---	----	----------

N4758 [3], the character meaning “death” which is removed before PDAM2.2 Shuishu.

But, contrasting cases are found too.

255	86			LYOG8-A	ljok ^s	禄	S8	
256	242			LYOG8-B	ljok ^s	六	S5	

N4758 [3], 2 characters with same pronunciation which are kept in PDAM2.2 Shuishu

18	243	𠄎	LYOG8XXEN2	ljok ⁸ xən ²	禄存星	s1 ^o	Variants, the same as six
----	-----	---	------------	------------------------------------	-----	-----------------	---------------------------

N4758 [3], the character with similar pronunciation which is removed in PDAM2.2 Shuishu

1B449	𠄎	SHUISHU LOGOGRAM LYOG8-B
	→ 516D 六	
1B41D	𠄎	SHUISHU LOGOGRAM LYOG8-A

two characters in PDAM2.2 [6] Shuishu with same pronunciation.

Considering the semantics are swapped in PDAM2.2 [6] (in N4638 [1] and N4758 [3], U+1B41D-like shape was described as numeral 6, but in PDAM2.2, U+1B41D is looking like as if it were *not* numeral 6), it causes a concern that these 2 characters are semantically interchangeable but coded separately.





Some experts would have the objection; *it should be discussed in future, and no need to postpone the standardization of the initial set, because we can change the policy from semantic identity to glyphic identity in future*. If the glyph shapes in PDAM2.2 are all significantly different and never been interchangeable, it would be right attitude. But, PDAM2.2 Shuishu have some glyphs with similar shapes, and there is a concern that the representative glyph shapes are not designed to be sufficiently distinguishable.

For example, U+1B48C and U+1B48D have quite similar glyph shape, and it is unclear for the implementers how they should be designed distinctively. The flat or slant horizontal stroke? Checking other documents, there is a different glyph shape for U+1B48C (see it is unclear whether “fa:n⁵⁵” is corresponding to U+1B48D). If such pairs must be coded separately in the initial set, it is recommended to improve the glyph shapes more distinguishable.

Radical-40

1B48C	𠄎	SHUISHU LOGOGRAM SEE1
	→ 1B52B 𠄎	shuishu radical-40
1B48D	𠄎	SHUISHU LOGOGRAM FAAM4

Example of two characters in PDAM2.2 Shuishu with similar shape.

			
se ¹³	sui ³³	fa:m ⁵²	fa:n ⁵⁵
梳子	篋子	犯	线

水书中的时间名称及
由此化用而成的水书条
目名称，有“中伤、陷害”

The glyphs similar to U+1B48C, U+1B48D in 水書常用字典[9]

Some experts might insist as; *both of U+1B4BC, U+1B4BD are most frequently used shapes, so they should be authorized* (as described in the section “**variants**” in N4758 [3]). But the variant table in N4758 lacks which glyph from which context. There is a possibility there are so many documents using U+1B4BC only, or U+1B4BD only, and the documents using both characters with distinctively designed shapes are hidden by them.

For example,

4. Shuishu looking like Hanzi should be coded separately or unified?

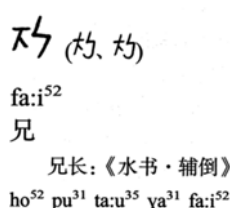
It is known that some Shuishu characters are quite similar to Hanzi and they are supposed to be imported characters from Hanzi to Shuishu. Therefore, PDAM2.2 Shuishu provides some relationship between Shuishu and Hanzi. Some of them are quite similar.

Radical-20		
1B3DC	𠂇	SHUISHU LOGOGRAM SUB8 → 1B514 𠂇 shuishu radical-20 → 5341 𠂇
1B3DD	𠂈	SHUISHU LOGOGRAM XED7-A → 4E03 𠂈
1B3DE	𠂉	SHUISHU LOGOGRAM SUB8 QYEI
1B3DF	𠂊	SHUISHU LOGOGRAM GAM3
1B3E0	𠂋	SHUISHU LOGOGRAM GUB7
1B3E1	𠂌	SHUISHU LOGOGRAM JU3 → 4E5D 𠂌

