

Title: Sorting all RTL scripts together
Author: Roozbeh Pournader (Google)
Date: 2012-11-03
Action: For UTC’s action

Suggested action

In UCA data files, move all right-to-left scripts next to each other, and keep this as a guideline for new right-to-left scripts added to the standard.

Background

In indexes, bibliographies, and various other sorted lists, right-to-left strings are sorted together, either at the beginning or the end of the list. The location at the beginning or the end of the list is usually determined by the direction of the main document or interface, but may need to be configurable. Some authors or systems may even split the list based on the direction of the elements, coming up with two separately-sorted right-to-left and left-to-right lists.

This was done for the visual benefit of readers. Such lists are far easier to both layout and look through. Otherwise, we end up with lists like this:

- Left-to-right item 1
 - Left-to-right item 2
 - Left-to-right item 3
- 1 Right-to-left item ●
 - 2 Right-to-left-item ●

As opposed to:

- Left-to-right item 1
 - Left-to-right item 2
 - Left-to-right item 3
- 1 Right-to-left item ●
 - 2 Right-to-left item ●

which is both more visually appealing and browsable.

The same guidelines apply to user experience in modern applications, where strings in the direction of the user’s preferred language may appear first, followed by strings in the opposing direction.

Achieving this with the current data in UCA would be hard. Currently, right-to-left scripts are disorganized in the UCA. There are two large blocks of scripts separated by various left-to-right ones, with five other scripts scattered through the rest. Here are the current blocks of right-to-left scripts, with the intervening left-to-right scripts shown in brackets:

- [Latin..Armenian]*
- Hebrew, Phoenician, Samaritan, Arabic, Syriac, Mandaic, Thaana, N’Ko
- [Tifinagh..Brahmi]*
- Kharoshthi
- [Thai..Runic]*
- Old Turkic
- [Vai..Carian]*
- Lydian
- [Old Italic..Linear B]*
- Cypriot, Old South Arabian, Avestan, Imperial Aramaic, Inscriptional Parthian, Inscriptional Pahlavi
- [Ugaritic..Egyptian Hieroglyphs]*
- Meroitic Hieroglyphs and Meroitic Cursive (interspersed together)
- [Han]*

Moving all right-to-left scripts together would also make it easier for CLDR to have defaults or parameters that move scripts based on the direction of the locale. For example, an Arabic user could have all RTL strings come up first, while an English users could have all RTL strings come up last.

It would also be useful if inside such a continuous block of right-to-left scripts, we keep the scripts in active modern use before the scripts with academic or historic-only usage. The order of the right-to-left scripts could be kept almost as is, but we would at least need to move Phoenician after N’Ko. That would also result in easier customization in CLDR.