Formation of the Reph in Bengali Script

Submitted by Paul Nelson, Microsoft Corporation

The preview document of Chapter 9 of the Unicode 4.0 book has text that was added since the Unicode 3.0 book. The section 9.2 Bengali provides solutions for some issues that are faced with Bengali script. One of these deals with the Ya-Phala.

The last two examples in that sub-section are shown below. This behavior specified causes the formation of the reph form to become Ra + Halanth + ZWJ, which is different than the behavior previously defined for the reph (second line).

র + ্ + য + ◌া = র্যা [ræ]

র + ্ + [ZWJ] + য + ◌া = র্যা [rjaː]

When processing Indic script text, the reph form is specified in R2 of the Rules for Rendering in section 9.1 of the Unicode book. Additionally, the reph form is always processed before the yaphala form. The situation which the change to the Unicode Book was meant to address is important. However, there are some problems it generates.

The issues with this change are:

1. Existing documents are broken. The combination Ra + Halanth + Ya previously caused the Ya_reph to be shaped.
2. An inconsistent manner to form the reph is introduced which is confusing for people entering text and inefficient for rendering agents unless the reph is always formed by requiring the ZWJ after the Ra + Halanth.

To resolve this issue, I propose correcting the Unicode 4.0 document to the following:

র + ্ + য + ◌া = র্যা [rjaː]

র + [ZWJ] + ্ + য + ◌া = র্যা [ræ]

As the Unicode Book is most likely nearly printed, it is urgent that this matter be addressed immediately to prevent confusion which is caused by the example in the preview book.