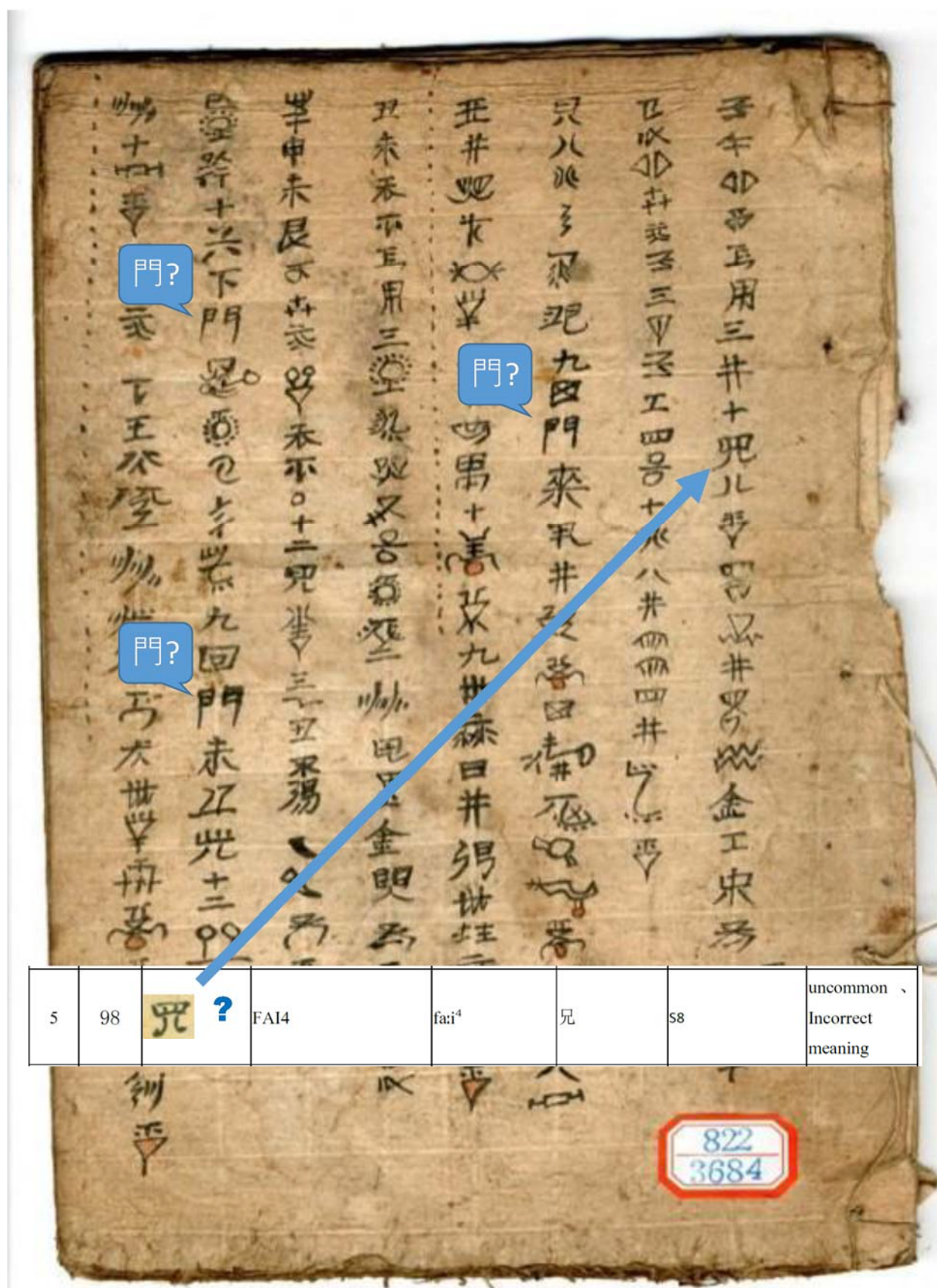
10 and 7 in PDAM2.2 Shuishu [6]


Checking the photos of the source materials used in the proposal to PDAM2.2 [6], there are several Hanzi-like characters which had never found in N4758 [3], although N4758 has a list of updated list and the differential list from the past submission.

