

## Revised Regular Expressions in UAX #38

John H. Jenkins

7 January 2019

There are a number of regular expressions giving the syntax for field values in UAX #38 which are of the form  $U\backslash+2?[0-9A-F]\{4\}$ . In order to accommodate Extension G, these will all need to change to  $U\backslash+[23]?[0-9A-F]\{4\}$ .

This is just a formal document noting this fact so that we can get a formal action item to get the work done. The fields affected are:

kCompatibilityVariant\*  
kHDZRadBreak\*  
kJinmeiyōKanji\*  
kJōyōKanji\*  
kKoreanName\*  
kSemanticVariant  
kSimplifiedVariant  
kSpecializedSemanticVariant  
kTraditionalVariant  
kZVariant

The fields marked with an asterisk do not actually require the change because their data would not be affected by Extension G.