

ISO/IEC JTC1/SC2/WG2/IRG N2637 Kushim Feedback

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по Стандартизации

Doc Type: Working Group Document
Title: Feedback on hybrid characters
Source: Kushim JIANG (姜兆勤)
Status: Individual Contribution
Action: For consideration by Gen KOJITANI and IRG
Date: 2023-10-05

The issues related to the hybrid characters contains following materials:

和 hybrid character 有关的议题包含如下材料:

- Online Review Tool, WS2021-00021 𠄎 <hc.jsecs.org/irg/ws2021/app/?id=00021>.
- Online Review Tool, WS2021-00029 𠄎 <hc.jsecs.org/irg/ws2021/app/?id=00029>.
 - Based on the evidence, the left component should be analyzed as a Latin capital letter X and is therefore included in the list.
依据字证，左侧构件应分析为大写拉丁字母 X，因此纳入本列表。
- Online Review Tool, WS2021-00718 𠄎 <hc.jsecs.org/irg/ws2021/app/?id=00718>.
 - Sergeevich characterized the form as a combination of multiple Siddham grapheme clusters, and is therefore included in the list.
Sergeevich 在审查意见中将该形刻画为多个悉昙形位簇，因此纳入本列表。
- Online Review Tool, WS2021-00760 𠄎 <hc.jsecs.org/irg/ws2021/app/?id=00760>.
 - Note that Eiso has overridden the relevant comments in IRG N2413 R2 by using contrary conclusion in its review comment (#11329) through a more careful analysis.
请注意 Eiso 在审查意见中通过更细致的证据分析，使用相反的结论覆盖了 IRG N2413 R2 中与该字有关的意见。
- Online Review Tool, WS2021-00765 𠄎 <hc.jsecs.org/irg/ws2021/app/?id=00765>.
- Online Review Tool, WS2021-00770 𠄎 <hc.jsecs.org/irg/ws2021/app/?id=00770>.
- WANG Yifan (2019). *On Encoding Policy of Gongche Notations and Upcoming Para-ideographs (tentative)*. L2/19-346.
- Eiso CHAN (2019). *Request to clarify some IDS issues*. IRG N2413 R2.
- Gen KOJITANI (2023). *Script-hybrid and abbreviated CJK letters*. IRG N2637.
- Gen KOJITANI (2023). *Proposal to Encode Abbreviated Ideographs (Kanjis) in Japanese*. L2/23-139.
- Ken LUNDE. *UniHan Database Property "kStrange"*. UTN #43.
 - Informative. Please focus on Category H, Category K and Category O.
资料性。请着重浏览 Category H、Category K 与 Category O。
- Andrew WEST. *BabelStone Fonts: BabelStone Han PUA*. <babelstone.co.uk/Fonts/PUA.html>
 - Informative. Please focus on the end of the repertoire.
资料性。请着重浏览字汇集尾部分。

There are four levels of issues here.

这里有四个层面的问题。

The first level is whether the evidence meets the *IRG PnP* requirements. That is, whether the character is in Han script, whether the script is in Kaishu, and whether it has textuality.


第一层为字符是否符合 *IRG PnP* 要求。即文种是否为广义汉字、书体是否为楷书、是否具有文本性。


Textuality is easily recognized from the evidence, but the first two require additional consideration.

文本性从字证中容易获证，但前两者需要另外考察。

Han script does not require the language recorded, but requires the stroke system, the component system, and the cognition system be broadly consistent with Han ideographs. Therefore, if these hybrid characters belong to a writing system that uses Han ideographs, they can be generalized as belonging to Han script. So, the more important question is to determine whether these hybrid characters are or must be part of these writing systems.

广义现代汉文字对图符系统所记录的语言不作要求，但要求笔画系统、构件系统和理据系统与汉字大体一致。因此如果这些 hybrid character 处于使用广义汉字的书写系统中（方块壮文、方块布依文、方块苗文……），可以笼统地将其视作属于广义汉字。因此更重要的问题在于判断这些 hybrid character 是否属于或必须属于这些书写系统的一部分。

For example, should the pictographs in the text of Zhuang ideographs (mention in Liu, 2022; see **Figure 1**) be regarded as part of Han script? If not, the pictograph  mentioned in L2/19–346 (page 9) should not be regarded as part of Han script either.

