

**Title:** Proposal to enhance the provisional *kBigFive* property

**Author:** Ken Lunde

**Date:** 2022-11-30

Traditional Chinese experts are well aware that while Big Five with its two levels and CNS 11643 Planes 1 and 2 are considered equivalent, there are some key differences. This document is a proposal to flag one key difference, specifically a mapping difference that affects two ideograph pairs, by enhancing the syntax of the provisional *kBigFive* property.

## Big Five versus CNS 11643 Planes 1 & 2

The Unicode coverage of the provisional *kBigFive* property (13,062 ideographs) is identical to that of the HB0 (nine ideographs), HB1 (5,401 ideographs), and HB2 (7,652 ideographs) source prefixes of the normative *kIRG\_HSource* property, and if the source prefixes of the latter are ignored, their property values, which correspond to Big Five code points, are identical. This is because Big Five serves as the foundation for HKSCS (*Hong Kong Supplementary Character Set*). This document will therefore reference only the provisional *kBigFive* property from here when referencing Big Five.

When the Unicode coverage of the provisional *kBigFive* property is compared with that of the T1 (5,413 ideographs) and T2 (7,650 ideographs) source prefixes of the normative *kIRG\_TSource* property, which correspond to CNS 11643 Planes 1 and 2, respectively, the following are the three key differences:

- CNS 11643 Plane 1 includes 213—not the expected 214—Kangxi Radicals in Rows 7 through 9, 210 of which are duplicates of ideographs in CNS 11643 Planes 1 and 2, and by extension are duplicates of ideographs in Big Five. The three that are mapped are U+4EA0 𠄎 (Radical #8; *kIRG\_TSource* = T1-2728), U+51AB 𠄎 (Radical #15; *kIRG\_TSource* = T1-272F), and U+52F9 𠄎 (Radical #20; *kIRG\_TSource* = T1-2734). While these three ideographs are not present in Big Five, they are present in HKSCS, and have the following *kIRG\_HSource* property values: H-C6C2 (U+4EA0 𠄎), H-C6C5 (U+51AB 𠄎), and H-C6C6 (U+52F9 𠄎).
- The two duplicate ideographs in Big Five Level 2, which are mapped to CJK Compatibility Ideographs, U+FA0C 𠄎 (*kBigFive* = C94A) and U+FA0D 𠄎 (*kBigFive* = DDFC), are not present in CNS 11643 Plane 2. Their canonical equivalents are U+5140 𠄎 (*kBigFive* = A461; *kIRG\_TSource* = T1-4442) and U+55C0 𠄎 (*kBigFive* = DCD1; *kIRG\_TSource* = T2-4176), respectively.
- Two ideograph pairs have different mappings due to different structures or different components. U+5284 𠄎 maps to Big Five (*kBigFive* = E35A), but U+7B9A 𠄎 maps to CNS 11643 Plane 2 (*kIRG\_TSource* = T2-4C61). Note that U+5284 𠄎 maps to CNS 11643 Plane 14 (*kIRG\_TSource* = TE-656C). U+5F5D 𠄎 maps to Big Five (*kBigFive* = C255), but U+5F5E 𠄎 maps to CNS 11643 Plane 1 (*kIRG\_TSource* = T1-7641). Note that U+5F5D 𠄎 maps to CNS 11643 Plane 3 (*kIRG\_TSource* = T3-5764).

It is the third key difference that is the focus of this proposal.

## The Proposal

Implementations, in particular fonts, tend to treat Big Five and CNS 11643 Planes 1 and 2 as being equivalent in terms of their coverage of the Unicode Standard, which means that it is generally prudent for such implementations to include glyphs for both ideograph pairs whose mappings differ between those two regional standards, specifically U+5284 劓 / U+7B9A 劓 and U+5F5D 彛 / U+5F5E 彛. In order to flag these ideographs, I propose that an apostrophe be appended to the equivalent Big Five mappings as follows (new *kBigFive* property records are highlighted in **red**):

U+5284	kBigFive	E35A
U+5F5D	kBigFive	C255
<b>U+5F5E</b>	<b>kBigFive</b>	<b>C255'</b>
<b>U+7B9A</b>	<b>kBigFive</b>	<b>E35A'</b>

The Syntax and Description of the *kBigFive* property as reflected in UAX #38 would need to be changed as follows (additions are highlighted in **red**):

Field	Text
Syntax	[0-9A-F]{4}\' <b>?</b>
Description	The Big Five mapping for this character in hexadecimal; note that this does not cover any of the Big Five extensions in common use, including the ETEN extensions. <b>An apostrophe (') at the end of the property value indicates an alternate Big Five mapping for two ideographs that map differently in CNS 11643, specifically U+5284 劓 (Big Five) versus U+7B9A 劓 (CNS 11643) and U+5F5D 彛 (Big Five) versus U+5F5E 彛 (CNS 11643).</b>

## Data file

The *kBigFive.txt* data file, which is a PDF attachment, provides the proposed *kBigFive* property value additions.

That is all.