

## IRGN2474 Feedback: Proposal to update IRG PnP for Non-cognate and UCV handling

Source: Henry Chan, Eiso Chan, Yi Bai

Status: Individual Contribution

Action: To be considered by IRG

Date: 2021-09-01

Pages: 2

### Introduction

In previous IRG discussions, there is considerable controversy of the addition of new UCV rules which unify current forms of components with historical forms where the same historical form may be shared by different components in different historical periods.

The first recorded case of this issue is in IRG #51, where the UCV rule unifying 𠄎 and 𠄏 was controversial because 𠄎 is also a common variant form of 𠄎. 𠄎 and 𠄏 were added as UCV rule #188a with the description "Only for rain".

The second recorded case of this issue is in IRG #56, where the UCV rule unifying 𠄎 and 𠄏 was controversial because 𠄎 form as a component is not always etymologically related to 𠄎. 𠄎 and 𠄏 were added as UCV rule #439 with the description "Only for etymological relation with 𠄎(tile)".

There is some resistance to the addition of these rules because of the concern of over-application of UCV rules leading to increased misunified or over-unified characters. Also there is a concern that the proof of burden was not clear.

### Proposal

This document proposes that:

1. The meaning of non-cognate rule should be updated to reflect actual usage by IRG. The PnP should clarify the long standing convention that characters related in historical derivation are also applicable for the non-cognate rule as long as the semantics are sufficiently different. It is suggested to add the following text in 2.1.3. *Non-cognate Rule*:

*Characters that are related in historical derivation may also be disunified as long the difference in glyph shape is sufficient for reflecting different semantics. An example is 間 and 間 which are related in historical derivation but are disunified because they have different semantics and typically used in mutually exclusive contexts in present day use. For the purpose of IRG processes, these characters have been and are still considered applicable for the non-cognate rule even if they are related in historical derivation and technically cognate.*

2. The prior precedent for constraining the application of a particular UCV rule to an etymological relationship be written as a standard procedure into the PnP. It is suggested to add the following text under section 2.5.2 *Discussion Procedure*: (1) *For unification issues* to clarify the circumstances of enforcing this requirement:

*To avoid misunification of ideographs, IRG may specify etymological constraints to the application of a particular UCV rule, i.e. an etymological relationship must be proven between a proposed character and a coded or another Working Set character for the rule to be applicable.*

3. The circumstances for adding an etymological relationship for a UCV rule be specified. It is suggested to add the following text subsequent to the previous paragraph:

*An etymological constraint will only be specified for a particular UCV rule when one or more of the suggested unifiable forms may typically be etymologically related to another radical or component, thus the rule is at high risk of causing misunification of unrelated characters.*

4. The proof of burden be specified. It is suggested to add following text subsequent to the previous paragraph:

*The unification shall only apply when there is sufficient evidence to prove that the two characters in discussion are etymologically related, i.e. the proof of burden lies in the reviewer instead of the submitter. A proposed character will also not be postponed unless there is reasonable doubt that the character is etymologically related with another coded or Working Set character.*

5. Re-iterate that characters with different semantics are still applicable for disunification under the non-cognate rule even if it satisfies the etymological requirement. It is suggested to add following text subsequent to the previous paragraph:

*Even if an etymological relationship can be proved between two characters, the non-cognate rule still applies, i.e. as long as there is sufficient evidence to show that the two characters are used with mutually exclusive semantics in a certain region or language, the characters will not be unified.*