例如，刘建琼《云南方块壮字的构造》中提及的方块壮文中的“象形字”（参见 **Figure 1**）是否应当视作广义现代汉文字的一部分？若否，则 L2/19–346 所提象形字“”等也不应视作广义现代汉文字。










文字的遗存。但在云南宁蒗县壮族《麽经》文本中却看到了不少的象形字，如：“”读作 fu⁴²，意为“火”；“”读作 huŋ⁴⁵，意为“猪肚子”；“”读作 n:uŋ²⁴、“”读作 ja:ŋ²⁴，都意为“粮食”；“”读作 ma⁵⁵，意为“马”；“”，读作 tha:u³¹，意为“桃子”；“”，读作 ŋu:i²⁴，意为“山”。从这些实例可以看出，宁蒗壮族方块壮

Figure 1 Pictographs in Zhuang text

Liu J. (2022) “The structure of Zhuang characters of Yunnan”, in *Journal of Honghe University*, 20(1).

Kaishu is one of the scripts that describes (narrow) Han ideographs. Thus, for one, non-Kaishu components will not be allowed. For example,  in IRG N2637 (page 4, no. 24) directly uses non-Kaishu component and may not be allowed (Kaishu context cannot decide that the character belongs to Kaishu). For two, non-Han components will not be able to define Kaishu.

楷书是描述狭义现代汉文字的书体之一。从而，其一，非楷书的构件将不被允许。如 IRG N2637 中“”直接使用非楷书构件，可能不被允许（楷书上下文无法决定该字同属于楷书）。其二，非汉字的构件将无法定义楷书。

The second level is whether the evidence meets the *IRG PnP* requirements. That is, whether the evidence is reliable (confirmed by IRG), whether there are multiple pieces of evidence, whether function and cognition are clarified, whether context is included, and whether it is for justifiable public interest. These requirements are very clear, so there is only discussion in the resolutions and no additional clarification is needed.

第二层为字证是否符合 *IRG PnP* 要求。即字证是否可靠（由会议确认）、是否为多字证、是否包含功能与字理、是否包含上下文、是否出于正当公共利益。这些要求均非常明晰，因此只有决议上的讨论，没有需要额外说明的地方。

The third level is whether the abstract shape can be analyzed. If it can be determined that the character does use the grapheme cluster as a component, then the abstract shape can be analyzed, otherwise the abstract shape is unknown.

第三层为抽象构形是否可被分析。若能确定该字确实使用该形位簇作为构件，则可分析出抽象构形，否则抽象构形不明。

For example, L2/19-346 points out that 𠃉 (now encoded as U+2CF36 𠃉) uses a component like digital three (or 了 shape) to represent the image of a person, and more examples are given in Hu (2006). The abstract shape of the actual shape “3” or “了” in these characters must therefore not be analyzed as the digital three or [了] class, but rather as a hieroglyphic component.

如 L2/19-346 指出“𠃉”（已编码为“𠃉”）使用的阿拉伯数字 3 的构件（或“了”形构件）实际为人象形，胡惠《方块壮字字体类型研究》中给出更多例字。这些字中的 3 形（或“了”形）的抽象构形因此不得分析为阿拉伯数字 3（或“了”类），而是某象形构件。

𠃉	aemq	背	象人形背后背有东西,表示背的动作。	指示象事字
𠃊	naengh	坐	“3”表人形,“一”表示人坐于其上	指示象事字
𠃋	ndwn	站立	“、”于上,表人直立	指示象事字
𠃌	ngaem	低(頭);頰(首)	“、”于人形下表低头状	指示象事字

Figure 2 Pictographs in Zhuang text

Hu H. (2006) *A study on the type of Zhuang ideographs*, Master thesis, Guangxi University.

The last layer is whether the value for attributes can be defined, mainly in the RS and IDS information. Non-Kaishu components and non-Han components are unable to define radical and stroke information, while IDS may have to be extended, where each cell is a legal grapheme cluster.

最后一层为属性值是否可被定义，主要在于 RS 信息与 IDS 信息。非楷书构件与非广义汉字构件无法定义部首信息与笔顺信息；同时 IDS 可能须扩展，其中每个单元是一合法形位簇。

(End of Document)