

2009-04-29

Title: Possibly incorrect Bidi Class for Rumi numerals
Author: Roozbeh Pournader
Date: 2009-04-29

Rumi numerals have been accepted in Unicode in the UTC #114 meeting (February 2008):

[114-C41] Consensus: Accept 31 Rumi numeral symbols at U+10E60..U+10E7E for encoding in a future version of the standard in a new block "Rumi Symbols" at U+10E60..U+10E7F, with names and properties as given in L2/07-235. [L2/07-235, L2/06-291]

The properties document, L2/07-235 gives the Bidi Class as “L” (Left-to-Right). This is odd.

First, there is an assumption that there is supposed to be no “L” characters encoded in the range U+10800..U+10FFF (see Mark Davis, “Allocating Unicode Characters”, L2/09-120).

Then, shouldn't these perhaps be “AN”s (Arabic Number)? That would keep them left-to-right, but would also stop them from acting *strongly* when the bidirectional algorithm is applied to them.

The proposals for Rumi numerals (L2/06-291 and L2/07-235) do not say much on how and why the specific character properties were chosen. There is a chance that “L” was chosen without enough consideration of the possible impacts of choosing strong bidi types.

The author believes that this should be investigated and either fixed, or explained in the text of the future version of the Unicode standard.