Title: Feedback on IRGN2731 and IRGN2742 (Revised)
Author: FAN Ming
Status: Individual Contribution

China NB submitted a response to the IRGN2717 on issues related to the encoding of "script-hybrid Han ideographs", in which they stated that these ideographs are not ready to be encoded as CJK-UIs, and listed some reasons in support of their argument. Here I provided some feedback and contrary opinions on these arguments, and explain why it's perfect fine to encode them as CJK-UIs.

中国代表提交了一份对 IRGN2717 中关于"混合文种汉字"的回应意见，认为这种字不应该作为中日韩统一表意文字编码，列出了一些理由作为论据。但我对此的意见则刚好相反，所以我在这里针对这份文件提出反馈，并指出为什么完全可以将它们像普通汉字那样编码。

1. Technical issues

In IRGN2731, China NB argues that "these characters exceed the general understanding and definition of Chinese characters, and go beyond the scope of technical processing of Chinese characters". However, ideographs do not need to be in the "general understanding and definition of Chinese characters" to be encoded as CJK-UIs. Most of these so-called "script-hybrid Han ideographs" are created with the rationality, and their creation process is not fundamentally different from that of Han ideographs. Just as Chinese articles mixed with English abbreviations will still be considered Chinese, it's perfect fine to treat these "script-hybrid Han ideographs" as normal Han characters practically, though this may not be in line with the view of perhaps most Chinese(or East Asians?) nowadays. Also, it's not necessary to localize these non-Han components to convert them to "normal" Han characters, just like it's not necessary to convert every English abbreviations in Chinese articles to Chinese -- again, that's just preconception of conservative views. When we understood and accepted these ideas, the technical problems will not be technical problems any more -- we will have sufficiently motivation to solve related technical problems, instead of blocking the encoding of them as CJK-UIs on the grounds of "technical problems".

在 IRGN2731 中，中国代表认为"这些字符超出了对汉字的一般理解和定义，以及相应的技术处理范畴"。但是，表意文字不需要符合"对汉字的一般理解和定义"才能像普通汉字那样编码。大多数所谓的"混合文种汉字"其创造具备合理性，创造过程与"正常"汉字没有本质区别。正如混入英文缩写的中文文章仍将被视为中文一样，在实际使用当中将这些"混合文种汉字"视为普通汉字是完全可以的，尽管这可能不符合当今大多数中国人（或者东亚人？）的观点。此外，没有必要将这些非汉字组件本地化成汉字部件来将其转换为"正常"汉字，就像没有必要将中文文章中的每个英文缩写都转换为中文一样——我还是说这只是保守的先入为主的文化观念。当我们理解并接受了这些观念之后，技术问题就不会是技术问题——我们会有足够的动力去解决相关的技术问题，而不是以"技术问题"为由阻止它们像普通汉字那样编码。

2. Why encoding them to separate blocks is not necessary

As explained above, it's completely fine to treat these "script-hybrid Han ideographs" as normal

CJK-UIs, so it's not necessary to encode them to separate blocks. Moreover, most of these characters are scattered throughout the text of "normal" Chinese characters in small numbers, and do not systemically form new or independent scripts. Therefore, it would be pretty weird to extract them separately to new blocks.

正如之前所说的，将这些"混合文种汉字"看作一般的汉字没有任何问题，所以没必要单独将它们编码在新区块中。更重要的是，大部分这种字都是零零散散地夹杂在"正常"汉字文本中，且数目很少，并没有系统性地形成新的独立的文种。因此如果将它们单独摘出来编码在新区块是很别扭的。

3. Possible concerns

In IRGN2731, China NB raised concerns that "The Chinese expert group believes that there is a general lack of understanding … provide methods agreed upon by experts to determine its standardized forms, radicals, strokes, variant forms, and unification rules". However, most of this kind of ideographs have clear shapes and sources, and is easy to be reviewed, that is, they are not involve in any such kind of controversies, so it's not a risky behavior to accept and encode them. We can postpone and discuss controversial ideographs in great detail in case of any risky move, but this does not affect the encoding of most others among them. As of the issues related to radical/strokes, again, once we accept them as normal Han ideographs, these issues will not be issues any more.

在 IRGN2731 中，中国代表提出了担忧，称"文字学和编码专家对汉字–表音文字混合体文字尚普遍缺乏认知……能够给出专家一致同意的确定其规范字形、部首、笔画、异体字及统合方式等的方法后进行"。然而大部分这种字的来源和字形明确且易于审查，并不涉及任何此种争议，将它们接受编码并不是冒险的举措。对于部分有争议的字可以推迟、详细讨论，避免冒险的举措，但这并不影响它们中的大部分的正常编码。至于部首笔画之类的问题，一旦我们接受这种文字像一般汉字一样编码之后，这些问题就不再是问题了。

~~(End of Document)~~(Document continue on next page)

Additional Feedback to IRGN2742

Wang Xieyang submitted a new feedback regarding the encoding of hybrid-script Han ideographs, in which he raised four key issues. All these 4 issues have clear and simple answers, which will be explained here in details.

