ISO/IEC JTC 1/SC 2/WG2 N4796

Title: **Proposal to add new block "CJK Complex Ideographic Symbols"**

Source: **suzuki toshiya**

Type: **individual contribution**

Date: **2017/04/01**

Abstract

Hanzi and its derivertive ideographs in East Asia are natively open character set, and new ideographs are invented even just for a novelty. Some of them have unclear range of the glyphic variations (because of the difficulty of the description, memorization and reproduction of inexperienced complex glyphs), the existing unification rule for CJK Unified Ideograph could be inappropriate. To resolve this difficulty, new block "CJK Complex Ideographic Symbols" is proposed to encode the symbols without too fine-tuned unification rules for CJK Unified Ideograph.

1. Limited Usecase But Non-Unifiable Variants

Today, when Asian governments include unpopular CJK Ideographs into their law systems, they are discussed very carefully, to minimize the confusion by their additions. However, the private synthesis of CJK Ideographs is not, on the contrary, some synthesis were suspected to be confusing by design. The "biang" could be one of them [1]; it is said that there is a song to memorize its glyph shape, but most components have no essential role in the semantics. In fact, some glyphic variations have different number of the components, even in UTC proposal submission[2] (see Figure 1). Also GlyphWiki website have more variants with incompatible component differences (see Figure 2).

UTC-01312 evidence UTC-00791 evidence

Figure 1: Variants found in UTC Proposal [2]

var-006	Var-009	var-010	var-014

Figure 2: Variants found in GlyphWiki (kamiyo_chars-biang-var-006, 009, 010, 014)

The different number of the components is usually recognized as “too significant difference to unify” in the past IRG discussions, so, the unification of these variants are not easily justified in the context of CJK Unified Ideograph. It is not good idea to encode these variants separately, because there would be many users using an instance without covering all variants (see Figure 3). The Similar issues might be found in the characters synthesized for the game looking for the character with the biggest stroke count.

Figure 3: A Trouble by Separate Encoding of the Variants for Complex Proper Name Hanzi

2. Outline of Proposed Solution

In the above, the conflict between conventional identification of CJK Unified Ideograph and some synthesized CJK Ideographs are described. To encode such characters without breaking the consistency of the unification rules for the general purpose CJK Unified Ideographs, new block “CJK Complex Ideographic Symbols” with more greedy unification rule is proposed (see Figure 4).

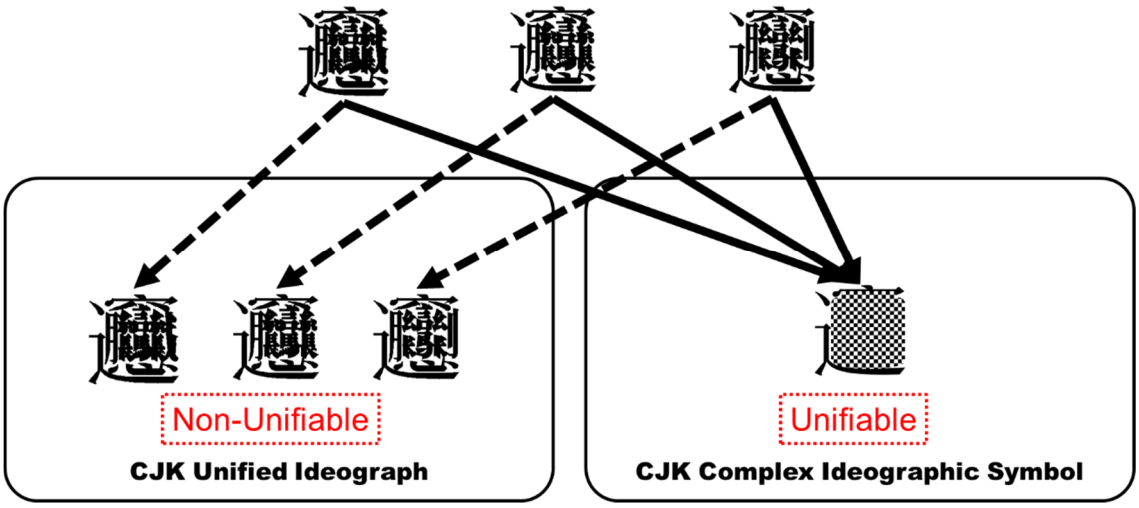

Figure 4: Difference of CJK Unified Ideograph and CJK Complex Ideographic Symbol

The glyph which is hard to define the range of the interchangeable variants, or, the glyphs whose range is difficult to fit with existing CJK Unified Ideographs should be coded in this block. The evaluation of the difficulty should be discussed by WG2 experts, but the rough outline could be designed by following conditions;

A) The character is not used by multiple idioms (not multiple trademarks).

B) Some components of the character could not be justified as essential components by the semantics.

C) Some components of the character have non-unifiable coded variants and the semantics cannot help to choose any specific shape (like 長 versus 长), or, non-unifiable structural difference (like 荆 versus 荊).

D) The number of repeated components could not be deduced by the semantics.

If two or more conditions could be applied to a glyph proposed to CJK Unified Ideographs, the encoding by proposed new block could be a considerable option.

Also, due to the permissive unification rule in new block, some properties for CJK Unified Ideographs should be tailored.

- **Instead of the stroke count, the minimum stroke count should be defined.**
- **Instead of the single radical, the list of possible radicals should be defined.**
- **The representative glyph is not required to be composed by CJK Stokes, to permit the unclear glyph descriptions, like,  .**

[1] https://en.wikipedia.org/wiki/Biangbiang_noodles

[2] IRG N2091, Submission for IRG Working Set 2015
http://appsrv.cse.cuhk.edu.hk/%7Eirg/irg/irg45/IRGN2091_UTC.pdf

[3] http://glyphwiki.org/wiki/Special:Search?search=biang&buttons=%E6%A4%9C%E7%B4%A2

(end of document)