For this reference, another interesting observation is that the removed character meaning elder brother (兄, fa:i⁵²), or something looking like that, is found in this document, and current PDAM2.2 has no character to mean elder brother. This observation causes a concern about the completeness of the proposed set. In fact, we can find the character meaning elder brother in 水書常用字典.



Shuishu character meaning “elder brother” in 水書常用字典 [9].



5	98	 ?	FAI4	fai ⁴	兄	S8	uncommon 、 Incorrect meaning
---	----	---	------	------------------	---	----	------------------------------------

A Reference used for Shuishu proposal (S6 Tanju: Nine Stars) and non-proposed characters

5. Summary

In summary, this document described 5 points to be resolved for the standardization;

- A) the assignment of the radical should be reconsidered, to be more stable among the variants.
 - the radical assigned by the first stroked component is not generic among the variants.
 - it is discussed how to handle the rotated glyphs with modified semantics.
- B) why composite glyphs are included but their components are not included?
 - there are past researches showing some of components characters, are they really missing in the source document of the current proposal?
- C) all representative glyphs are sufficiently distinctive?
 - the methodology by variant table in N4758 could be insufficient, because it does not consider the difference between “documents using one of 2 similar characters” and “documents using both of 2 similar characters”.
- D) the characters looking like Hanzi should be coded separately?
 - if the discussion was already done, please describe the criteria what kind of Hanzi-like Shuishu was proposed and what kind were excluded (and if there is any plan to submit more in future).

The proposals to resolve these issues are following:

- A) provide the variant table and discuss about the most stable component which is most generally found in the variants. if it is too long term work, or, the database is already lost, more algorithmic assignment is expected;
 - define the stable sub-collections for numerals, zodiac symbols etc as the earliest submission N4638 [1], and the characters for such sub collections should be divided to such sub-blocks and their own ordering, instead of radical plus first stroke ordering. also see proposal D).
 - make the list of components for each proposed characters, aslike IDS for CJK Unified Ideograph. the structural description is not needed, just the components are needed.
 - pick the most complicated (e.g. the component with the maximum number of strokes, or, the component with the maximum number of corners) as the radical to order the character.
- B) confirm whether the component characters are really missing in the source documents.
 - if they are really missing, the postpone of the composite characters for the idiom is suggested.
- C) improvement of the variant table, with the consideration of the document using one or all of similar glyphs.
 - if it is too long term, the removal of similar glyphs from the first set is suggested, regardless with their semantics.
- D) it would not be practical to exclude all Hanzi-derived or Hanzi-like characters. thus, please define the several important categories which are very frequently used (like nine stars, lunar mansions, numerals, heavenly stems, earthly branches, time, wuxing/ba-gua/yin-yang) and allow Hanzi-like characters in them, because the number of Hanzi-derived characters in these categories would not have an explosion. for general purpose characters, the discussion about the criteria should be discussed as future issue.

References

- [1] WG2 N4638, “Proposal for encoding Shuishu in the SMP of the UCS”, China NB (2014-08-20)
- [2] WG2 N4696, “Preliminary Proposal for encoding Shuishu in the SMP of the UCS”, China NB (2015-10-19)
- [3] WG2 N4758, “Updated Proposal for encoding Shuishu in the SMP of the UCS”, China NB (2016-09-03)
- [4] WG2 N4839, “Towards the ordering of the Shuishu script in the UCS”, Michael Everson (2017-07-26)
- [5] WG2 N4894, “Results of the ad-hoc meeting on Shuishu in Hohhot, 2017-09-21”, (2017-09-22)
- [6] WG2 N4922, “Additional repertoire for ISO/IEC 10646:2017 (5th ed.) Amendment 2”, (2017-11-22)
- [7] WG2 N4938, “Draft Dispositions of comments on PDAM2 to ISO/IEC 10646 5th edition”, (2018-03-13)
- [8] 《中國水書》編審指導委員會：『中國水書』，四川出版集團 (2006.12-)
- [9] 韋世方：『水書常用字典』，貴州民族出版社, (2007-09), ISBN 9787541214783
- [10] 西田龍夫：『言語学大辞典 別冊 世界文字字典』，水書, p.523-526 (2001), ISBN 4385151776)