对于混合文种汉字的编码问题，王谢杨提交了一份新的反馈，当中列出了四个关键的问题。这四个问题均有明确清晰的答案，此处将对此展开详细的讨论。

1. Localization of non-Han components

In IRGN2742, Wang Xieyang wrote:

> *For example, can the following three sets of character shapes be unified to each other? ⬜X 也 and ⬜⼂也 ⬜门T and ⬜门丁 ⬜木⬜⬜ and ⬜木才 If they can be unified, then what is the standard of the unification? Is the standard similarity of glyphs?*

The answer is clear, that is, the usage of glyph should simply follow the evidence – use what their sources use. Take ⬜门 T versus ⬜门丁 as an example, if the evidence shows the glyph with localized component, that is, ⬜门丁, then we should simply use it instead of ⬜门 T. If both glyphs were found in different sources, choose the one with more authoritative and notable sources. If both were from authoritative and notable sources, then whoever comes first will be encoded. If they were submitted simultaneously, then let the submitter decide which to use – in fact, this standard is no different from deciding which variant of ordinary Han ideographs to encode, which is already well-established in the IRG procedures. This also answered the issue 3 raised by Wang Xieyang (In fact, I also wrote "*Also, it's not necessary to localize these non-Han components to convert them to "normal" Han characters, just like it's not necessary to convert every English abbreviations in Chinese articles to Chinese*" in this document above)

在 IRGN2742 中，王谢杨写道：

> *比如，以下三组字形是否可以统合？ ⬜X 也和⬜⼂也 ⬜门T 和⬜门丁 ⬜木⬜⬜ and ⬜木才 如果应该统合，那么统合的标准是什么？字形相近吗？*

答案很明确，那就是字形的使用应该按照证据来——来源用啥就是啥。以⬜门 T 和⬜门丁为例子，如果来源使用的是已经经过本地化的部件（也就是说是⬜门丁），那么我们就不用⬜门 T。如果两个字形在不同的来源都有出现，那么谁的来源更具有权威性和知名度就选谁。如果都来自比较权威有知名度的来源，那么哪个先提交的就选哪个。如果是同时提交的，那么就让提交者决定编码哪一个——事实上这和一般汉字决定编码哪个异体的过程没有不同，而这些在 IRG 的工作流程中已经非常明确了。这里同时也回答了王谢杨提出的第三个问题（其实我在这个文件的上面也已经说了"此外，没有必要将这些非汉字组件本地化成汉字部件来将其转换为"正常"汉字，就像没有必要将中文文章中的每个英文缩写都转换为中文一样"）。

2. The criteria for determining hybrid-script Han-ideograph

As the name suggests, the "hybrid-script Han-ideograph" refers to Han ideographs created from

different scripts. The kana and Bopomofo characters list in IRGN2742 did not meet the requirement, so they are definitely out of the scope of hybrid-script Han ideograph discussed here. In fact, the principle of their creation and the usage of them is totally different from that of Han ideographs, so they are different scripts, and can also be excluded based on this principle. Currently, the ideographs that meet the two requirements all contain Han components. However, even one day in the future we found very special examples, this does not affect the encoding of most of them since they themselves did not involve in such controversies.

This also answered the first question raised in IRG2742. That is, once these basic issues were resolved and the principles were established, most of them can be encoded accordingly without any controversies, and it doesn't matter whether their numbers are large or not, as they can be clearly reviewed by principles and sources just like ordinary Han ideographs.

正如名字所示，"混合文种汉字"指的是由不同文种所创造的汉字。文件 IRGN2742 中列举的假名和注音字符组成的符号不符合此定义，所以它们当然不属于这里讨论的混合文种汉字。事实上，它们的创造原理和用途也和汉字完全不同，所以是不同的文种，从这一点上也能够排除。目前我们所见的符合这两个条件的字都包含汉字部件。然而，即使我们在将来遇到了很特殊的例子，这也不影响它们中的大部分的编码，因为这些本身不涉及此类争议。

这也同时回答了 IRGN2742 举出的第一个问题，那就是说，一旦这些基本的问题解决了，原则确定下来了，相应地它们中的大部分就能够没有争议地编码了，数量的多寡其实并不重要，因为它们就能够像普通汉字那样根据原则和来源进行审查。

(End of Document)