SAT Comments to IRGN2225

1. On general principles

According to the document, the component list is given with the intention "to increase quality of machine checking of IDS", along with that "improve the quality of IDS database" (IRGN2220) and "enhance the ability to use intuitive Ideographic Description Sequences" (IRGN2204). However, we find the current status of the list not optimally designed in light of its ideal and motivation. As we observed, the issues in general fall under several types:

(Actual review for individual components follows in the next section.)

a) Etymological separation of similar shapes

Although it is crucial for encoding ideographs to refer to their historical background, emphasizing etymological correctness in IDS decomposition may sometimes complicate comparison and matching workflow, as graphically intuitive division of a modern ideograph is not necessarily backed by its etymology. For example, 香 can be easily described like ⿳ 禾日, where the bottom part does not derive from 日 nor 曰 (but 甘). Therefore we should be restrictive from loading etymological semantics on components without radical use.

b) Unification criteria

We find that the criteria which divide some components in the main list (§2) and those in reference components list to be unified (§3) is not enough clear. Some §3 components do have minimal pairs in current standard while some of §2 do not. Any justification or instruction will be appreciated. Supposedly unifiable components are mentioned in the next section.

c) Insufficient isolation of compound shapes

Some shapes are seemingly too big to be monolithic components. While they may have been understandably induced from the set in BMP, further decomposition would contribute to flexibility, future utility, and fuller coverage, with which the original objective of describing commonest characters remains unaffected.

2. Individual issues

| H9-01 | 臣亡 | CJK COMPONENT SIMPLIFIED OVERSIGHT
Example: top of 監 / 覧 / 豎
= CDP-8CC9 |
| H14-01 | 臣盇 | CJK COMPONENT OVERSIGHT
Example: top of 鑒 /寛
= CDP-8D50 |

These shapes are easily decomposable (1.c). Singling out the right half of H9-01 will allow more usability and coverage. H14-01 seems redundant.

S4-01	刂又	CJK COMPONENT C-SIMPLIFIED LEAD Example: top of 坚 / 竖 / 紧 / 肾 / 贤 = CDPO-81EF / UTC-02964

This shape is easily decomposable (1.c). The left part can be described using S2-01, which is also a simplified shape that etymological distinction are not pointful (1.a).

S5-01	刂ケ	CJK COMPONENT C-SIMPLIFIED OVERSIGHT Example: top of 监 / 览 = CDPO-81F0 / UTC-02965

This shape is easily decomposable (1.c). It shares left part with S4-01 (S2-01) and right part with H9-01.

P4-02	夕	CJK RADICAL MEAT TWO Example: top of 炙 / 畓 / 然 / 將
P4-03	夕	CJK RADICAL MOON TWO Example: top right of 望 = CDP-8B5E

The distinction between the two shapes may be blurred when they do not serve for meaning (1.a). They are however still useful considering the universal recognition of MEAT-MOON distinction among ideographic regions.

P5-01	𠚫刀	CJK COMPONENT LIU Example: top of 留 / top left of 劉 = CDP-8C69

This shape is decomposable (1.c). Singling out the left half will allow more usability and coverage.

P7-01	𣆙	CJK COMPONENT FOOD VESSEL TWO Example: left of 即 / 既

Z5-01	𠮷	CJK COMPONENT FOOD VESSEL ONE Example: left of 即 / 既 = CDP-8B7C

Needs more clarification about distinction and/or use cases regarding the two shapes (1.b).

P15-01	𧮫 𧮫 𧮫	CJK COMPONENT COMPLEX POURING VESSEL Example: top of 釁 / 曡

This shape is decomposable (1.c). Making a component without 同 in the middle (=CDP-8B64) will enable to include other shapes in place of it.

| D6-01 | 亦 | CJK COMPONENT C-SIMPLIFIED CHAOS
Example: top of 恋 / 娈 / 峦.
Note: use 4EA6 亦 instead where appropriate, such as 5955 奕
= UTC-02967 |

Distinction with 亦 is not meaningful being a simplified shape (1.a, 1.b).

| D7-02 | 宀夂 | CJK COMPONENT GUEST
Example: top of 賓
= CDP-8D74 |

This shape is easily decomposable (1.c). Singling out the bottom half will allow more usability and coverage.

| Z3-02 | 丮 | CJK COMPONENT HOLD ONE
Example: right top of 恐 / 鞏 etc |
| Z3-03[3] | 丸 | CJK COMPONENT HOLD TWO (variant of 丮)
Example: right of 執 / 埶 / 狃 / 秇
Note: use 4E38 丸 (wan2) where appropriate, e.g. 7D08 紈. |

The distinction between these shapes and existing characters such as 丮, 丸, 凡, 凡 etc. may be blurred when they do not serve for meaning (1.a). Addition of these components would not likely to provide much more usefulness.

| Z4-01 | 矢 | CJK COMPONENT TILTED HEAD
Example: bottom of 吳
= CDP-8CD5 |

This glyph is not significantly different from 矢, especially compared with unification examples, thus should be subject to unification (1.b).

3. Proposal for addition

Shape	Description	Examples
士	A shape mainly derives from 宋 (U+233B5). = CDP-8D52	索, 孛, 橐 etc.
立	A shape mainly derives from the upper part of 辛 (U+41C2). = CDP-8BAE	帝, 商, 啇, 产 etc.

4. On notation

The suggested format of provisional components in IDS is such as ⬚H7-01 日. We however

propose more formally (and visually) significant notation that using HTML entity reference format, such as ⊟&H7-01;⊟. It also agrees with the convention in some of Taichi Kawabata's documents.

The advantages are:

- more visually segmented (e.g. ⊟&H7-01;&H3-04;)
- easier to define in program
- easier to detect mistypes with program

The disadvantages are:

- type more letters
- possibly more mistypes