

Title: Proposal to enhance the **kRSUnicode** property for non-Chinese simplified radicals

Author: Ken Lunde

Date: 2022-09-02

This is a proposal to enhance the *informative* **kRSUnicode** property of the Unihan database to accommodate non-Chinese simplified radicals, and includes proposed property value changes based on this proposed enhancement.

A limited number of the 214 *Kāngxī Dictionary* (康熙字典 *kāngxī zìdiǎn*) radicals have non-Chinese simplified forms, but there is currently no way to indicate them in the *kRSUnicode* property. Chinese simplified forms of the radicals are reflected by appending an *apostrophe* (U+0027 ' APOSTROPHE) to the radical number, such as 159' for 车, which is the Chinese simplified form of Radical #159 (車).

Note that two simplified radicals are used for both Chinese and Japanese, specifically Radical #90 (丩) and Radical #199 (麥). These radicals are already accommodated by the existing apostrophe notation of the *kRSUnicode* property.

The proposal is to 1) enhance the *kRSUnicode* property by simply appending a *quotation mark* (U+0022 " QUOTATION MARK) to the radical number of non-Chinese simplified radicals in property values; 2) update the relevant sections of [UAX #38](#); and 3) update the relevant *kRSUnicode* property values to reflect this new notation.

The following table lists the six radicals that have non-Chinese simplified forms that are not yet accommodated by the *kRSUnicode* property, and the Traditional and Chinese simplified forms are shown for comparison:

Radical	Traditional Form	Chinese Simplified Form	Non-Chinese Simplified Form
182	U+2FB5 風 U+98A8 風	U+2EDB 风 U+98CE 风	U+322C4 𠂇
208	U+2FCF 鼠 U+9F20 鼠	N A	U+9F21 𠂇
210	U+2FD1 齊 U+9F4A 齊	U+2EEC 齐 U+9F50 齐	U+2EEB 𠂇 U+6589 𠂇
211	U+2FD2 齒 U+9F52 齒	U+2EEE 齿 U+9F7F 齿	U+2EED 𠂇 U+6B6F 𠂇
212	U+2FD3 龍 U+9F8D 龍	U+2EF0 龙 U+9F99 龙	U+2EEF 𠂇 U+7ADC 𠂇
213	U+2FD4 龜 U+9F9C 龜	U+2EF3 龟 U+9F9F 龟	U+2EF2 𠂇 U+4E80 龜

Note that Radical #208 does not have a Chinese simplified form, that the [CJK Radicals Supplement](#) block does not include the non-Chinese simplified forms of Radicals #182 and #208, and that the non-Chinese simplified form of Radical #182 is in the [CJK Unified Ideographs Extension H](#) block.

The following table provides a revised description of Bits 28 through 31 of the sorting algorithm as specified in [Section 2.1.2 of UAX #38](#) (changes and additions are highlighted in **red**):

Bits	Text
28-31	Bits 28-31 are used to indicate whether the entry has a simplified form for the radical or not. A value of 1 indicates the Chinese simplified form of the radical (for example, 齿); a value of 2 indicates the non-Chinese simplified form of the radical (for example, 𠂇); and a value of 0 indicates the traditional form for the radical (for example, 齒).

The following table provides the revised Syntax and Description for the *kRSUnicode* property as reflected in [Section 4.1 of UAX #38](#) (changes and additions are highlighted in **red**):

Field	Text
Syntax	[1-9][0-9]{0,2}[\'\"]?\.-[0-9]{1,2}
Description	<p>The standard radical-stroke count for this character in the form “radical.additional strokes”. The radical is indicated by a number in the range (1..214) inclusive. An apostrophe (') after the radical indicates a Chinese simplified version of the given radical. A quotation mark (") after the radical indicates a non-Chinese simplified version of the given radical. The “additional strokes” value is the residual stroke-count, the count of all strokes remaining after eliminating all strokes associated with the radical.</p> <p>This field is also used for additional radical-stroke indices where either a character may be reasonably classified under more than one radical, or alternate stroke count algorithms may provide different stroke counts.</p> <p>The residual stroke count may be negative. This is because some characters (for example, U+225A9 and U+29C0A) are constructed by removing strokes from a standard radical.</p>

In addition to 1) updating Section 2.1.2 of UAX #38 to update the sorting algorithm as specified in the first table above; 2) updating Section 4.1 of UAX #38 to reflect the updated Syntax and Description of the *kRSUnicode* property as specified in the second table above; and 3) updating the relevant *kRSUnicode* property values in the Unihan database using the supplied data file, the tools that produce the code charts will need to be updated to support the new notation and to display the non-Chinese simplified form of the radical. The code chart excerpt below for U+9F50 齐 illustrates how this is currently done for Chinese simplified radicals:

9F50			
齐 210'.0			
文 67.2			
	G0-466B	H-A0FC	T3-2370

Data file

The *kRSUnicode-data-2.txt* data file, which is a PDF attachment, provides the proposed property value changes for the Unihan database.

That is all.