

IRG N713R

Title: Comments on JCS Proposals
Source: Unicode Technical Committee
Author: John H. Jenkins, UTC Liaison to IRG
Distribution: IRG
Action: For Review and Response by IRG
Date: 8 December 1999

Clarification of the Non-Cognate Rule

While reviewing document N709, which discusses the status of JPN095 and its unification with U+6747, the subject of the status of the Non-Cognate Rule was raised in the meeting of the Editorial Committee. Japan argues in N709 that JPN095 should be deunified from U+6747 on the basis of its having a different etymology, pronunciation, and meaning. This would invoke the non-cognate rule and require the two to be separated.

There was concern expressed that Extension B might be seriously destabilized if IRG members were to start invoking the non-cognate rule at this point in order to deunify current unifications found in Extension B. There might potentially be thousands of candidates for deunification on this basis.

In the past, the editorial committee has determined whether or not to invoke the non-cognate rule based solely upon the meaning of characters as defined in the KangXi dictionary. JPN095, however, is a character of Japanese origin and isn't found in KangXi.

Another problem arose in connection with JPN053. Inclusion of this character would require the deunification of the shapes used for U+XXXX. Such deunification is justifiable under the non-cognate rule, but it seems generally inadvisable.

The following recommendations are brought to the attention of the plenary for consideration:

- 1) The non-cognate rule may be applied by use of any of the standard IRG dictionaries, in the same order as is used for the four dictionary sorting algorithm. That is, if two characters are both found in the KangXi dictionary, then the KangXi dictionary is to be used to determine whether the characters are cognate or not. If the two characters are not both in the KangXi dictionary, then the Morohashi dictionary is to be used, and so on. The plenary may further define how to determine whether or not two characters are cognate if the two are not both found in any of the standard IRG dictionaries.
- 2) In order to avoid destabilizing Extension B and delaying the adoption of ISO/IEC 10646-2, no further candidates for deunification based upon the non-cognate rule

should be accepted at this point. Deunifications based upon the non-cognate rule should be deferred until work begins on Extension C.

The Editorial Committee requests that the IRG plenary adopt resolutions to clarify how the non-cognate rule is to be invoked.