

Title: Proposal for updated rules for assigning Primary and Secondary Radicals
Author: Henry Chan
Source: Individual Contribution
Date: 1 March 2025

1. Background

In WS2017 and WS2021, there has been an ongoing debate on the assignment of radicals for characters where the radical selection was not intuitive or obvious.

Some experts favoured selecting a radical on semantic grounds, while others favoured selecting the radical on structural grounds.

The debate was partially settled by the introduction of secondary radicals as a formal part of IRG's review process, with preferences for an "intuitive" primary radical.

Section 2.2.1.d.(5).c) of the IRG P&P currently states the following:

- > If the technically correct (aka semantic) radical for an ideograph hampers its
- > discoverability, or is region-dependent, the primary radical shall be assigned as though
- > made by an ideograph expert who is neither a specialist in the history of the Han script nor
- > familiar with ideograph etymology. The technically correct radical can be assigned as a
- > second radical. Both are shown in the code charts, though the primary one serves as the
- > basis for ordering within a CJK Unified Ideographs block.

However, after the introduction of secondary radicals, a substantial amount of meeting time is now spent on discussing the addition of secondary radicals.

The large number of secondary radicals being recorded greatly increases the time for the ORT manager to process the meeting minutes.

2. Problem 1 - Additional radicals added for sake of completeness

Additional secondary radicals were suggested even in cases where the originally primary radical was “intuitive”, and added just for completeness.

All characters submitted for encoding will be assigned a Kangxi Radical between 1 - 214. The primary radical determines its position in the code chart. The secondary radical is purely informative.

Adding extra potential radicals to facilitate easier lookup is not within IRG’s scope.

IRG’s meeting should concern itself primarily with encoding matters, not the compilation of dictionary data.

Additional dictionary data, while useful for the general public, could be discussed within the CJK & Unihan Working Group.

Action Items:

- The ORT manager will amend the ORT to have the capability to suggest additional radicals for Unihan.

- Reviewers are requested to resolve the comments for the radicals suggested for completeness, and re-submit the suggested radicals under the newly added comment type.

3. Problem 2 - Idea of “technically correct (aka semantic) radical”

3.1 Explainer - Part 1

The text alludes that there is a concept of a “technically correct radical” and equates it with a “semantic radical”.

This understanding seems to stem from an over-extension of the existing radical system in Kangxi Dictionary.

The original radical system in Shuowen had 540 radicals. This was kept roughly stable up until and including the Sung Dynasty. These radicals were based by grouping characters according to shared semantic elements.

The radicals were gradually whittled down for user convenience by dictionary authors. We arrived at the 214 radical system in the Ming Dynasty with 《字彙》. Kangxi Dictionary also adopted this convention.

In this radical system, where characters belonged to radicals which were removed, characters were re-assigned to other radicals and often were no longer assigned by semantic elements.

For example:

舅: 臼 phonetic + 男 semantic => 男 radical was removed => 臼 is assigned as the radical.

甥: 生 phonetic + 男 semantic => 男 radical was removed => 生 is assigned as the radical.

Using the actual application of the “technically correct (aka semantic) radical” as practiced by the IRG in previous meetings, this would result in the above those characters assigned as 田, as 男 itself is assigned radical 田. This would result in an un-intuitive radical.

IRG adopts the Kangxi radical system, and the Kangxi radical system is not strictly semantic based, so the idea of “technically correct (aka semantic) radical” is flawed.

3.2 Explainer - Part 2

The vast majority of characters in the Kangxi dictionary and other dictionary for characters used in the Chinese language are phono-semantic characters, composed of a 形符 “form element” (a type of semantic element) and 聲符 “phonetic element”.

Of these “form elements”, the majority are non-composed characters (獨體字) and are radicals themselves. This means selecting the semantic element to assign the radical will nearly always result in an intuitive radical.

The vast majority of characters which caused a non-intuitive radical to be assigned in WS2017 and WS2021, while their grouping can also be noted as phono-semantic characters, are a different class of characters. These characters are composed of one 義符 “meaning element” (another type of semantic element) and one 聲符 “phonetic element”. These character classes were more commonly seen in characters of Vietnamese and Zhuang origin.

Of these “meaning element”, the majority are composed characters (合體字), and are not radicals themselves. More than often extracting the radical of these elements to be the radical of the whole character would yield a radical that is nested several levels deep in the character, and not be intuitive at all.

These are by nature two different classes of characters*. The Kangxi radical system is built for the first class of phono-semantic characters, and not the latter class of phono-semantic characters.

The problem was actually caused by IRG experts which attempted to extend existing conventions to characters the conventions are fundamentally not suitable for.

* In reality, the distinction between the two is not always clear cut, but the proposal below will result in the same radical if it is analyzed either way.

3.3 Explainer - Conclusion

Given that Kangxi radical system also assigns radicals to the phonetic element, I suggest we dispose of the idea of a “technically correct (aka semantic) radical”, and simply establish easy to follow systematic rules for assigning a default radical, which should yield an intuitive radical in the majority of cases.

3.4. Proposal

3.4.1 Goals of this proposal

- Standardize the assignment of the primary radical in a way that is mostly consistent with the radicals assigned by the Kangxi Dictionary
- Reduce the cases where secondary radicals can be assigned

3.3. Proposal for selecting primary radicals

I am proposing the following process for assigning the primary and secondary radical for new characters.

This can be applied from the next working set, or current working set, subject to discussion from IRG.

1. If the character can be trivially analyzed as a traditional phono-semantic character (形聲字) composed of a form element + phonetic element:

If the character cannot be **trivially** analyzed, these rules should not be used.

- a. If the form part is a radical, then the form part will be selected to be the sole radical.

Example: 雞: 奚 phonetic + 隹 form => 隹 is the radical.

Example: 花: 艹 form + 化 phonetic => 艹 is the radical.

Example: 羸: 羸 phonetic + 貝 form => 貝 is the radical.

Example: 條: 攸 phonetic + 木 form => 木 is the radical.

Example: 勝: 朕 phonetic + 力 form => 力 is the radical.

- b. Otherwise, if the form part is not a radical, but the phonetic part is, then the phonetic part will be selected to be the sole radical.

Example: 舅: 臼 phonetic + 男 form => 臼 is the radical.

Example: 甥: 生 phonetic + 男 form => 生 is the radical.

- c. Otherwise, the radical of the form part will be used as the sole radical.

Example: 饑: 會 form + 卑 phonetic => 日 is the radical.

A secondary radical can be assigned only if there is evidence to analyze a character both ways. Theoretical decompositions are not considered.

2. Otherwise, if a character is a non-decomposable character (獨體字):

- a. If the character contains a center vertical line crossing the whole character, it shall be selected as the sole radical.

Example: 中: vertical centerline => | is the radical.

- b. The closest looking radical can be selected.

Example: 由: looks like 田 => 田 is the radical.

Example: 畚 looks like 谷 => 谷 is the radical.

- c. The first N strokes or last N stroke may be selected as the sole radical, where N is any positive integer.

Example: 光: last two strokes => 儿 is the radical.

To match the behaviour of the Kangxi Dictionary, if the last stroke is a CJK Stroke N, it shall be filed under Radical 4.0 (丿), not Radical 3.0 (㇇).

Example: 之: last stroke ㇇ => 丿 is the radical.

3. Otherwise, if the character is a decomposable character (合體字):

- a. If the character is a surrounding structure (U+2FF4 to U+2FFD):

- i. If the surrounding component is a radical, it will be chosen as the sole radical.

Example: 圍: 口 outer component + 韋 inner component => 口 is the radical.

- ii. Otherwise, the radical of the surrounding component will be chosen as the sole radical.

Example: 𨾏: 更 outer component + 益 inner component => radical of 更, 日 is the radical.

The component being surrounded will not be used to produce any radical.

- b. If the character is a left-middle-right / top-center-bottom structure (U+2FF2, U+2FF3):

- i. If the left and right components or the top and bottom components can be combined into a radical, it will be chosen as the sole radical.

Example: 𠄎: 二 outer component + 月 inner component => 二 is the radical.

Example: 衡: 行 outer component + 角 & 大 inner components => 行 is the radical.

The component being surrounded will not be used to produce any radical.

- c. Otherwise, any top-level component which is a radical may be selected to be the radical.

If there is more than one candidate, the radical with the most strokes will be chosen as the sole radical.

Example: 酒: 氵 + 酉 => take 酉 with most strokes.

- d. Otherwise, a top level component with the smallest amount of residual strokes should be selected to produce the radical.

Example: 陵: 巴 + 陵

巴: Radical 己 with 1 residual stroke

陵: Radical 阜 with 10 residual strokes

=> take 己 as it produces the least residual strokes

3.4 Benefits of the proposal

- The proposal eliminates semantic assignment except in the trivial case, where the rules guarantee it will produce an intuitive assignment while preserving semantic bias where possible.
- The proposal does not guarantee to produce the exact same result as the Kangxi Dictionary, but would be in agreement with a majority of the commonly used characters.
- This proposal would result in an intuitive radical in the majority of cases.

There are cases where it is impossible to give an intuitive radical. Above rules strive to guarantee a deterministic radical.

There are cases where it will generate a less “intuitive” radical for a character:

𠃉 now gives radical 己 instead of 二.

However, considering all characters with the same structure:

𠃉 𠃊 𠃋 𠃌 𠃍 𠃎 all now give radical 己 instead of 二 一 己 支 皮 角 角 respectively.

3.5 Effects on WS2024

If the above rules were to be applied to the characters in WS2024, here is the effects on an excerpt of characters which currently have suggested secondary radical changes:

1. [03144](#) - will result in a sole radical of 长 based on rule 3(c). Current radical 肉.
2. [03586](#) - will result in a sole radical of 羊 based on rule 3(c). Current radical 貝.
3. [04278](#) - will result in a sole radical of 骨 based on rule 1(b). Current radicals 骨, 虎.
4. [04548](#) - will result in a sole radical of 鳥 based on rule 3(c). Current radical 鳥.
5. [04570](#) - will result in a sole radical of 彡 based on rule 3(c). Current radical 鳥.
6. [04589](#) - will result in a sole radical of 鸟 based on rule 1(a). Current radical 鸟.
7. [04595](#) - will result in a sole radical of 鸟 based on rule 1(a). Current radical 鸟.

Case 1, 2, 3 and 5 results in the elimination of the current radical and replaced with an intuitive radical with the non-obvious semantic radical being discarded.

Case 3, 6 and 7 result in the elimination of a secondary radical and the existing (intuitive) primary radical kept.

EOF